Multimodal Meta-Learning for Time Series Regression

First Ideas Talk

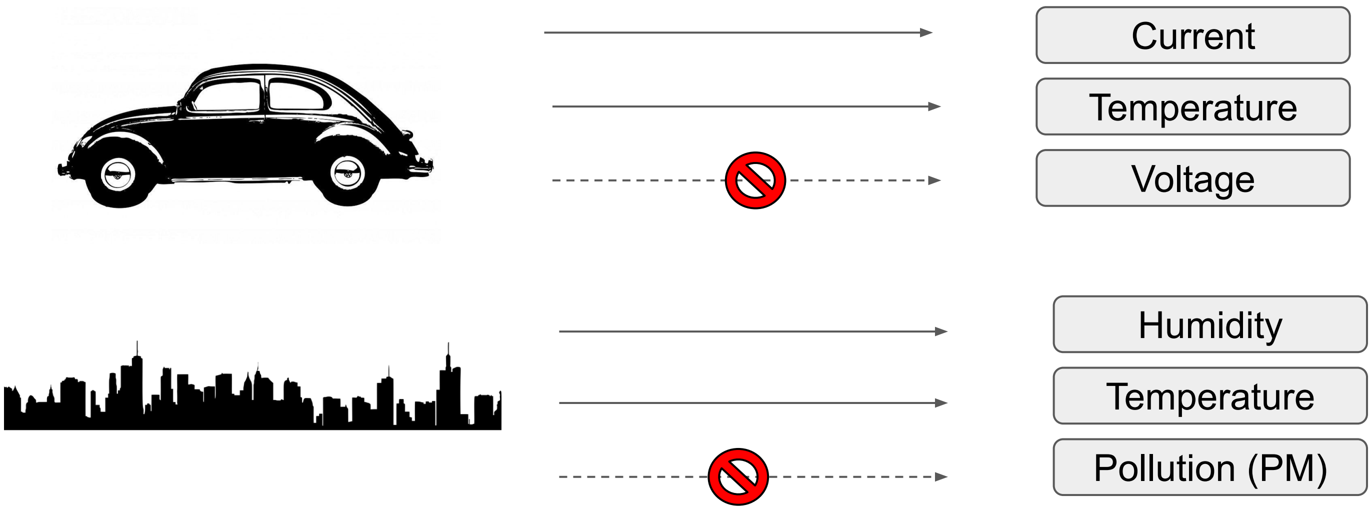
Sebastian Pineda Arango
Supervised by: Vijaya Krishna Yalavarthi and Felix Heinrich

Stiftung Universität Hildesheim 2003

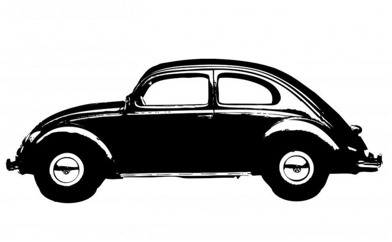
Motivation

Motivation

Regression is interesting when we do not have sensors for some variables.



Motivation



Current

Temperature

Voltage

Current

Temperature

Model

Voltage

Historical data

Current

Temperature

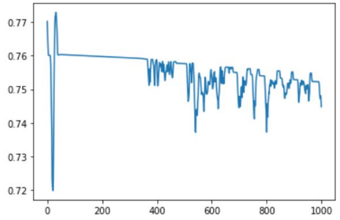
Model

Voltage

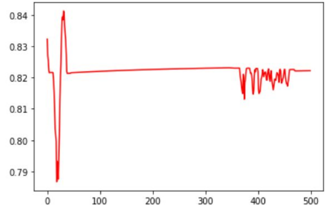
Motivation

We want to **leverage** the data from other **similar** sources (tasks) so that we can learn with less data in **new** sources (tasks).

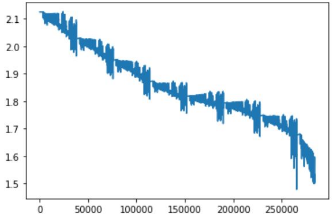
Berlin pollution measurements



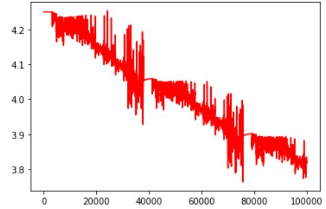
Paris pollution measurements



Battery signals (First year)

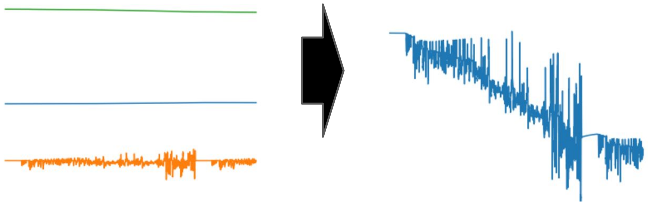


Battery signals (10 years)



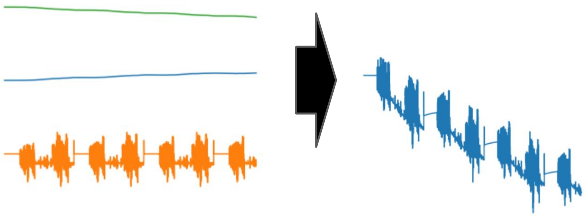
Few-shot Learning for Multivariate Time Series

Task 1: Electric signals under condition 1



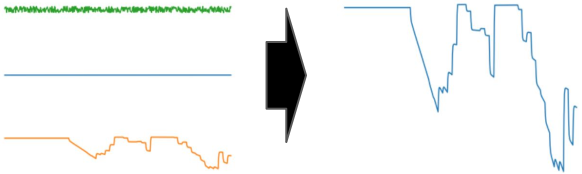
$L = 10000$

Task 2: Electric signals under condition 2



$L = 10000$

Task 3: Electric signals under unseen conditions

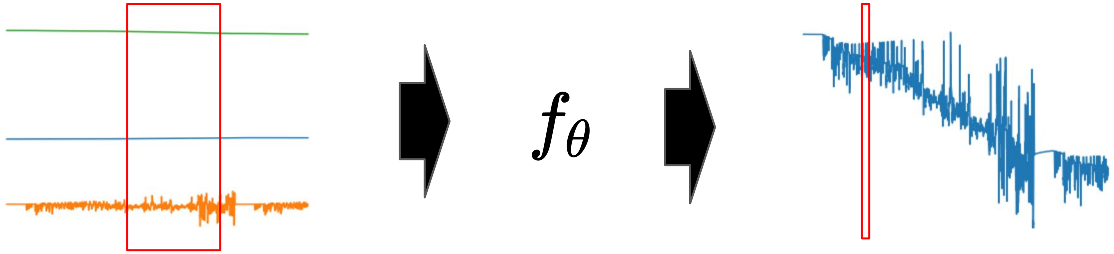


$L = 100$

Problem formulation:
Time Series Regression (TSR) Task

A **time series regression task** is formulated as:

- Given a multivariate time series $\mathbf{s}_i \in \mathrm{R}^{L \times C}$, learn the parameters θ of a model $f_\theta : \mathrm{R}^{L \times C} \rightarrow \mathrm{R}$ that predict the current value Y_i of another time series (target time series).



