Deep Representation Learning for Cluster Level Time Series Forecasting

8th International Conference on Time Series and Forecasting (ITISE 2022)

Authors:

Tsegamlak T. Debella, Bethelhem S. Shawel, Maxime Devanne

Jonathan Weber, Dereje H. Woldegebreal, Sofie Pollin , Germain Forestier

Presentation Outline

Introduction

- In today's data-driven world, time series forecasting is one of the intensively investigated temporal data mining technique.
- In practice, time series are defined by taking a set of ordered observations where the ordering is based on one of the common independent variables.
- The ordering of individual observations in turn introduces shapes (patterns) that are key inputs to temporal data mining techniques.

Introduction

- Time series defined by ordering angular distance measurements.

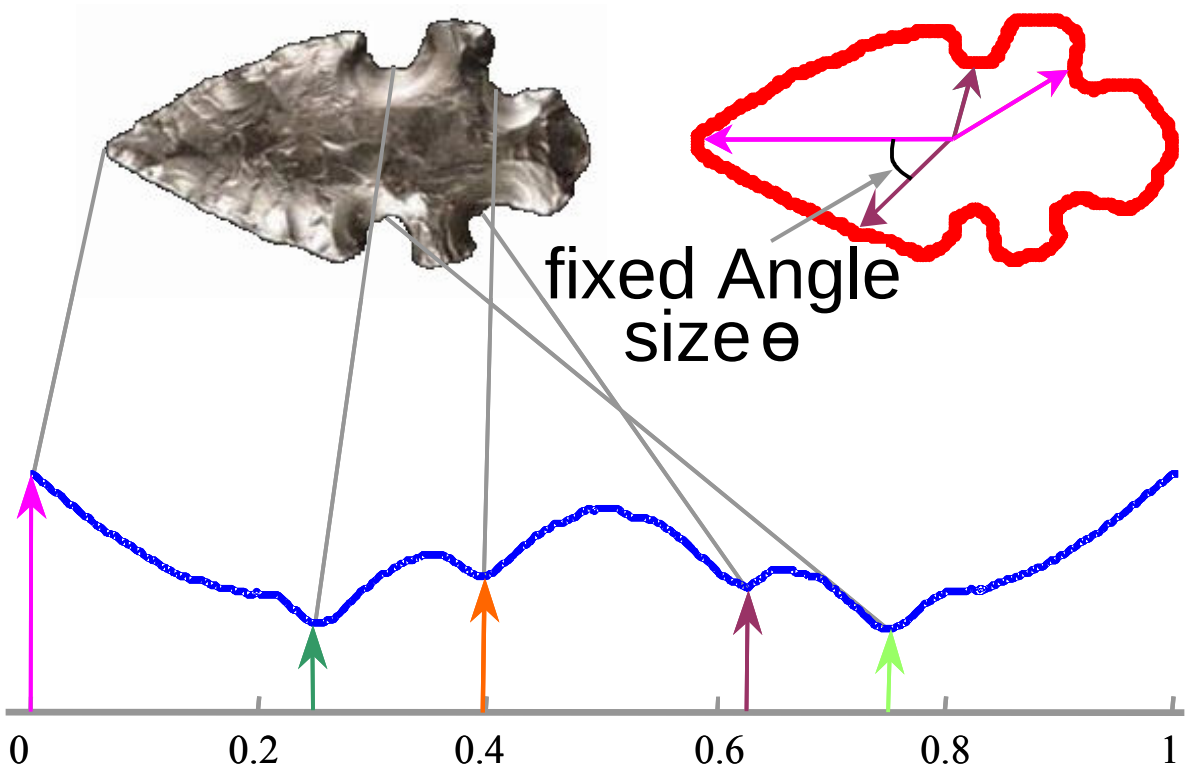


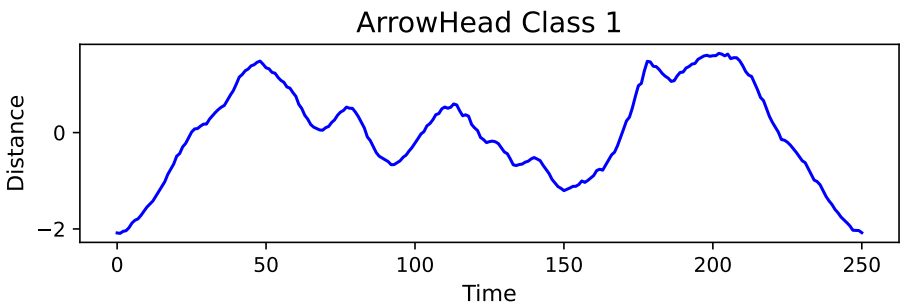
Figure 1: Extraction of time series from an image of an arrowhead [4]

