

# Multi-perspective embedding for non-metric time series classification

Maximilian Münch<sup>1,2</sup>, Simon Heilig<sup>1</sup> and Frank-Michael Schleif<sup>1 \*</sup>

1- University of Applied Sciences Würzburg-Schweinfurt,  
Faculty of Computer Science and Business Information Systems  
Würzburg, Germany

2- Bernoulli Institute for Mathematics, Computer Science  
and Artificial Intelligence, University of Groningen,  
Groningen, The Netherlands

**Abstract.** The interest in time series analysis is rapidly increasing, providing new challenges for machine learning. Over many decades, Dynamic Time Warping (DTW) is referred to as the de facto standard distance measure for time series and the tool of choice when analyzing such data. Nevertheless, DTW has two major drawbacks: (a) it is non-metric and therefore hard to handle by standard machine learning techniques, and (b) it is not well suited for multi-dimensional time series. For this purpose, we propose a multi-perspective embedding of the time series into a complex-valued vector space and the evaluation by a model that is able to handle complex-valued data. The approach is evaluated on various multi-dimensional time series data and with different classifier techniques.

## 1 Introduction

Time series analysis traditionally occupies a special role in machine learning. The majority of algorithms critically depends on fixed-length real-valued vectorial input data and can not be applied to data with various lengths. One common approach today is to embed any data to a vectorial representation using embedding techniques of deep learning, requiring tons of training data without any guarantee that the essential input information is preserved [1]. Large data sets are not always available, and for different scenarios proximity-based measures are still very popular [2], including time series analysis [3]. For each pair of input data, i.e., time series, a comparison measure provides a proximity score characterizing the degree of similarity or dissimilarity between two inputs [4].

Considering an input set of  $N$  time series objects with varying lengths and an appropriate dissimilarity measure, we can represent the data by calculating all pairwise comparisons and obtain an  $N \times N$  dissimilarity matrix. The respective matrix can be directly used for a nearest-mean classifier [4]. In case of a similarity function, like the Euclidean inner product, an  $N \times N$  kernel matrix is obtained. If the dissimilarities are obtained by a metric measure or the similarities are

---

\*MM is supported by the Bavarian HighTech agenda and the Würzburg Center for Artificial Intelligence and Robotics (CAIRO). SH is supported by the ESF (WiT-HuB 4/2014-2020), project KI-trifft-KMU, StMBW-W-IX.4-6-190065.

positive semi-definite (psd), a variety of machine learning techniques like branch-and-bound approaches or dual formulations of classifier models, like the support vector machines [5], can be used.

It is desirable to use a measure which (a) can manage input data of varying length and (b) is meaningful by preserving the relevant information between input data. Over the last 50 years, *Dynamic Time Warping* (DTW) and derived measures have proven to be useful for time series and are actively used [3]. While DTW is still the distance of choice, it has two significant limitations in real-world scenarios: (a) data are often multi-dimensional to describe a problem from different perspectives, and (b) the measure itself is non-metric, which limits the direct use in various algorithms. So far, only a few techniques for the holistic study of multi-dimensional time series with DTW [6] have been suggested. For similarities, particular psd time series kernels have been proposed [7]. In particular non-metric measures like DTW can be shaped into non-psd similarities by means of a double centering technique [4]. In our setting, the term multi-perspective is used synonymously to characterize data, given in multiple dimensions as well as data which are provided in different representations, e.g. by using multiple kernel functions or dissimilarity measures. In this view, combining these multiple perspectives serves as a nonlinear feature transformation of the input data, but in a flexible but still interpretable formulation. In the context of multi-perspective representations, multiple kernel learning could be used, which, however, cannot be applied to non-psd kernels. To overcome these limitations, we propose a multivariate embedding of the various time series dimensions into several complex-valued vector spaces that are later combined to one space. On own previous work on complex-valued embeddings in [8] and using ideas from [2], this can be done at low costs with moderate approximations. The recombined complex-valued data can be processed by an effective classifier model for complex-valued data suggested in [9].

## 2 Background and Related Work

In this work, we focus on time series classification, which is challenging due to the non-fixed length representation of the individual time series objects. Additionally, time series data are frequently multi-dimensional or multi-channel data, such that many classical approaches can not directly be used.