Problem formulation: Few-shot TSR

A **few-shot time series regression problem** is formulated as:

Given a set of tuples of multivariate time series and the target time series:

$$\mathcal{S}^{Tr} = \{(\mathbf{S}_i^{Tr}, Y_i^{Tr}) |\ \mathbf{S}_i^{Tr} \in \mathrm{R}^{L^{Tr} \times C}, Y_i^{Tr} \in \mathrm{R}^{L^{Tr}} i = 1, \ldots, N^{Tr}\}$$

we want to learn how to adapt fastly the parameters of a model $f_\theta : \mathrm{R}^{L \times C} \rightarrow \mathrm{R}$ to a new set of time series with less data (less history).

$$\mathcal{S}^{Te} = \{(\mathbf{S}_i^{Te}, Y_i^{Tr}) |\ \mathbf{S}_i^{Te} \in \mathrm{R}^{L^{Te} \times C}, T_i^{Te} \in \mathrm{R}^{L^{Te}}, i = 1, \ldots, N^{Te}\}$$

Related work

Model-Agnostic Meta-Learning (MAML)

Authors	Finn et al.
Conference	ICML
Year	2017
Main idea	Find initial parameters that enable fast adaptation to new tasks (less data and less iterations).
Remarks	Applied to image classification

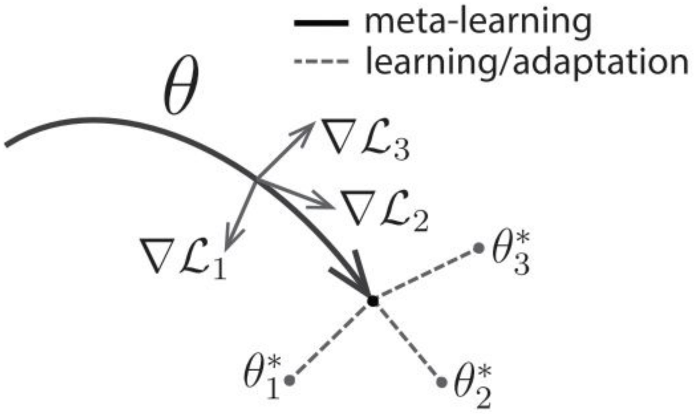


Image taken from the original paper

Multimodal Model-Agnostic Meta-Learning (MMAML)

Authors	Vuorio et al.
Conference	NIPS
Year	2019
Main idea	Find initial parameters that enable fast adaptation to new tasks and by conditioning the parameters to task-specific embeddings.
Remarks	Apply to image classification, RL and regression

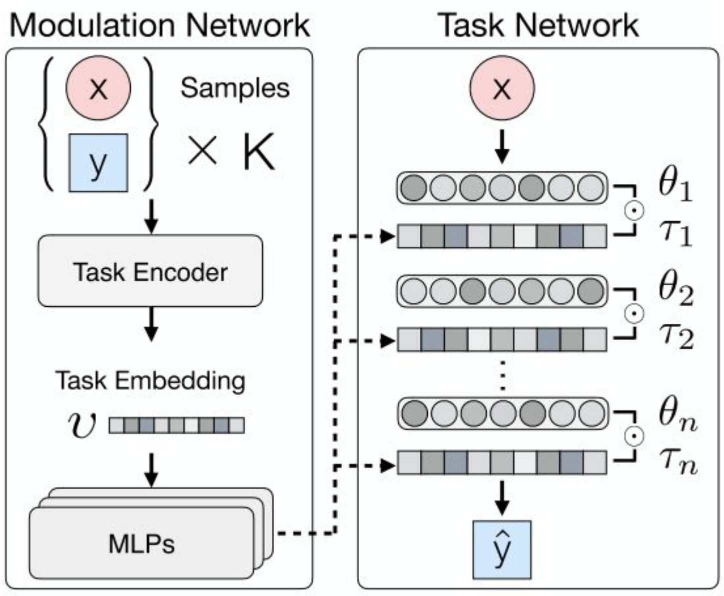


Image taken from the original paper

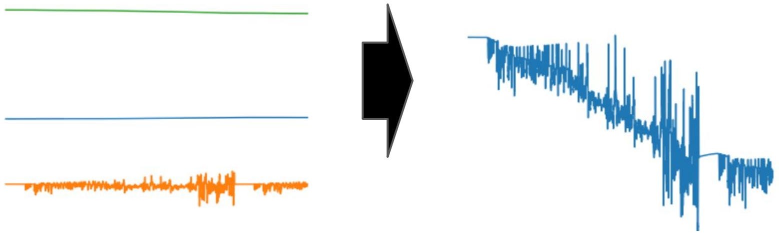
More on related work

- **Transfer learning on time series**
  - *Transfer learning for time series classification*, by Fawaz et al. (2018) in IEEE International Conference on Big Data.
- **Data augmentation on time series**
  - *Data Augmentation for Time Series Classification using Convolutional Networks*, by Guennect et al. (2016) in ECML.
- **More about meta-learning**
  - *Tadam: Task dependent adaptive metric for improved few shot learning*, by Oreshkin et al. (2018), in NIPS.
  - *LEO. Meta-learning with latent embedding optimization*, by Rusu et al. (2019) in ICLR.