Introduction

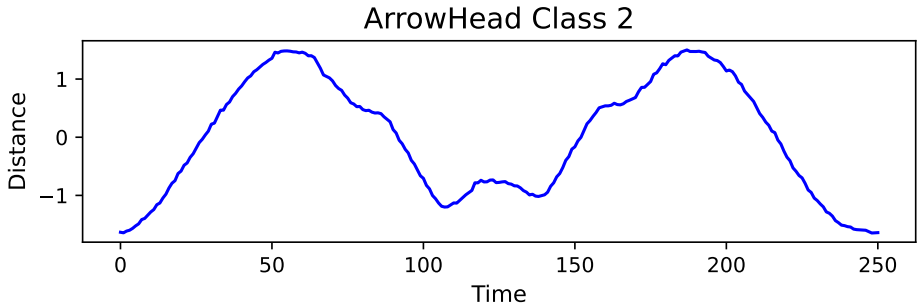
- Shapes, features, or patterns could be used in the context of different objectives

(a) *Avonlea* Arrowhead

(b) *Clovis* Arrowhead



(c) Time series for *Avonlea* Arrowhead



(d) Time series for *Clovis* Arrowhead

Figure 1: Shape based categorization of stone Arrowheads [2]

Introduction

- In this aspect, time series forecasting aims to optimize using (1).

$$\hat{X}_{i,t:t+p} = f(X_{i,t-L:t-1}, Y_{i,t-L:t-1}) \tag{1}$$

- Consequently, given a forecasting model f(.):
  - $\hat{X}_{i,t:t+p}$ implies predictions.
  - X ($X_{i,t-L:t-1}$ implies L previous observations.
  - $Y_{i,t-L:t-1}$ implies exogenous variables

Statement of the problem

- How do clustering and forecasting correlate?
  - Identification of exogenous variables.
  - Summarizing patterns within groups.
  - Reducing the number of required models.

.

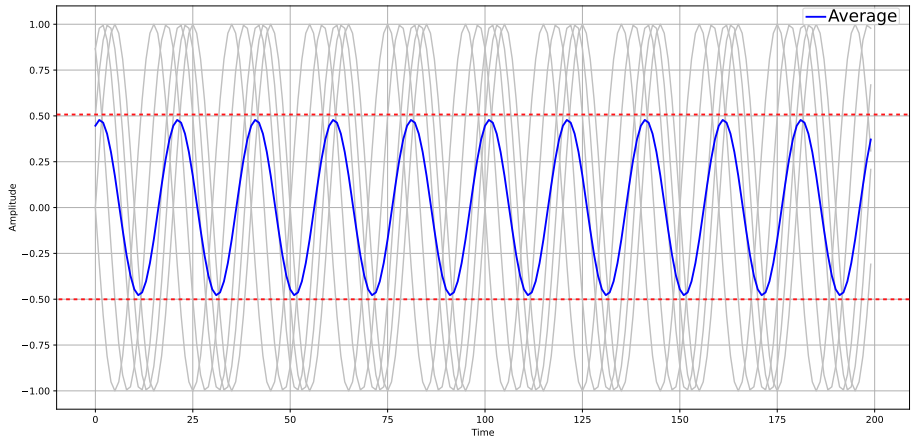
- For instance, lets consider temporal datasets representing [1, 5, 6].
  - Radio nodes are not isolated entities.
  - Hundreds if not thousands of radio stations.

Statement of the problem

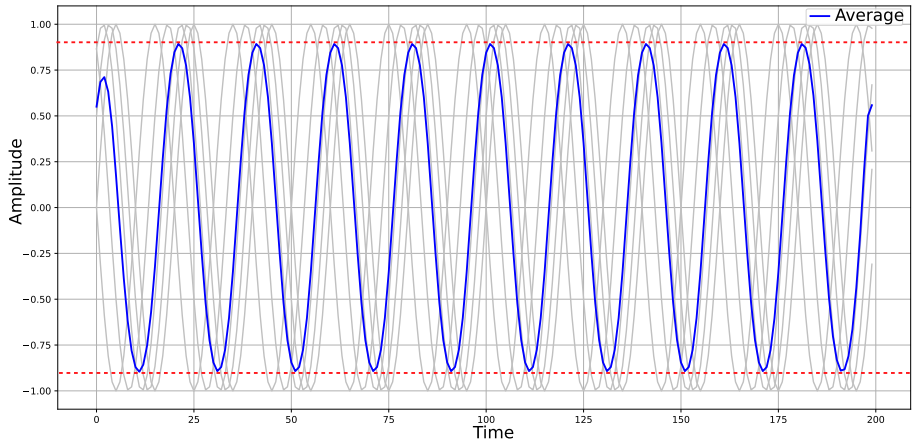
- An improper clustering approach could affect the quality of cluster level forecasts.

Example 2.1

- Let's consider sinusoids at phase shifts of $[0, \pi/3, \pi/6, \frac{2\pi}{3}, \textit{ and } \pi]$.



(a) An arithmetic mean



(b) SDBA mean [3]

Figure 2: Impact of temporal distortion on the estimation of cluster centorids

Proposed Approach

- In this work, we mainly aimed to show the potentials of deep representation learning in cluster formation and cluster centroid estimation.
- Consequently, we propose to:
  - Cluster datasets collected from a wireless network in the latent space of a Convoluational neural network.