Nowadays, for structured data, including time series, deep learning-based embedding techniques such as Word2Vec, Protein2Vec, etc., are highly competitive for learning a vectorial representation of the input data [10]. They perform reasonably well in certain domains but are computationally costly and require a tremendous amount of training data [1].

Often the available data is rather limited and additional it is desirable to obtain an interpretable predictor such that the contribution of input features or multiple perspectives can be easily accessed. Therefore, we do not further consider deep learning techniques but focus on the de facto standard of Dynamic Time Warping and related approaches as the methods of choice [3].

While deep learning may easily address multiple dimensions, DTW techniques are traditionally restricted to one-dimensional data. A multi-dimensional treatment of these objects is usually done by *Multiple Kernel Learning* (MKL).

Generally, MKL aims to derive one strong kernel as a combination of various weak base kernels as an input to an arbitrary kernel method. For this reason, multiple kernel learning is frequently used in information fusion where each kernel was derived by a different similarity measure or originates from different input sources (e.g. for analyzing text, video, and audio data at the same time)[11]. However, multiple kernel learning is highly limited by strong mathematical constraints (such as positive semi-definiteness [5]) that prevent the immediate usage of arbitrary distance or similarity measures in MKL. An overview of correction techniques to process non-psd data is given in [2] and [12]. By applying these techniques, MKL can be used again, but the choice of the appropriate correction approach is not easy and more important the overall runtime and memory complexity is quadratic or cubic.

In general, developing a suitable classification model for time series data with multiple perspectives is a challenging task with many obstacles.

### 3 Multi-perspective embedding of non-psd proximities

To address the aforementioned limitations, we propose a multi-perspective embedding into a new (and potentially complex-valued) vector space. The approach in [8] provides a method to embed a single non-metric matrix into a complex-valued vector space. This approach is extended by considering multiple non-metric matrices and by employing a metric relevance learning technique. An individual embedding of the various non-metric matrices may lead to complex-valued vectorial representations (as we will see later), which we call perspectives. The approach is referred to as *Multi-Perspective Embedding* (MPE). Thereby, each time series is finally represented by a large real-valued or complex-valued vector with respect to the chosen embedding variant. The individual dimensions of this vector can be traced back to the different perspectives, but also to the individual time series, which served as parameters of the embedding, as we explain in the following. Additionally, the so-far isolated perspectives (or subspaces) are linked by a Mahalanobis-like, learned interpretable distance measure.

Considering a set  $\mathcal{X} = \{t^1, \dots, t^N\}$  in some real-valued input space  $\mathbb{R}^F$ , with  $F$  features and  $N$  time series objects  $t^i$ . Each time series object  $t^i$  is represented by  $M^i$  sample points  $x_{ij} = t_j^i$  with  $x_{ij} \in \mathbb{R}^F$ . Let  $\mathcal{D}_f \in \mathbb{R}^{N \times N}$  be the symmetric dissimilarity matrix for the  $f$ -th feature by applying DTW on pairwise time series restricted to the input feature  $f$ . Let  $\mathcal{S}_f \in \mathbb{R}^{N \times N}$  be the symmetric similarity matrix derived from  $\mathcal{D}_f$  after *double centering*, as shown in [8]. If  $\mathcal{S}_f$  has negative eigenvalues, it is called an *indefinite matrix*. Such negative eigenvalues cause some serious issues when used in kernel methods like SVMs [5].

The complex-valued embedding space  $\mathcal{C}^L$  is obtained by embedding  $\mathcal{S}_f$  using the approach from [8], where  $L$  is the number of landmarks. As an alternative,

---

**Algorithm 1** (Complex-valued) embedding of multiple time series proximities

---

```

Multi-perspective Embedding( $S := [S_1, \dots, S_f], l, P^* := []$ )
for  $S_f$  in  $S$  do
   $K_1, K_2 := \text{Nyström-Approximation}(S_f, l)$ , using [2] for similarities and kmeans++
   $[C, A] := \text{eig}(K_2)$ ; with eigenvectors  $C$  and eigenvalues in  $A$  (diagonal)
  if complex-valued then
     $W := \text{diag}(\text{sqrt}(1./\text{diag}(A))) \cdot C^T$  complex-valued Nyström projection matrix
  else
     $W := \text{diag}(\text{sqrt}(1./\text{diag}(|A|))) \cdot C^T$  real-valued Nyström projection matrix
  end if
   $M := W \cdot K_1$        $P^* := [P^*, M]$       complex-valued embedding
end for
return  $P^*$        $\mathbf{x}_i$  := i-th row of  $P^*$ 