Proposed solution

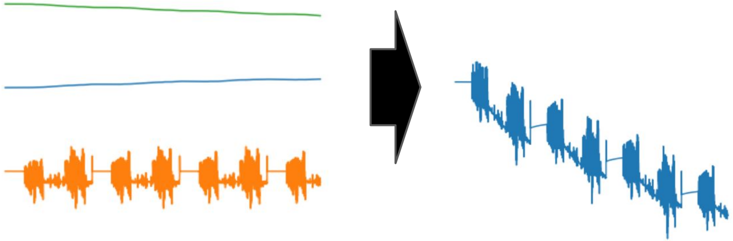
Few-shot Learning for Time Series

Task 1



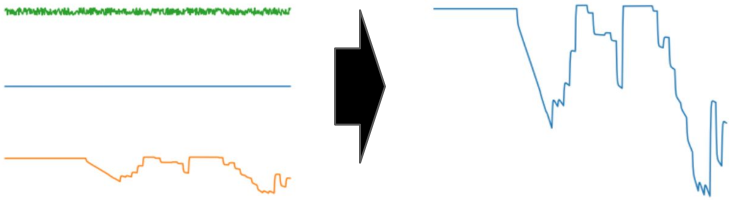
$L = 10000$

Task 2



$L = 10000$

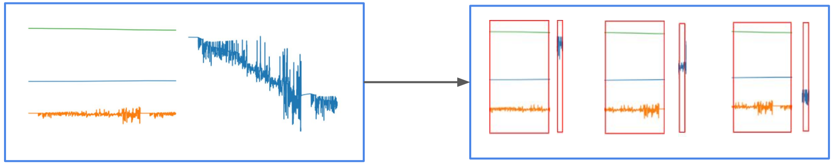
Task 3

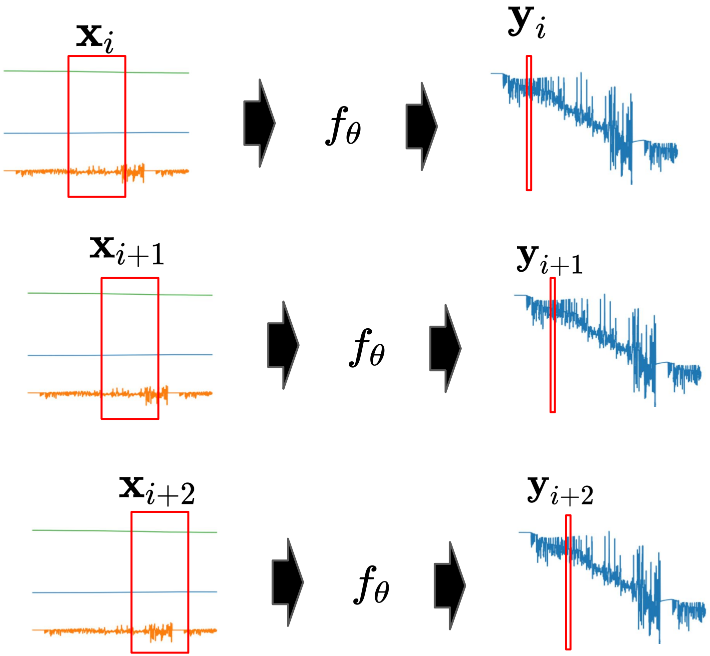


$L = 100$

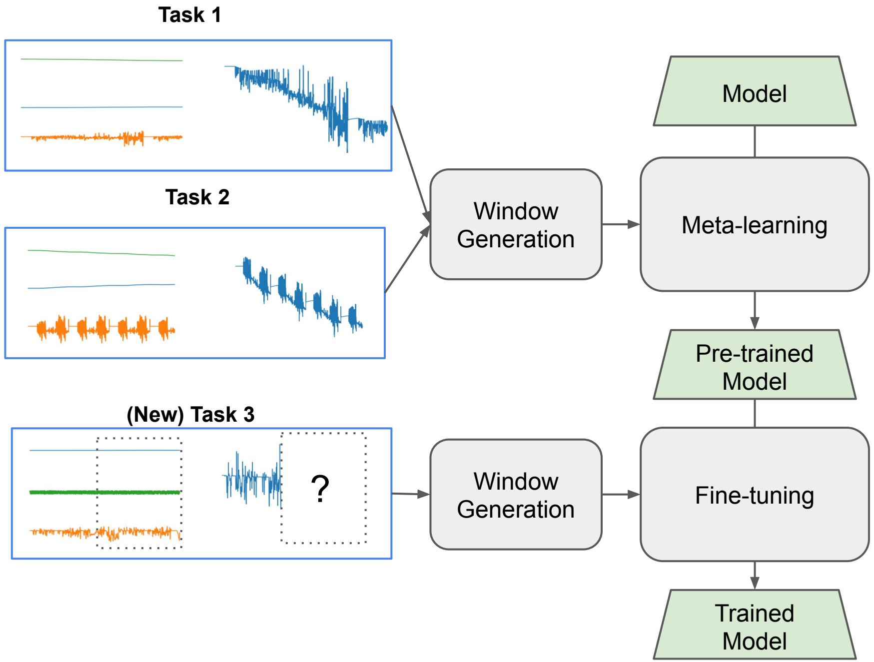
Windows Generation

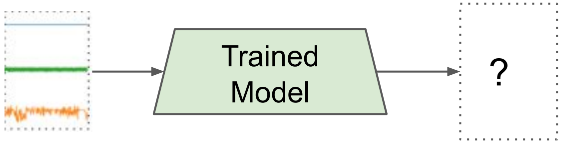
Traditionally, the time series learning implies to create a set of windows with the respective target.





MAML for TSR: Training & Testing





- Faster adaptation
- Less data hungry

MMAML for TSR: Concept

- Task: $\mathcal{T}_i$
- Task embedding: $v_i = h_{\phi^{enc}}(\mathcal{T}_i)$
- Generated parameters: $\tau_i = g_{\phi^{gen}}(v_i)$
- The generated parameters modify the task networks through FiLM layers.