Proposed Approach

- In this work, we mainly aimed to show the potentials of deep representation learning in cluster formation and cluster centroid estimation.
- Consequently, we propose to:
  - Cluster datasets collected from a wireless network in the latent space of a Convoluational neural network.
  - Estimate time domain cluster centroids using a multi-tasking Autoencoder.

Proposed Approach

- In this work, we mainly aimed to show the potentials of deep representation learning in cluster formation and cluster centroid estimation.
- Consequently, we propose to:
  - Cluster datasets collected from a wireless network in the latent space of a Convoluational neural network.
  - Estimate time domain cluster centroids using a multi-tasking Autoencoder.
  - Incorporate spatial information into a forecasting model.

Proposed Approach

- Datasets?
  - Data traffic load offered to 729 base stations located within Addis Ababa, Ethiopia.
  - Four months of measurement, i.e., September 2019 - December 2019.
  - Measurement is in Tera Bytes.
- Forecasting model?
  - K-Seasonal AutoRegressive Integrated Moving Average (K-SARIMA).

$$(\phi_p(L)\Phi_P(L^{s_k}))(1-L)^d(1-L)^{D_{s_k}}(X_t-\mu) = (\theta_q(L))(\Theta_Q(L^{s_k}))\varepsilon_t \quad (2)$$

Proposed Approach

- For the deep clustering, we propose the network arrangement shown in Figure 6.

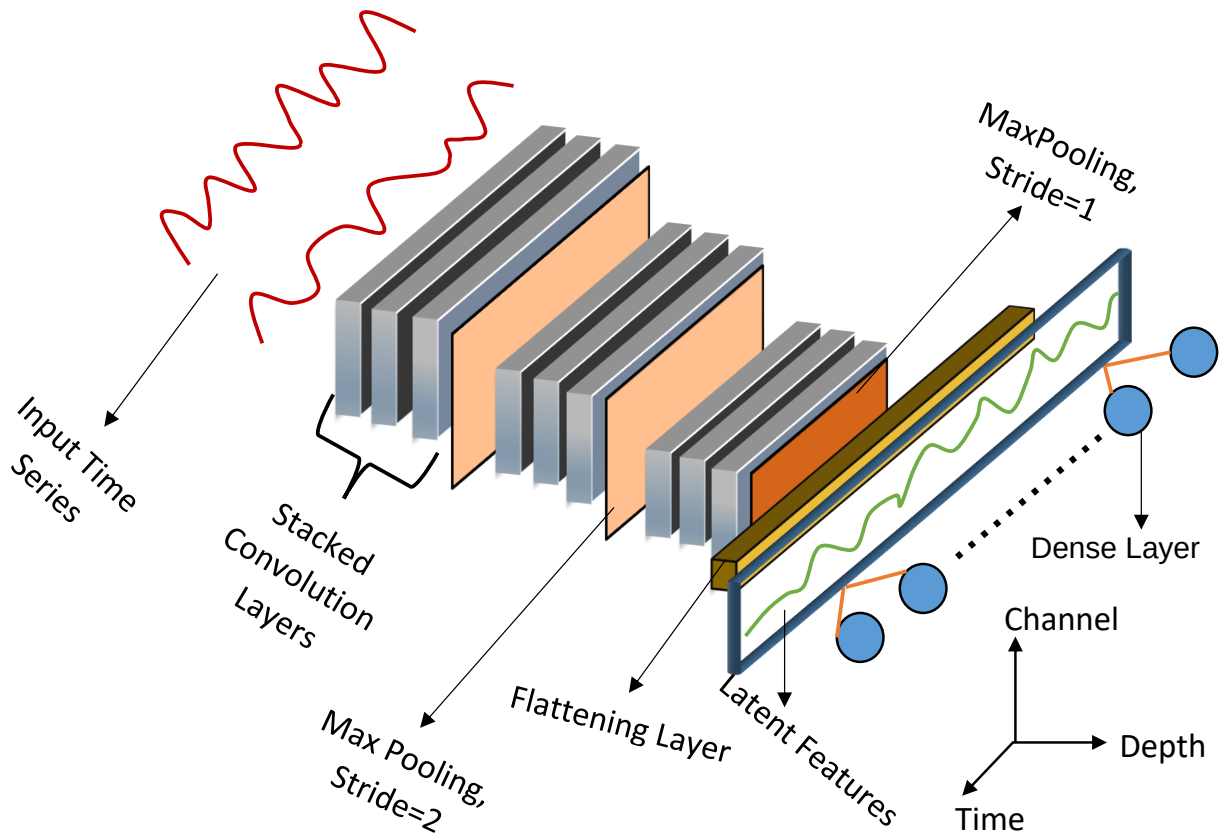


Figure 3: Proposed deep embedding clustering network

Proposed Approach

- The network was trained to optimize for the objective function given in (3) [8].

