Surrogate Loss Learning for Dynamic Time Warping(DTW)

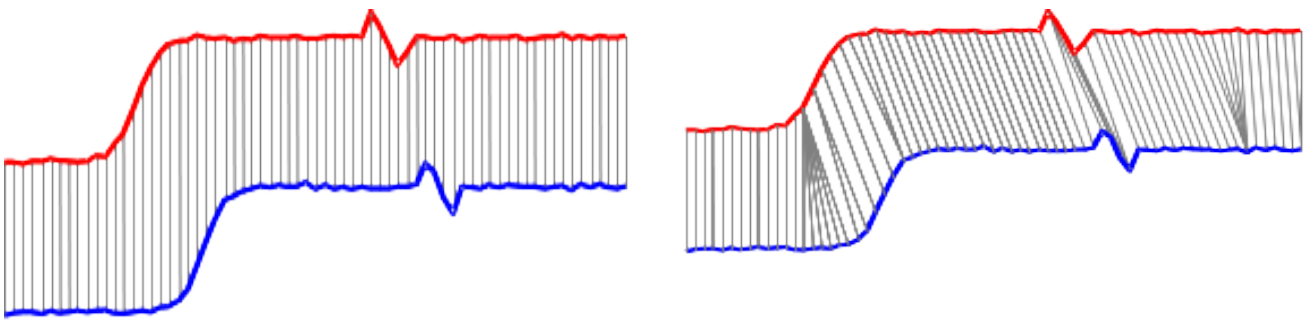
Supervisor: Prof. Dr. Dr. Lars Schmidt-Theime.
Co-Supervisor: Vijaya Yalavarthi

Tanvi Joglekar

Univesity of Hildesheim

Agenda

Motivation



(a) Euclidean Distance (b) Dynamic Time Warping

- **Dynamic Time Warping(DTW)** is a technique used to quantify the similarity between two time sequences varying in speed and length.
- DTW is used as an evaluation metric in time series forecasting.
- Can be applied to irregular time series.

Motivation

DTW Method:

- Calculate for each point in the series an easy distance (e.g. Euclidean).
- Iterate back through the series and add also the distance to the nearest a adjacent point.

For any two time-series,

$$X = (x_1, x_2, ..., x_N)$$

and

$$Y = (y_1, y_2, ..., y_M)$$

we compute a distance(cost) matrix as:

$$c_{i,j} = d(x_i, y_i) + min\{c(i-1, j-1), c(i, j-1), c(i-1, j)\}$$

Motivation

Research has manifested that we can use a metric as a loss function.

Problem: *DTW is non-differentiable because of* ***'min'*** *operation.*
Solution: *Surrogate Losses are a superior option.*

Goal: Develop a surrogate model which depicts DTW loss function as a meta-level neural network.

- Surrogate Loss network will be trained before forecasting model.

Literature Review

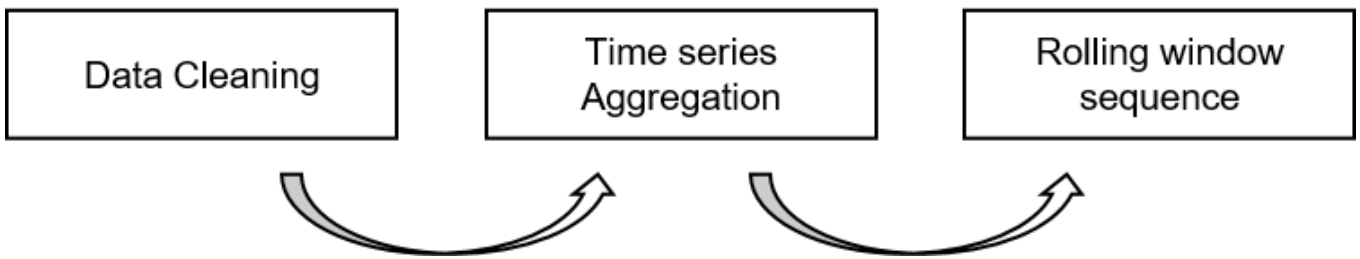
In time series forecasting following loss functions are used:

- MAE:
- MAPE
- MSE
- soft-DTW *

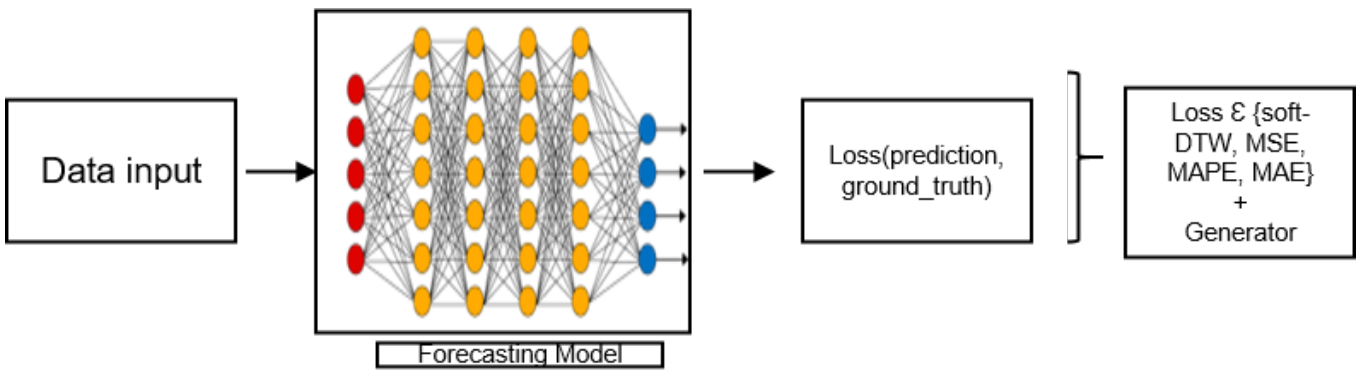
*soft-DTW: a differentiable loss function version of DTW which is highly depedent on hyperparameter λ.
Given as: $dtw_\gamma(x, y) := min_\gamma\{\langle A, \Delta(x, y)\rangle, A \epsilon A_{n,m}\}$.