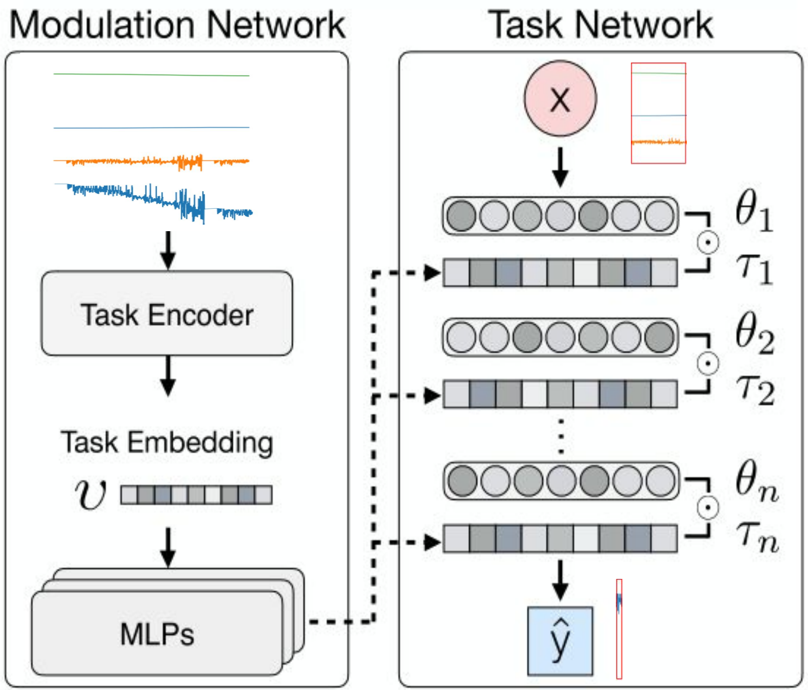
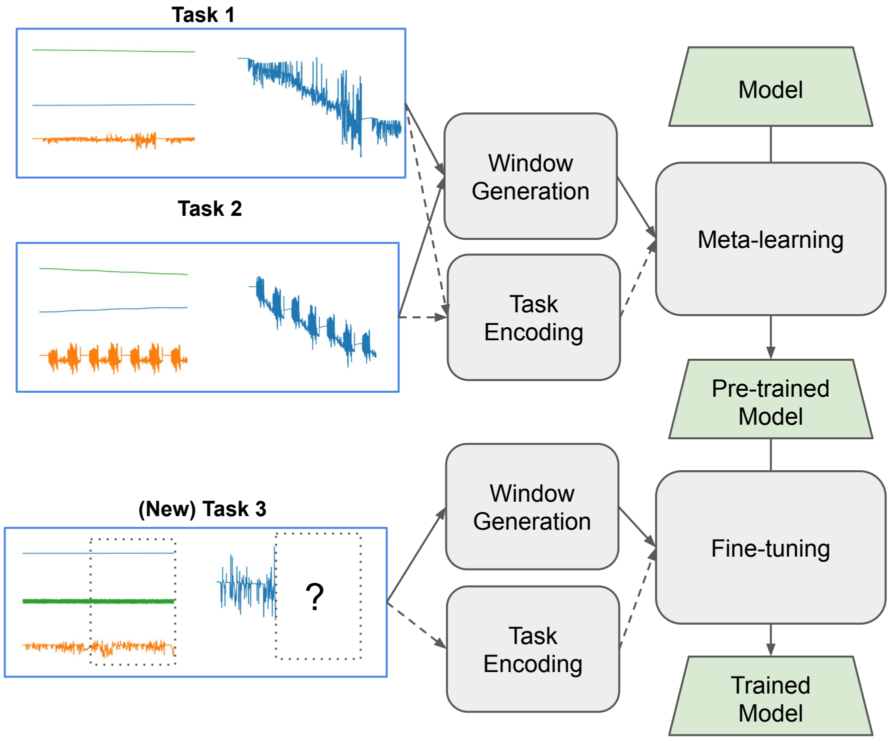
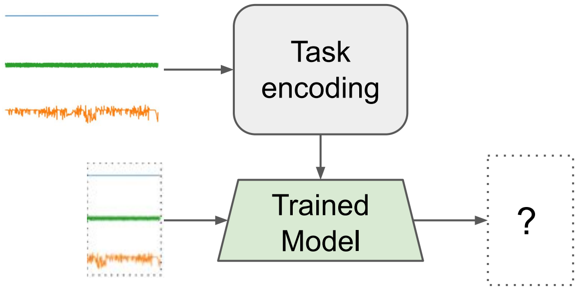


Image modified from the original paper

MMAML for TSR: Training & Testing





- Hierarchical information
- Cluster of tasks in embedding space

Methodological considerations

Objectives

- Formulate a framework that extends MAML formulation into Multivariate Time Series Regression.
- Show how task-conditioning improves generalization and adaptation in time series.
- Design a serie of few-shot tasks and use them to evaluate the framework by comparing with different baselines.

Baselines

- **Gaussian Processes:** simple and good for few data [6].
- **Rocket:** SOTA in Time Series Classification [5].
- **LSTNet**: SOTA in Multivariate Time Series Prediction [4].

vs. **Proposed Model**: **MMAML** over Task Network + Modulation Network (RVAE)

Data foundation

Data	Source	Description
Battery signals	Volkswagen	Measurements for different agings (0-10 years), under different driving conditions.
Five Cities PM Data	UCI Repository	Data about PM particles, 5 cities, 5 years.
EEG Visual Signals Data Set	UCI Repository	Visual experiments to create brain-interfaces

Timeline

Time	Tasks
August	Literature review, data exploration
September	Baselines implementation
October	Proposed model implementation
November	Experiments on models (Hyper. Tun., etc.)
December	Results evaluation and adjustments
January	Results report and thesis finalization

References

[1] Vuorio et al. *Multimodal Model-Agnostic Meta-Learning Via Task-Aware Modulation*, ICLR (2019).

[2] Finn et al. *Model-agnostic meta-learning for fast adaptation of deep networks*, ICML (2017).

[3] Pérez et al. *FiLM: Visual reasoning with a general conditioning layer*, AAAI (2018).

[4] Lai et al. *Modeling long- and short-term temporal patterns with deep neural networks*, ACM SIGIR (2018).