$$L = KL(P||Q) = \sum_i \sum_j p_{i,j} \log \frac{p_{i,j}}{q_{i,j}} \tag{3}$$

where,

$$p_{i,j} = \frac{\frac{q_{i,j}^2}{\sum_{j=1}^{K} q_{i,j}}}{\sum_{j=1}^{K} \frac{q_{i,j}^2}{\sum_{j=1}^{K} q_{i,j}}}, \qquad q_{i,j} = \frac{\left(\frac{1+||Z_i-\mu_j)||_{l2}}{\alpha}\right)^{\frac{-\alpha+1}{2}}}{\sum_{j=1}^{K}\left(\frac{1+||Z_i-\mu_j)||_{l2}}{\alpha}\right)^{\frac{-\alpha+1}{2}}}$$

- In order to define latent space cluster centroids ($\mu_j's$) a classic K-Means is used.

Proposed Approach

- The DEC arrangement can not generate time domain estimates. [7].

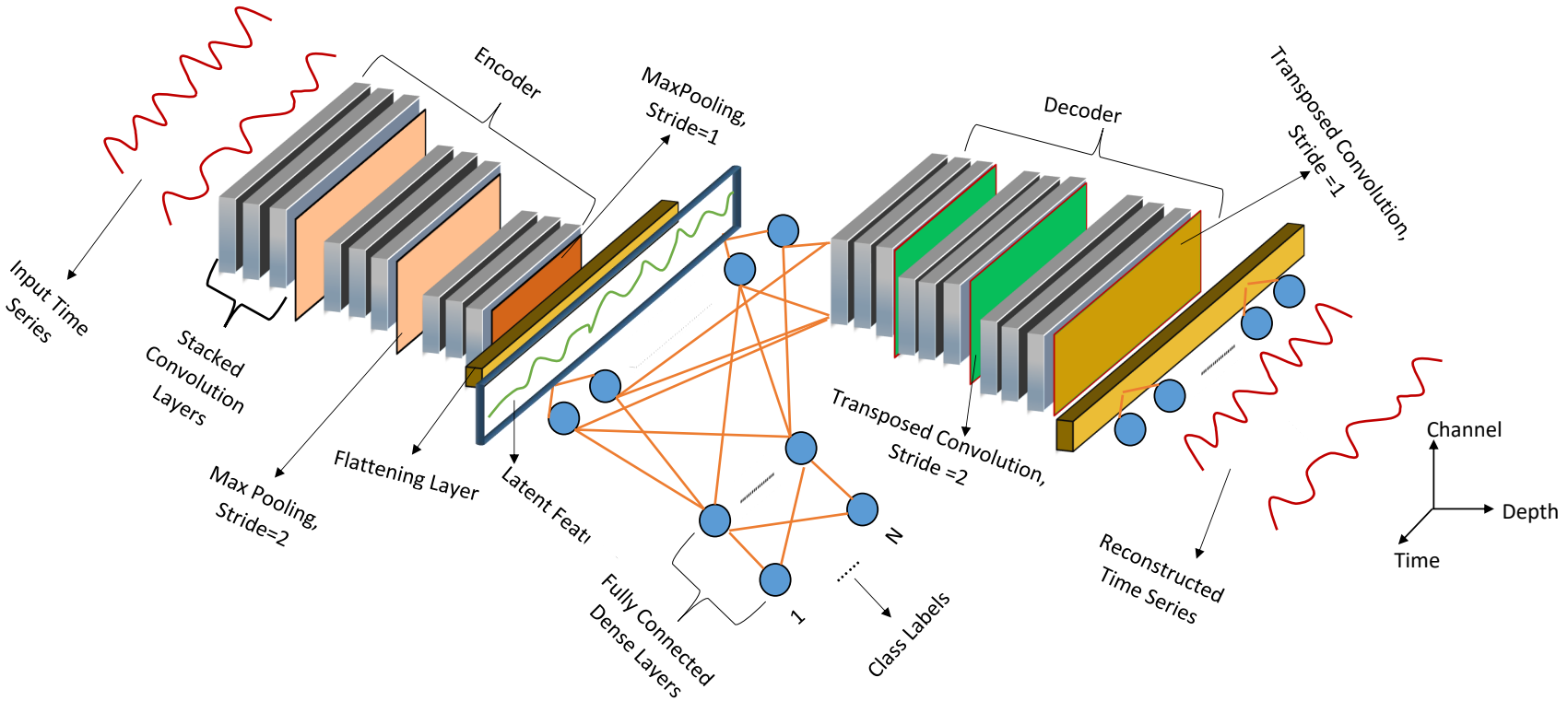


Figure 4: Proposed multi-tasking autoencoder

Proposed Approach

- The multi-tasking network was trained to optimize for the objective function given in (4) [8].
  - Latent means are estimated by taking an arithmetic mean.
  - We use the decoder to reproject latent means.