Research Idea

Pre-processing

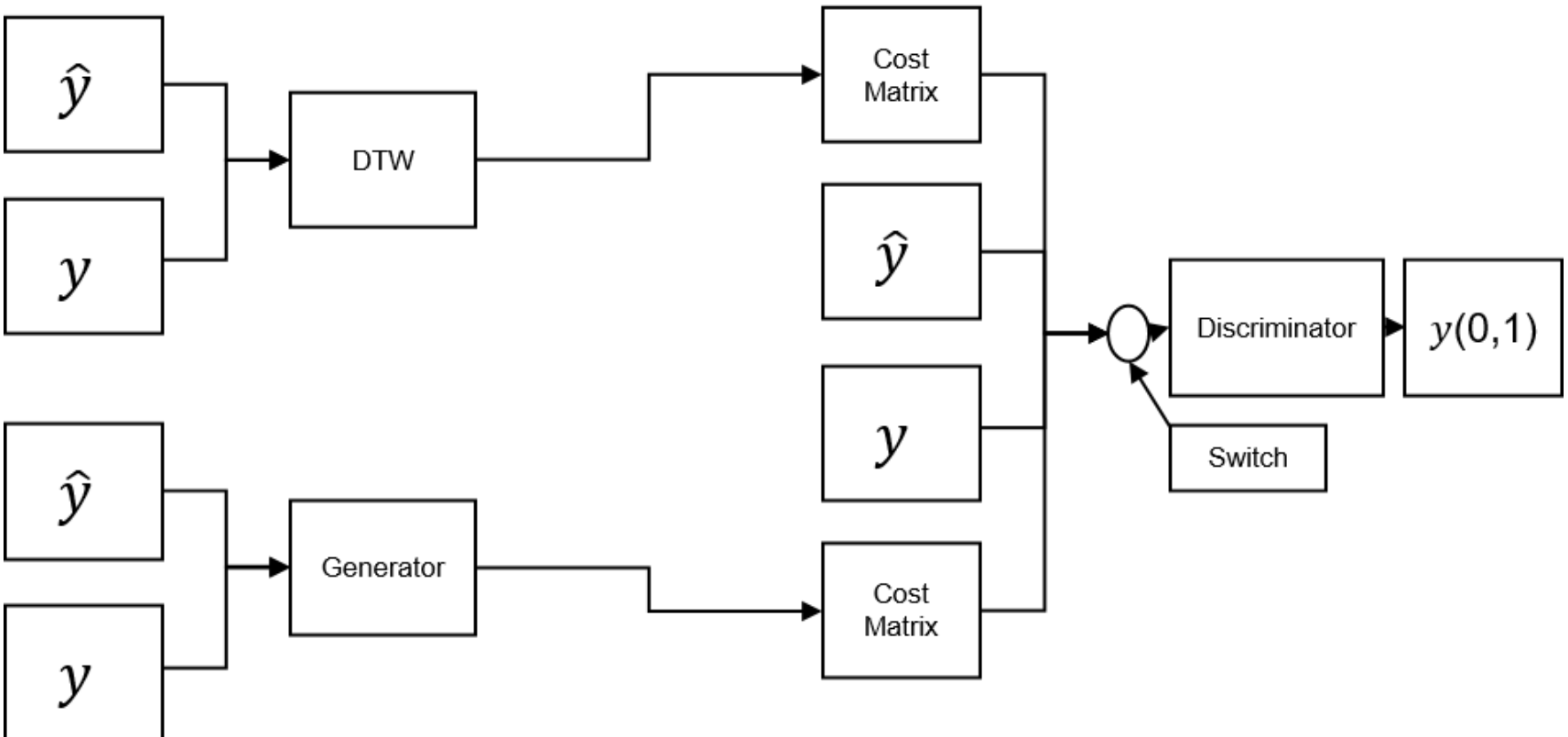


Baseline Idea



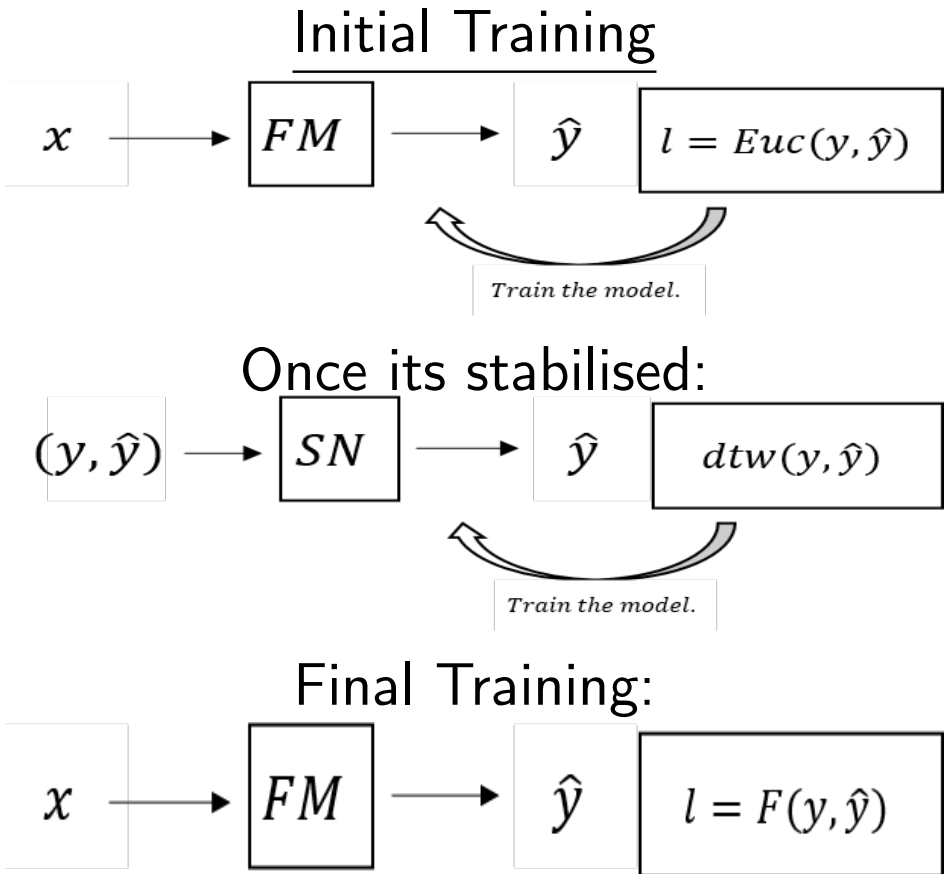
Research Idea

Proposed Model for Surrogate Loss Network- GAN:



Research Idea

Meta-data for training Surrogate Loss Network:
Forecasting Model → FM, Surrogate Function → SF,Input → X, Prediction → Y



Data Foundation

Public Data-sets:

- Traffic
- Electricity
- Retail

Private/ Company Data-set:
RBB Data-streams

Data Foundation

Dataset	Description	Prediction horizon	Properties	Pre-processing
Electricity Dataset	UCI Electricity Load Diagrams Dataset, containing electricity consumption of 370 customers	Use past week's data to forecast over the next 24 hours.	Target Type: R No. Of columns: 370 No. Of samples: 500k	Aggregate on hourly level.
RBB Data-streams	Time-related sensored data from production machinery.	Use past year's data to forecast over the next week.	Target Type: R No. Of columns: 165 No. Of samples: 17k	Aggregate on daily level.
Traffic Dataset	UCI PEM-SF Traffic Dataset which describes occupancy rate of 440 San Francisco Bay Area Freeways.	Use past week's data to forecast over the next 24 hours.	Target Type: $[0, 1]$ No. Of columns: 370 No. Of samples: 500k	Aggregate on hourly level.
Retail Dataset	Favorita Grocery Sales Dataset from the Kaggle competition is a metadata for different products and the stores sampled at the daily level.	Use 90 days of past data to forecast product sales of 30 days into the future.	Target Type: R No. Of columns: 130k No. Of samples: 500k	Use daily data.

Evaluation

Performance Evaluation:

- We evaluate the performances of proposed model, and compare it against above mentioned baselines.
- The results are evaluated using following metrics:
  1. DTW
  2. MSE