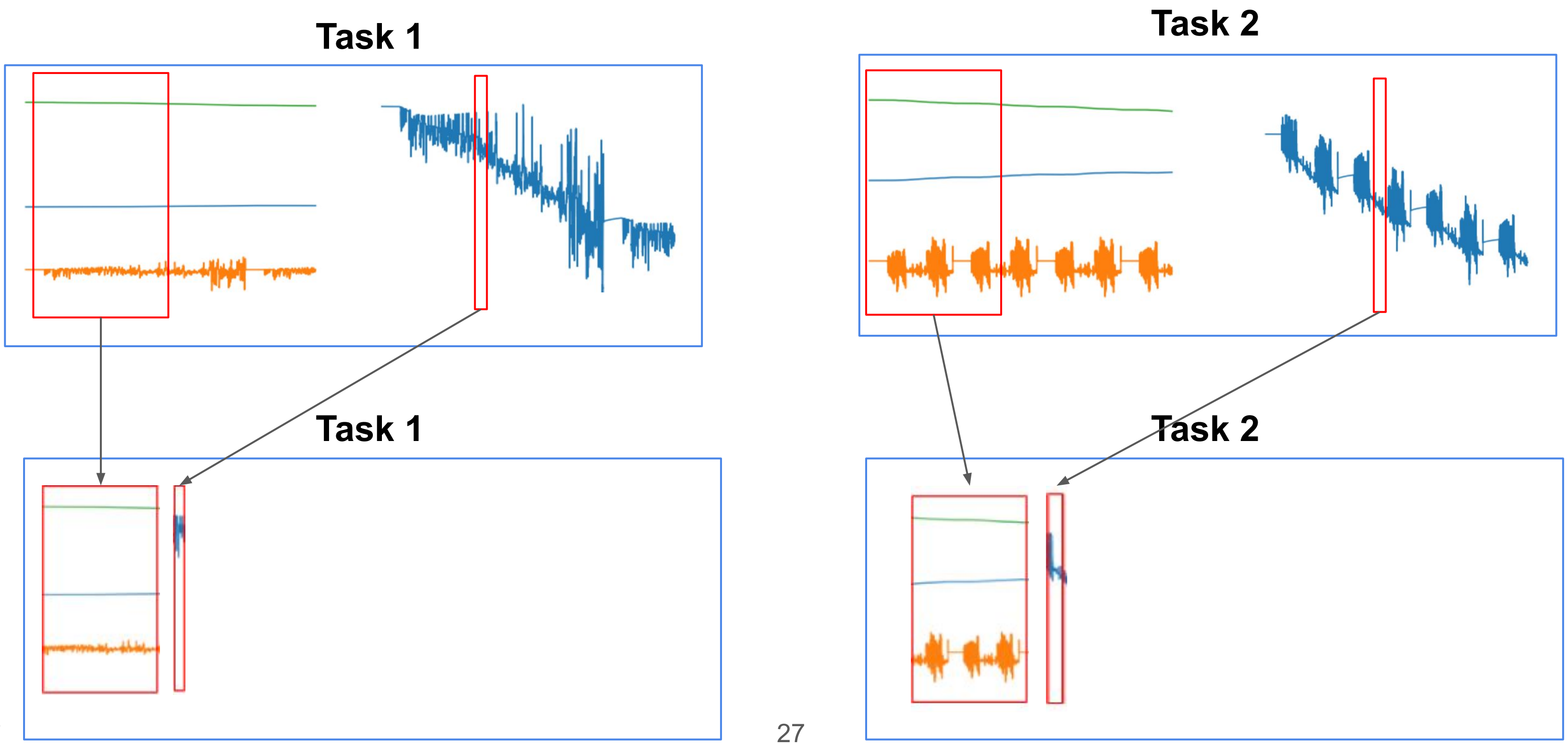
[5] Dempster, A., Petitjean, F. & Webb, G.I. *ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels.* Data Min Knowl Disc (2020)

[6] Roberts S. et al. *Gaussian processes for time series Modelling*. Phil. Trans. R. Soc. (2013)