$$L_{Multi}(X, \hat{X}, Cat, p_{cat}) = \frac{1}{N}\sum_{i=1}^{N}||X_i - \hat{X}_i||_{l2} - \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} Cat_{i,j} \ln^{p_{i,j}} \quad (4)$$

Proposed Approach

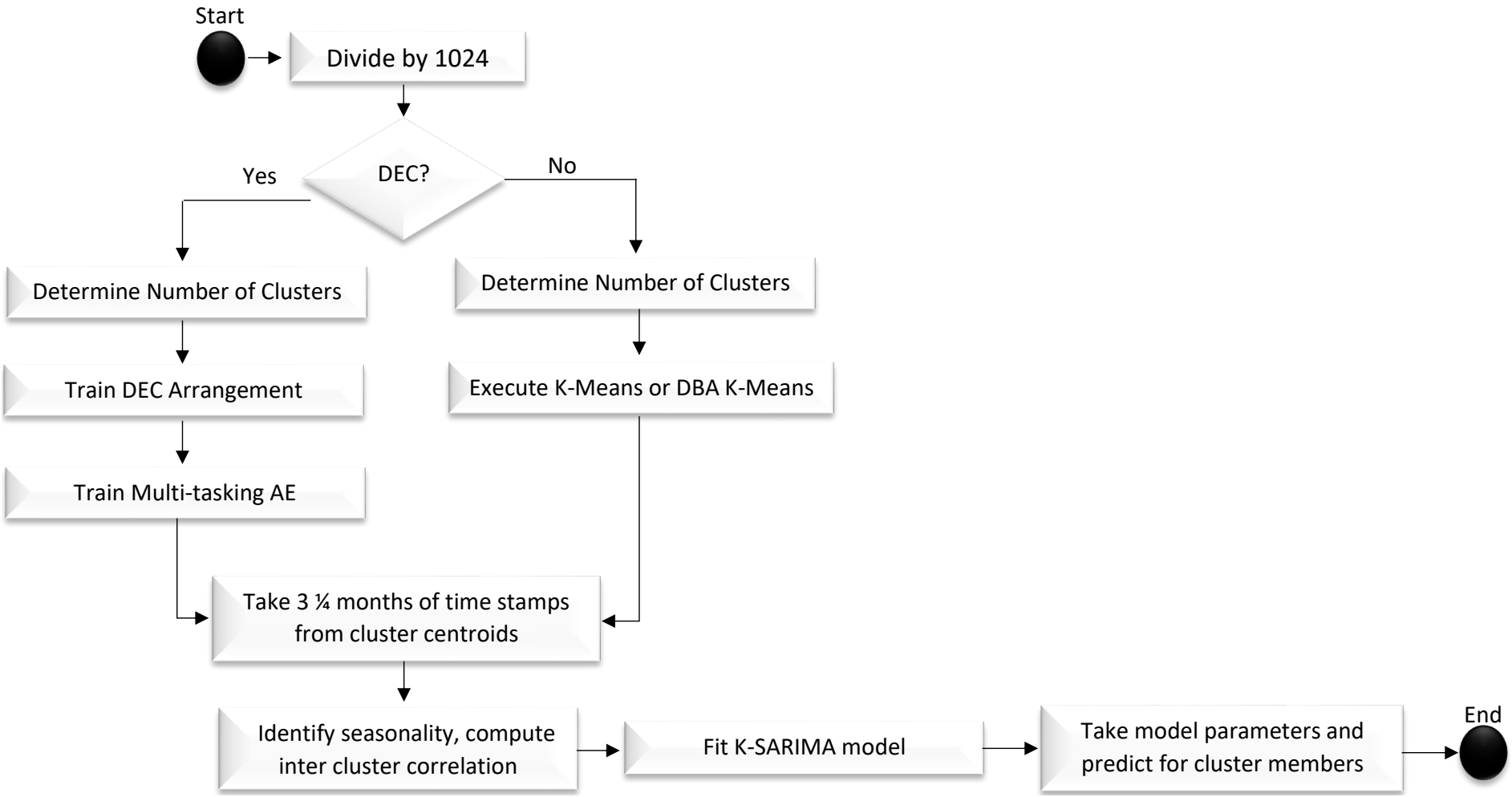


Figure 5: Experimental setup

Results and Discussion

- We started our analysis of the dataset by observing their intra cluster inertia, i.e., in order to determine the number of clusters.