```

---

the embedding can be calculated by an eigen decomposition and a projection:

$$\mathcal{S}_f = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \quad \mathbf{V} = \mathbf{U}|\mathbf{\Lambda}|^{\frac{1}{2}} \quad (1)$$

leading to a real-valued vector space, due to the so-called flipping strategy (the  $|\cdot|$  operator in Eq. (1)), as suggested in [4], or a complex-valued vector space if the  $|\cdot|$  operator is left, as suggested in [8]. Each embedded vector  $\mathbf{x}_k \in \mathcal{C}^{L \times F}$  can be related back to an original input time series  $t^k$  as encoded by pairwise score values using the various dissimilarity matrices  $\mathcal{D}_f$ . The comparison of two embedded vectors is done in a norm using:  $d(\mathbf{x}_k, \mathbf{x}_l) = \|\mathbf{\Omega}(\mathbf{x}_k - \mathbf{x}_l)\|_2^2$  with  $\mathbf{\Omega}$  a linear projection matrix. This matrix can be learned as outlined in [9]. In our application, this matrix links the multiple perspectives of the time series to each other and permits an interpretable importance weighting of the single perspectives and the relevance of individual time series within a perspective<sup>1</sup>. The entire pipeline for multi-perspective embedding is detailed in Algorithm 1.

The aforementioned procedure can be done with linear costs following ideas discussed in [2], by carefully interlinking the individual steps and using a Nyström matrix approximation to avoid squared matrices.

## 4 Experiments

We evaluate the efficiency of the approach on various time series data and multiple classifiers, results shown in Tab. 1. The multi-dimensional time series are embedded into a single (complex-valued) vector space by using Algorithm 1. The following multi-dimensional time series data are used in the experiments:

The **Arabic Spoken Digit** (AS) data set [13] contains time series about spoken Arabic digit recognition, with 10 repeated recordings of 10 digits for each of the 88 speakers, resulting in a total of 8800 series in 13 features.

The **Handwritten Characters** (HC) data set, taken from [13], comprises 2,858 time series entries of trajectory data in 20 classes and 3 features.

---

<sup>1</sup>Plotting  $\mathbf{\Omega}$ , the weights can be easily interpreted, skipped due to lack of space.

The **Presence Detection** (PD) data set is taken from [14] and deals with device-free presence detection via Bluetooth Low Energy sensors in lecture halls. The data are given as 83 time series in 96 dimensions in 4 classes.

#### 4.1 Classifier Methods

**Nearest neighbor for complex-valued data:** As a baseline model, we used the nearest neighbor classifier with DTW as distance measure (NN-DTW). Here, the proximity matrices are averaged and hence conducted into one single proximity matrix to infer an overall label prediction. This approach is computationally expensive since either all distances have to be recalculated after each iteration or the entire distance matrix has to be stored.

**Easy and Average Multiple Kernel Learning:** EasyMKL and AverageMKL are two methods from the multiple kernel learning domain, implemented in MKLpy [11] used in combination with a Support Vector Machine (SVM) whose parameters were optimized via a grid search. The MKL models have been calculated on the various  $\mathcal{S}_f$ .

**Complex-valued Generalized Learning Vector Quantization:** We employ the cGMLVQ algorithm proposed in [9] as a variant of a prototype-based learning algorithm for classifications. To interlink the multi-perspective embeddings and to scale the importance of the various contributions, a matrix relevance learning was used. For simplicity, we employ one prototype per class.

#### 4.2 Results

We evaluate our complex-valued multi-perspective embedding on the aforementioned data in a ten-fold cross-validation with a hold-out test set. All classification models were trained and evaluated on the same training and test splits or their embedding counterparts. Table 1 shows the performance of the considered methods and algorithms.