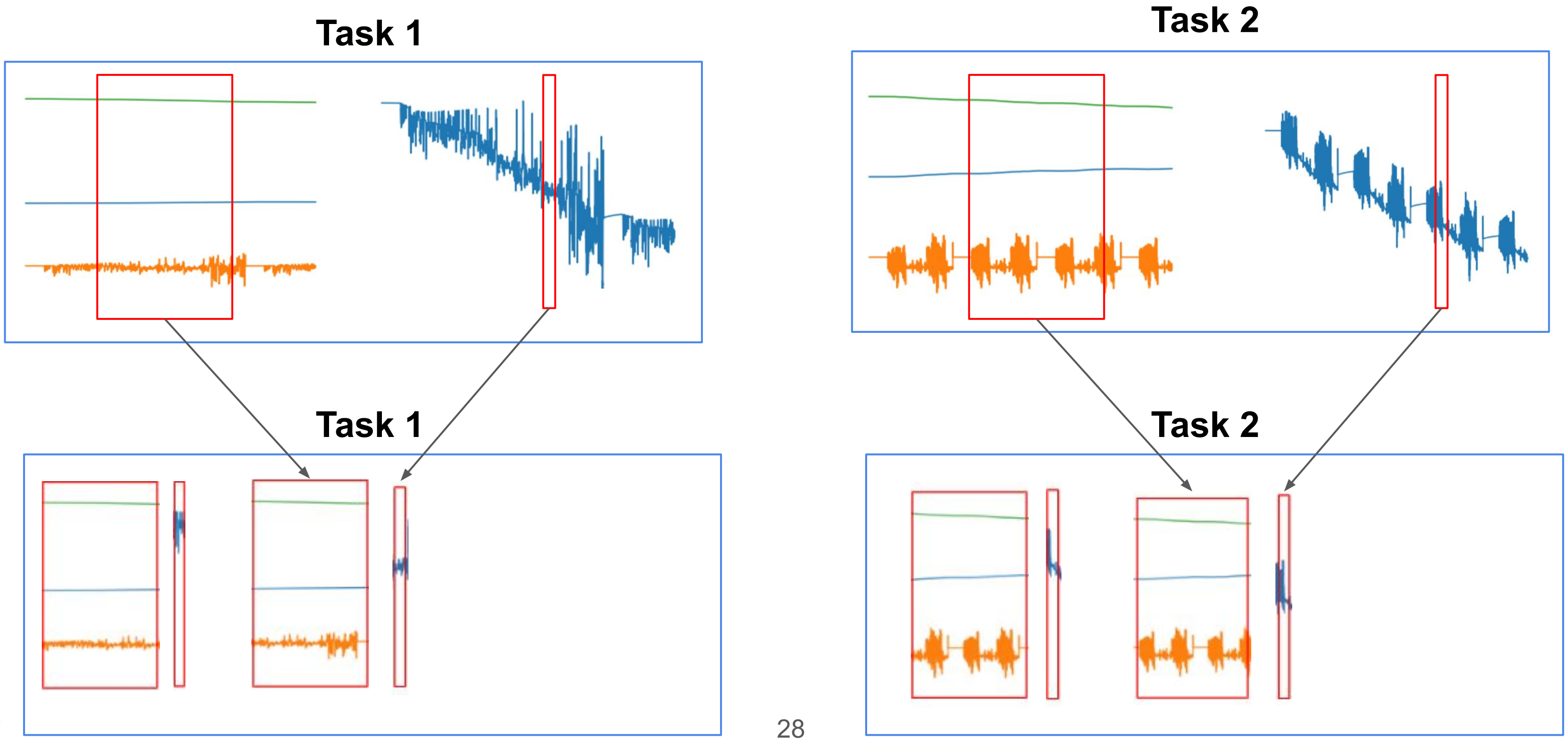
Thank you

Appendix

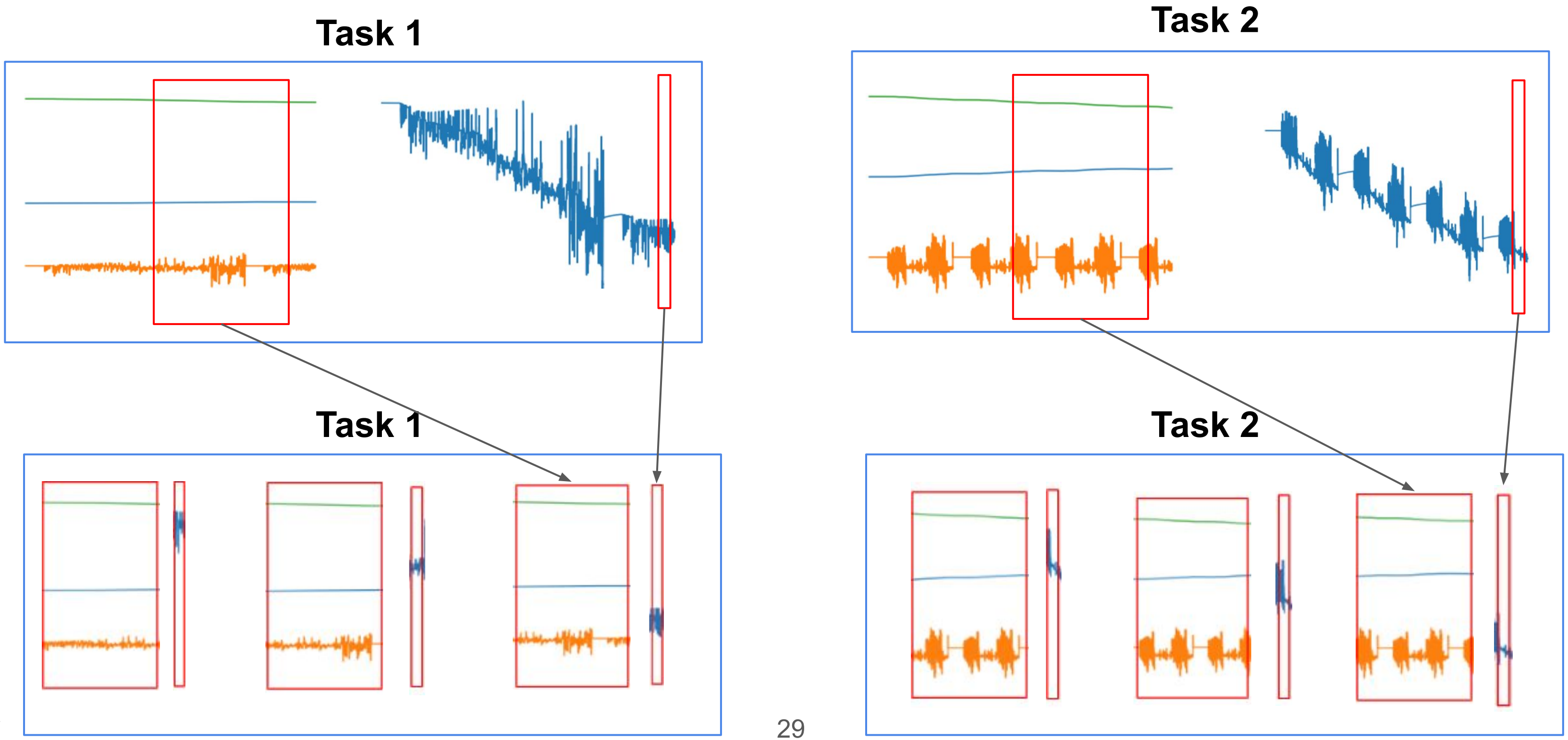
Time Series Regression (TSR) Task



Time Series Regression (TSR) Task

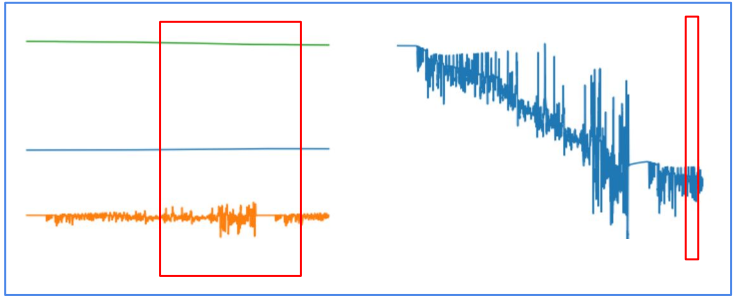


Time Series Regression (TSR) Task

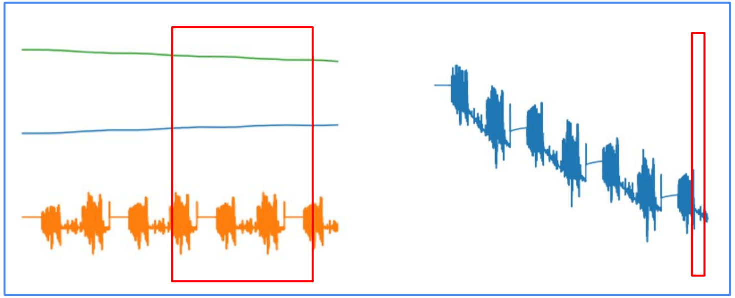


Time Series Regression (TSR) Task

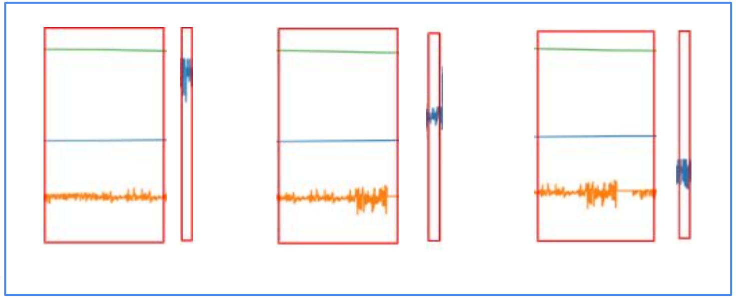
Task 1



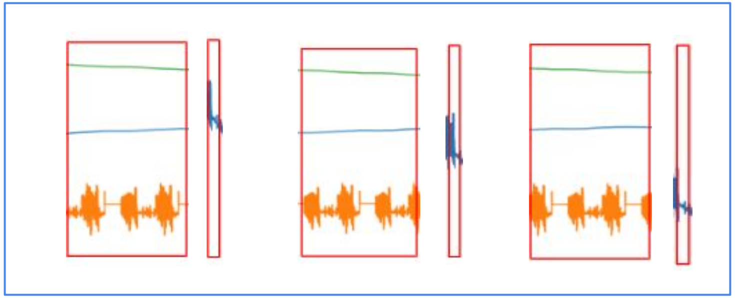
Task 2



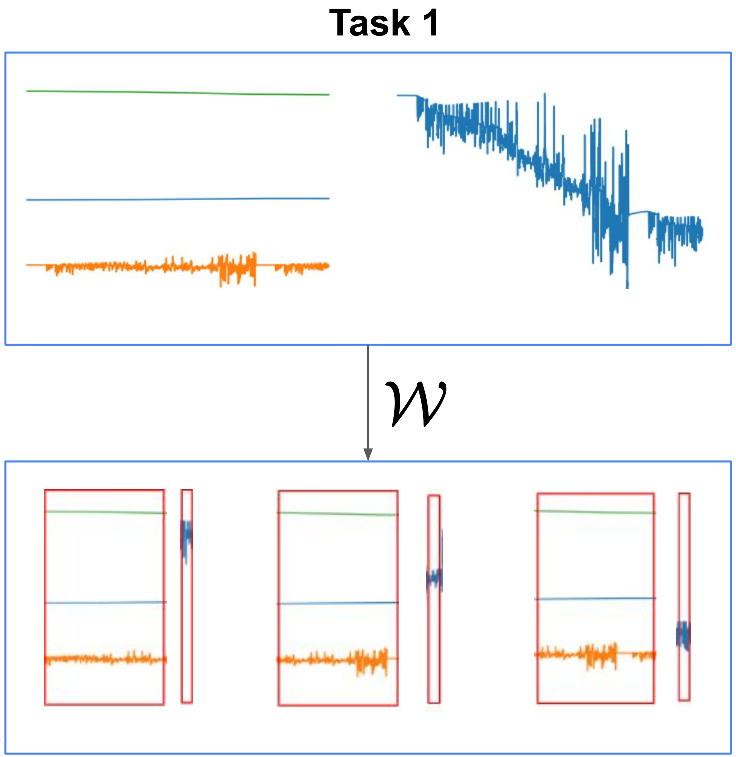
Task 1

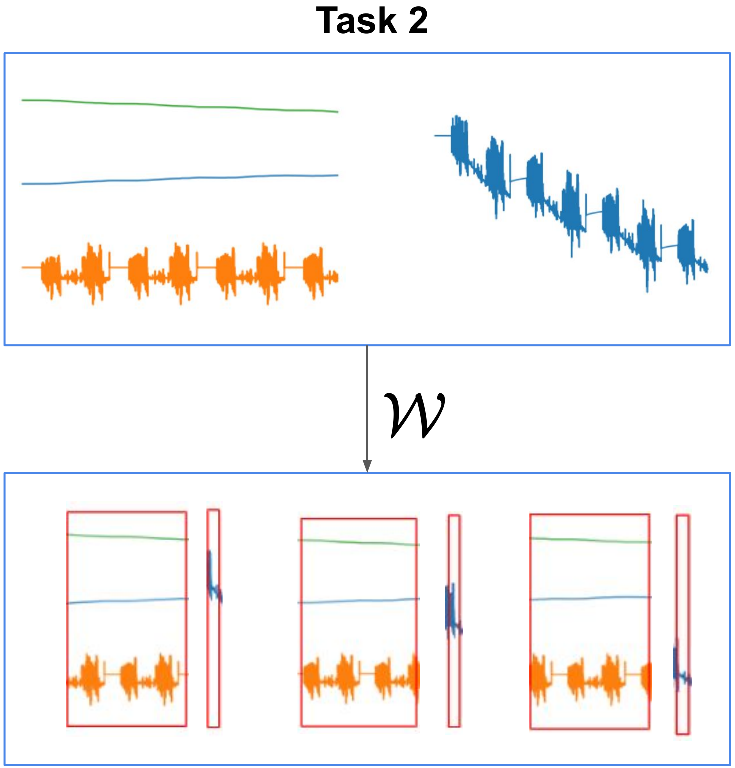


Task 2



Time Series Regression (TSR) Task





MAML for TSR: Concept

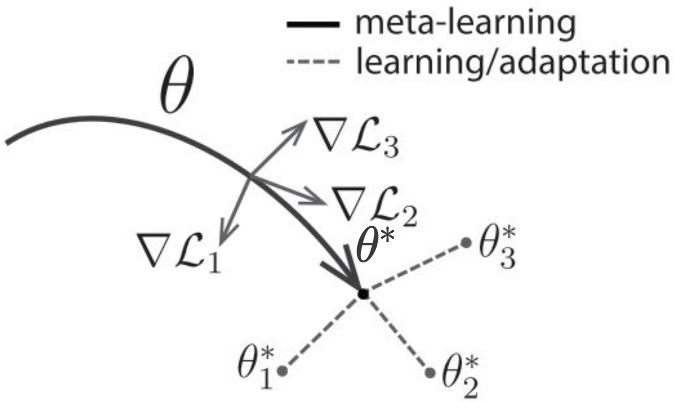


Image modified from the original paper

Initial parameters