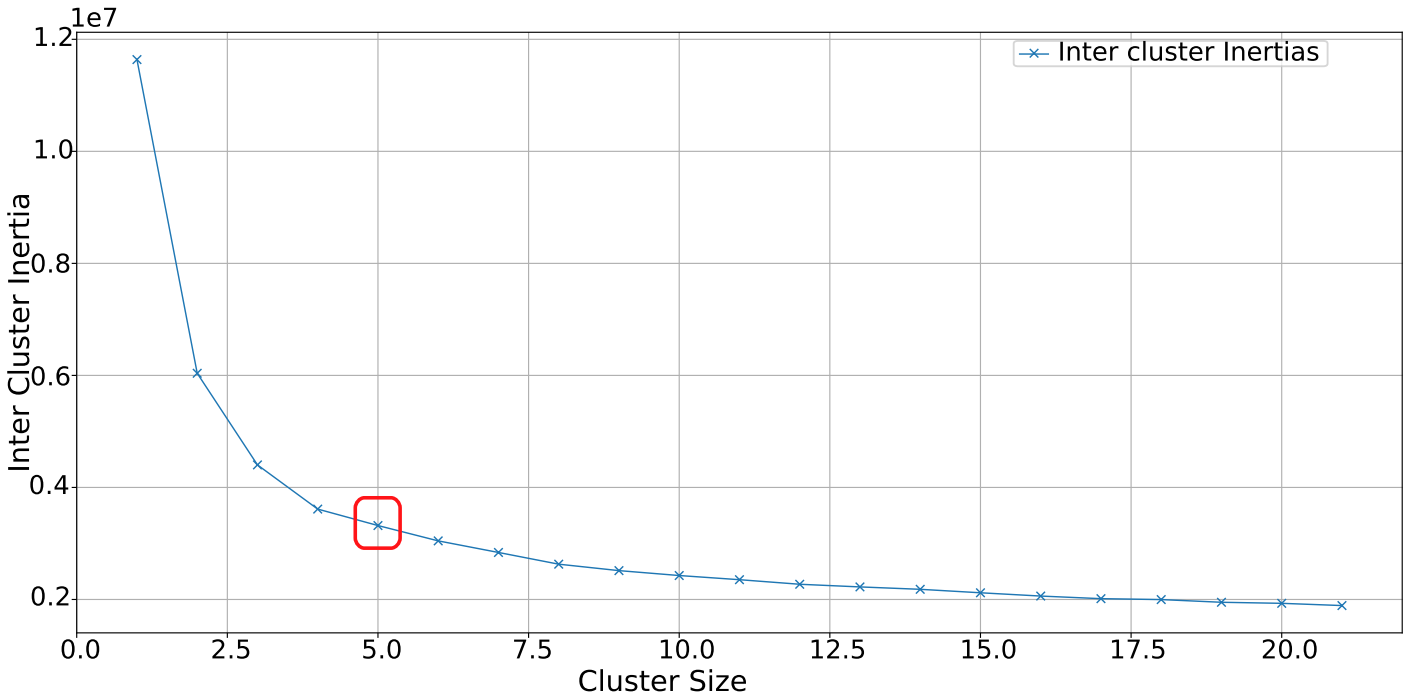
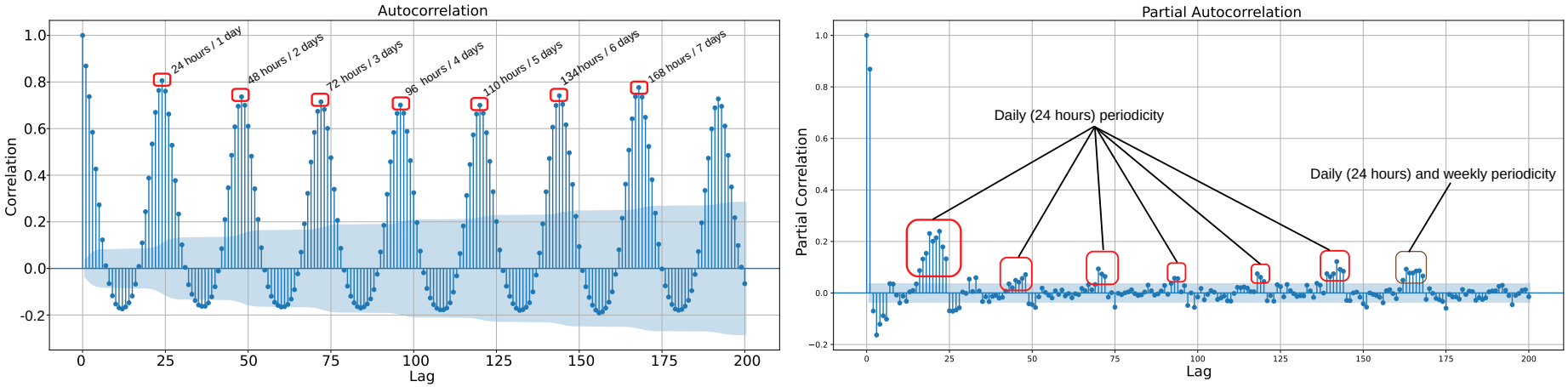


Figure 6: Intra cluster inertia

Results and Discussion

- In order to determine the order of the S-ARIMA model, i.e., (p,q,d) and (P,Q,D), we analyzed autocorrelation and partial autocorrelation of the cluster centroids.