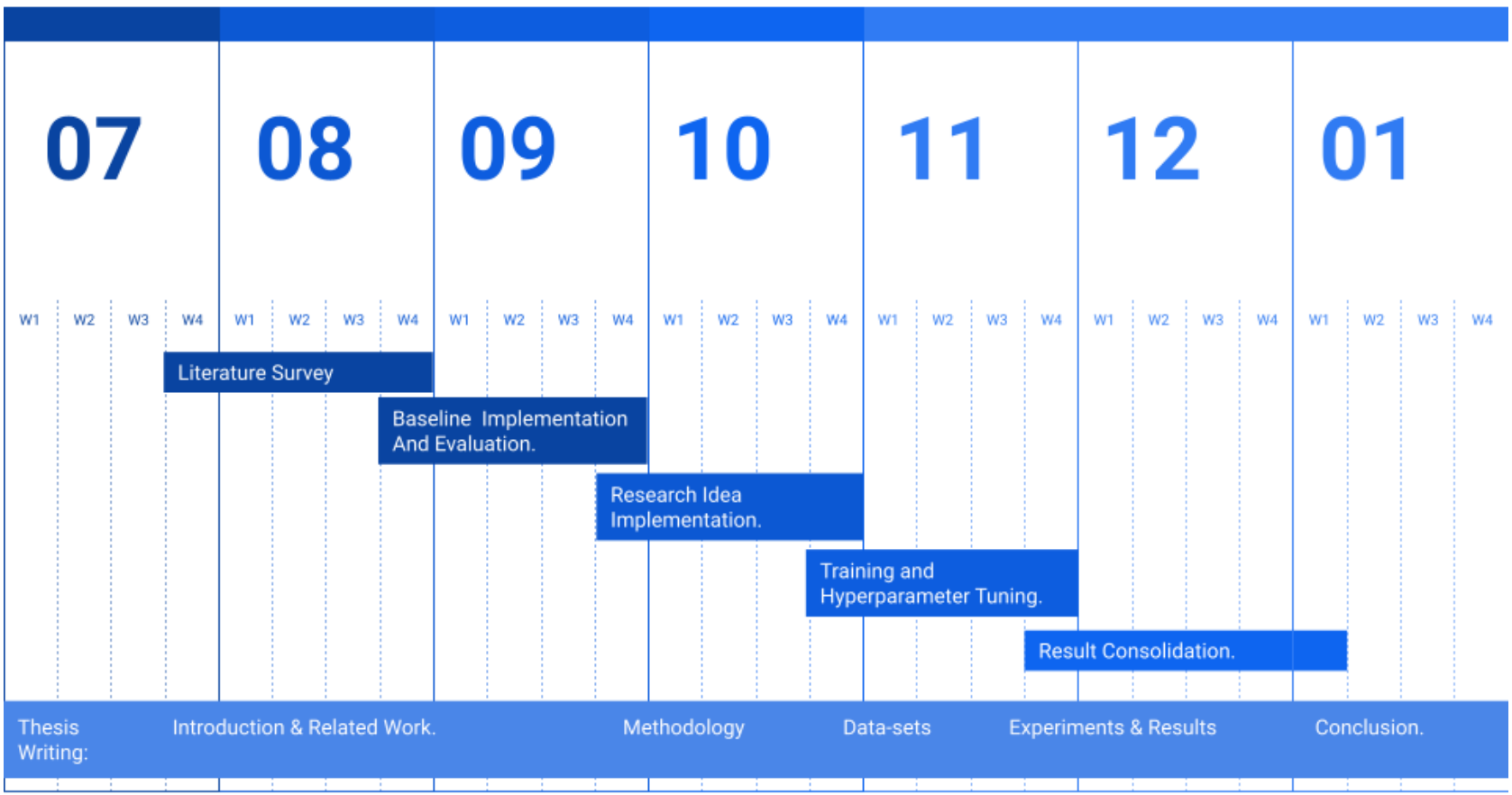
References

- Cuturi, Marco, and Mathieu BlondelCuturi, Marco, and Mathieu Blondel. Soft-DTW: a differentiable loss function for time-series, International Conference on Machine Learning. PMLR, 2017.
- Lima Bryan, Sercan O. Arık, Nicolas Loeffb, Tomas Pfister. Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting, https://arxiv.org/pdf/1912.09363.pdf.
- Grabocka, Josif, Randolf Scholz, and Lars Schmidt-Thieme. Learning surrogate losses, arXiv:1905.10108 (2019).
- Lamb Alex et al, Professor Forcing: A New Algorithm for Training Recurrent Networks 2016(https://arxiv.org/abs/1610.09038)
- Electricity Data-set: https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014

References

- Traffic Data-set: http://pems.dot.ca.gov/
- Favorita Data-set: https://www.kaggle.com/c/favorita-grocery-sales-forecasting/data
- RBB-Dataset: MunichRE

Timeline


07
08
09
10
11
12
01
W1 W2 W3 W4 W1 W2 W3 W4 W1 W2 W3 W4 W1 W2 W3 W4 W1 W2 W3 W4 W1 W2 W3 W4 W1 W2 W3 W4
Literature Survey
Baseline Implementation And Evaluation.
Research Idea Implementation.
Training and Hyperparameter Tuning.
Result Consolidation.
Thesis Writing:
Introduction & Related Work.
Methodology
Data-sets
Experiments & Results
Conclusion.