Adaptation update

$$\theta_3^* = \theta^* - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_3}(f_\theta)$$

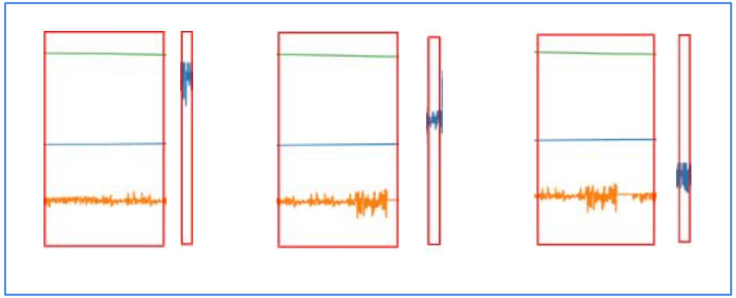
$$\min_\theta \sum_{\mathcal{T}_i \sim \mathcal{S}^{Tr}} \mathcal{L}_{\mathcal{T}_i}\left(f_{\theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)}\right)$$

$$\min_\theta \sum_{(\mathbf{S}_i, Y_i) \sim \mathcal{S}^{Tr}} \sum_{(\mathbf{x}_j, \mathbf{y}_j) \sim \mathcal{W}(\mathbf{S}_i, Y_i)} ||\mathbf{y}_j - f_{\theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)}(\mathbf{x}_j)||_2^2$$

Learning a model for TSR Task

$$\theta^* = \min_\theta \sum_{(\mathbf{x}_j, \mathbf{y}_j) \in W_i} ||\mathbf{y}_j - f_\theta(\mathbf{x}_j)||_2^2$$