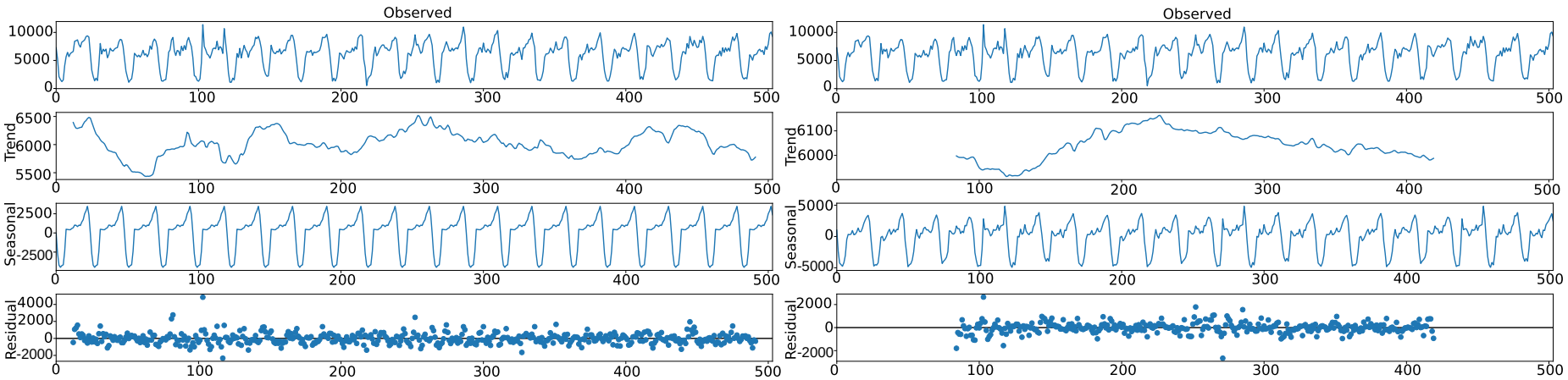
(a) Autocorrelation

(b) Partial autocorrelation

Figure 7: Seasonality analysis via autocorrelation and partial autocorrelation

Results and Discussion

- The daily and weakly seasonality is also evident in the decomposition of the cluster centorids and sample series taken from the entire dataset.



(a) Daily seasonality **(b)** Weekly seasonality

Figure 8: Decomposition of cluster centorids

- D-SARIMA $(2,\ 1,\ 2),\ (2,\ 0,\ 0)_{24},\ (2,\ 0,\ 0)_{168}$

Results and Discussion

Figure 9: Clusters identified by the DEC arrangement

Results and Discussion

Figure 9: Clusters identified by the DBA K-Means

Results and Discussion

- We used the parameters of the fitted D-SARIMA model to forecast for individual cluster members, i.e., for forecasts labeled with *CS*.

Table 1: Comparison of cluster level and data level forecasting

Techniques	Errors	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster4	Average
DEC_CS	RMSE	1.104	1.050	0.409	1.524	0.863	0.990
	MAE	0.816	0.778	0.305	1.145	0.671	0.743
DEC_Clust_BS		1.417	1.205	0.426	1.867	0.992	1.181
		1.092	0.932	0.308	1.158	0.758	0.909
DBA K-Means_CS	RMSE	0.714	2.129	1.079	0.416	1.468	1.161
	MAE	0.524	1.715	0.789	0.318	1.156	0.901
DBA_K-Means_Clust_BS		0.868	2.012	1.205	0.330	1.652	1.214
		0.664	0.565	0.919	0.239	1.288	0.935
K-Means_CS	RMSE	1.131	0.938	0.389	1.774	1.906	1.228
	MAE	0.859	0.739	0.293	1.452	1.522	0.973
K-Means_Clust_BS		1.263	0.961	0.388	1.664	1.993	1.254
		0.968	0.735	0.282	1.289	1.562	0.967

Results and Discussion

- DEC D-SARIMA, DBA K-Means, K-means (RMSE,MAE): (0.758, 0.596), (0.802, 0.617), (0.847, 0.669).