Data	No Embedding			Real Embedding		Complex Embedding	
	NN DTW	Easy MKL	Average MKL	MPE- NN	MPE- GMLVQ	MPE- NN	MPE- cGMLVQ
AS	0,94	0,88	0,88	0,91	0,92	0,94	<b>0,95</b>
HC	0,90	0,94	0,92	0,94	0,95	<b>0,96</b>	<b>0,96</b>
PD	0,79	0,82	0,67	0,94	0,96	0,94	<b>0,98</b>

Table 1: Classification results on multi-dimensional time series data using our approach in comparison to other strategies (Standard deviations always small).

In our experiments, the baseline classifier NN-DTW performed comparably to other methods. Apparently, NN-DTW is originally less efficient for multi-dimensional time series data, but by averaging all perspectives, this method becomes competitive. Slightly inferior results were achieved by EasyMKL and AverageMKL, motivating the use of multiple kernels in the computations. These models are, however, valid only for psd kernels. The best results over all data sets

have been obtained by the complex-valued multi-perspective embeddings. By learning the relevance of features in cGMLVQ, we get interpretable models and better insights. We not only discover which embedded feature are particularly important but also which original time series  $t^i$  are particularly relevant. Due to the low-rank approximation used in Algorithm 1, the proposed strategy also has low memory and runtime costs.

## 5 Conclusions

In this work, we presented an efficient and fast strategy for learning from multi-dimensional time series data, with promising results. Compared to other methods, our approach is competitive and, in some cases, significantly better (especially on data set PD). Our approach also provides better interpretability and less memory and computational costs. Here, we focused on non-metric proximity measures and time series analysis, but our proposal is more general, which will be addressed in a journal publication.

## References

- [1] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller. Deep learning for time series classification: a review. *Data Min. Knowl. Discov.*, 33(4):917–963, 2019.
- [2] Frank-Michael Schleif and Peter Tiño. Indefinite proximity learning: A review. *Neural Comput.*, 27(10):2039–2096, 2015.
- [3] A. Shifaz, C. Pelletier, F. Petitjean, and G. I. Webb. Elastic similarity measures for multivariate time series classification. *CoRR*, abs/2102.10231, 2021.
- [4] E. Pekalska and R. P. W. Duin. *The Dissimilarity Representation for Pattern Recognition - Foundations and Applications*. WorldScientific, 2005.
- [5] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.
- [6] M. Shokoohi-Yekta, J. Wang, and E. J. Keogh. On the non-trivial generalization of dynamic time warping to the multi-dimensional case. In *Proc. of the 2015 SIAM Intern. Conf. on Data Mining, Vancouver, Canada, 2015*, pages 289–297, 2015.
- [7] M. Cuturi, J. Vert, Ø. Birkenes, and T. Matsui. A kernel for time series based on global alignments. In *Proc. of ICASSP 2007*, pages 413–416. IEEE, 2007.
- [8] M. Münch, M. Straat, M. Biehl, and F.-M. Schleif. Complex-valued embeddings of generic proximity data. In *S+SSPR 2020*, volume 12644 of *LNCS*, pages 14–23. Springer, 2020.
- [9] M. Straat, M. Kaden, M. Gay, T. Villmann, A. Lampe, U. Seiffert, M. Biehl, and F. Melchert. Learning vector quantization and relevances in complex coefficient space. *Neural Computing and Applications*, Mar 2019.
- [10] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. In *1st ICLR 2013*, 2013.
- [11] I. Lauriola and F. Aioli. Mklpy: a python-based framework for multiple kernel learning. *CoRR*, abs/2007.09982, 2020.
- [12] S. Mehrkanoon, X. Huang, and J. A. K. Suykens. Indefinite kernel spectral learning. *Pattern Recognit.*, 78:144–153, 2018.
- [13] L. v. d. Maaten. Learning discriminative fisher kernels. In L. Getoor and T. Scheffer, editors, *Proc. of 28th ICML, 2011*, pages 217–224. Omnipress, 2011.
- [14] M. Münch, K. Huffstadt, and F.-M. Schleif. Towards a device-free passive presence detection system with bluetooth low energy beacons. In *27th Europ. Symp. on Artif. Neur. Netw., ESANN 2019, Bruges, Belgium, 2019*, 2019.