Task 1



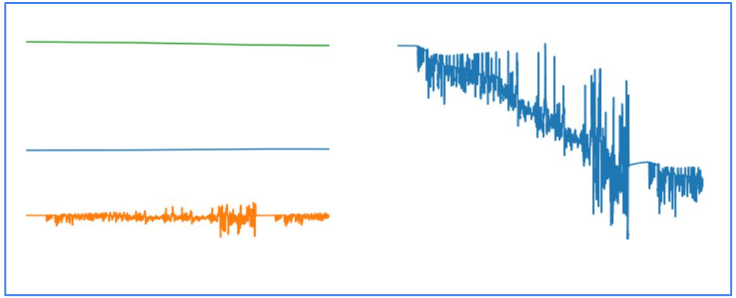
$\mathcal{W}$

Learning a model for several TSR Task

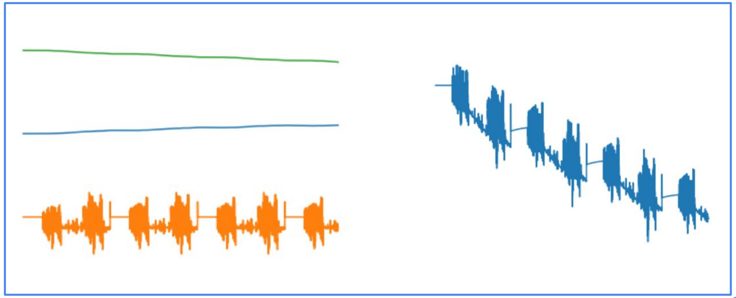
$$\theta^* = \min_\theta \sum_{\mathcal{T}_i \sim \mathcal{S}^{Tr}} \mathcal{L}_{\mathcal{T}_i}(f_\theta) \qquad \mathcal{T}_i = (\mathbf{S}_i, Y_i)$$

$$\theta^\star = \min_\theta \sum_{(\mathbf{S}_i, Y_i) \sim \mathcal{S}^{Tr}} \sum_{(\mathbf{x}_j, \mathbf{y}_j) \sim \mathcal{W}(\mathbf{S}_i, Y_i)} ||\mathbf{y}_j - f_\theta(\mathbf{x}_j)||_2^2$$

Task 1

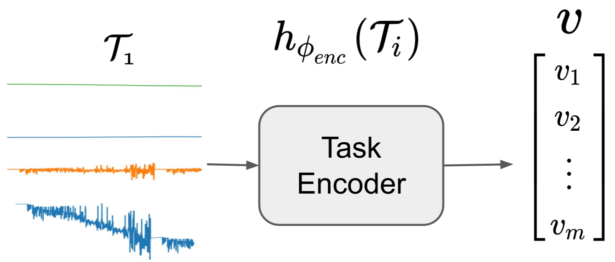


Task 2

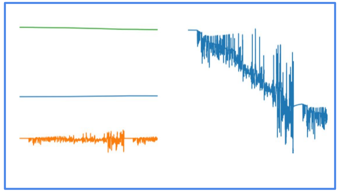


TSR Task Embedding

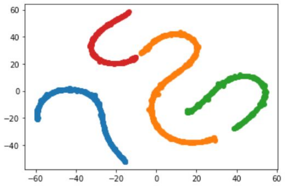
- RVAE leverages a Seq2Seq architecture to find the embeddings of time series.
- The encoder of RVAE can be used as task encoder for TSR.

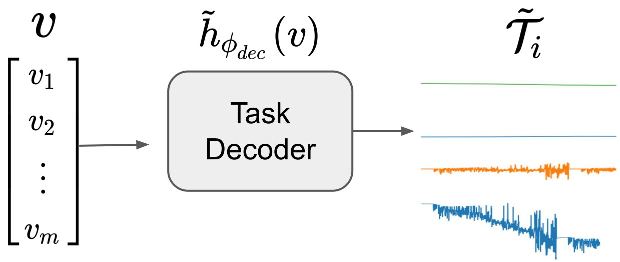


Task 1



T-SNE Plot of the embeddings





MMAML for TSR: Concept

- Task: $\mathcal{T}_i$
- Task embedding: $v_i = h_{\phi^{enc}}(\mathcal{T}_i)$
- Generated parameters: $\tau_i = g_{\phi^{gen}}(v_i)$
- Reconstructed task: $\tilde{\mathcal{T}}_i = \tilde{h}_{\phi^{dec}}(v_i)$
- The generated parameters modify the task networks through FiLM layers (...)
- Minimization objective:

$$\min_{\theta,\phi_{enc},\phi_{dec},\phi_{gen}} \sum_{\mathcal{T}_i \sim \mathcal{S}^{Tr}} \mathcal{L}_{s_i}\left(f_{\theta-\alpha\nabla_\theta \mathcal{L}_{s_i}\left(f_{\theta,\tau_i}\right),\tau_i}\right) + \lambda||\mathcal{T}_i - \tilde{\mathcal{T}}_i||_2^2$$

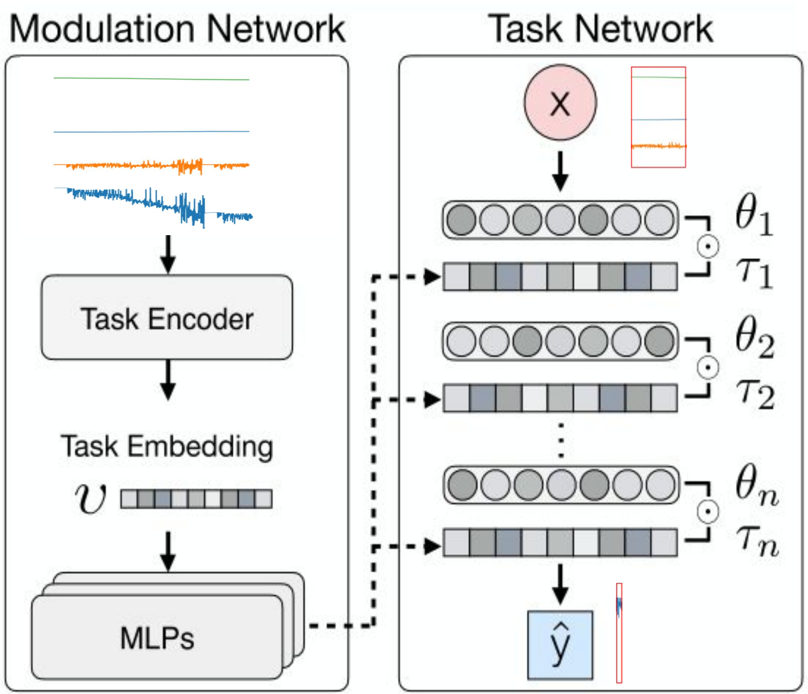


Image modified from the original paper

Ablation studies and other experiments

- Only task network
- Only MAML
- Task network + Data Augmentation
- Task network + Transfer learning
- Loss without the reconstruction loss