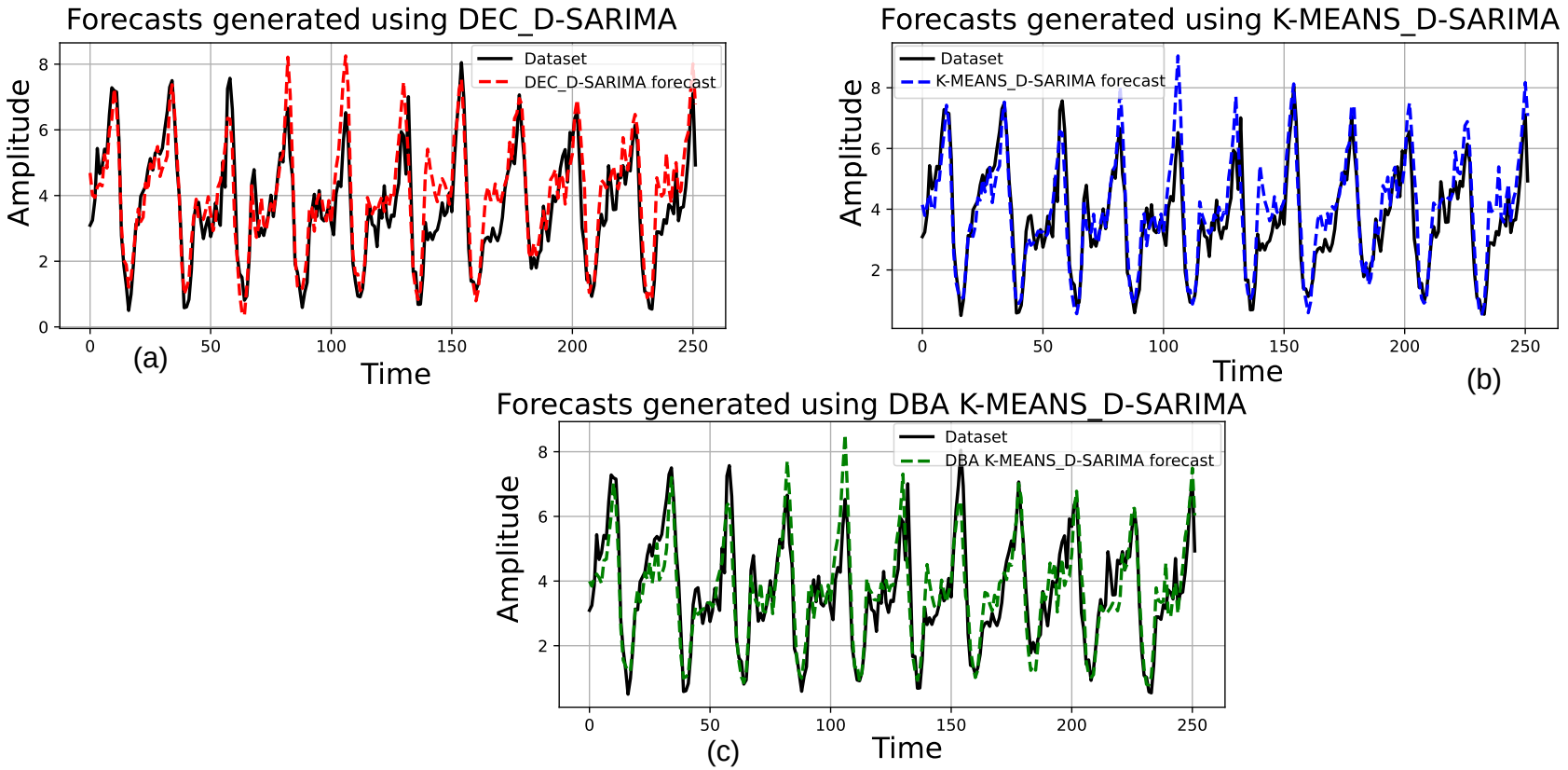


Figure 10: Sample cluster level forecasts

Conclusion

- In this paper, we argued that the effects of temporal distortion should be addressed in cluster level forecasting.
- In this regard, we showed the potential of of representation learning in improving the performances of readily available forecasting models.
- In reality, the potential is not limited to aiding linear models. In contrary, the reported results could further be improved with the integration of more resilient forecasting models.

Thank You!

List of References

[1] Guangshuo Chen. *Spatiotemporal Individual Mobile Data Traffic Prediction*. Technical Report RT-0497. INRIA Saclay - Ile-de-France, Feb. 2018. URL: https://hal.inria.fr/hal-01675573.

[2] Yanping Chen et al. *The UCR Time Series Classification Archive*. www.cs.ucr.edu/~eamonn/time_series_data/. July 2015.

[3] Marco Cuturi and Mathieu Blondel. "Soft-DTW: a Differentiable Loss Function for Time-Series". In: *Proceedings of the 34th International Conference on Machine Learning*. Sydney, NSW, Australia, July 2017, pp. 894–903.

[4] Ye Lexiang and Keogh Eamonn. "Time Series Shapelets: A New Primitive for Data Mining". In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2009, pp. 947–956.

List of References

[5] A. Sfetsos and C. Siriopoulos. "Time series forecasting with a hybrid clustering scheme and pattern recognition". In: *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 34.3 (2004), pp. 399–405.

[6] Bethelhem S. Shawel et al. "Hybrid Prediction Model for Mobile Data Traffic: A Cluster-level Approach". In: *2020 International Joint Conference on Neural Networks (IJCNN)*. 2020, pp. 1–8.

[7] Tsegamlak Terefe et al. "Time Series Averaging Using Multi-Tasking Autoencoder". In: *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)*. 2020, pp. 1065–1072.

List of References

[8] Junyuan Xie, Ross Girshick, and Ali Farhadi. "Unsupervised Deep Embedding for Clustering Analysis". In: *Proceedings of the 33rd International Conference on International Conference on Machine Learning*. Vol. 48. New York, NY, USA: JMLR.org, 2016, pp. 478–487.