Multiple-Kernel Dictionary Learning for Sparse Reconstruction of Unseen Multivariate Time-series *

Babak Hosseini[0000-0003-3335-4289] and Barbara Hammer

CITEC center of excellence, Bielefeld University, Bielefeld, Germany
{bhosseini,bhammer}@techfak.uni-bielefeld.de

Abstract. A considerable amount of research has been devoted to designing algorithms for description and recognition of unseen classes in datasets. One important application of such algorithms is multivariate time-series (MTS) such as motion data and audio signals, in which it is usually expected to observe partial similarities between specific dimensions of the unseen and seen classes. In this work, we propose a novel unsupervised framework for the analysis of the unseen categories in MTS datasets from the above perspective. Our framework learns a unique multiple-kernel dictionary (MKD) which selects and combines specific dimensions of the MTS taken from the seen classes. We incorporate the designed MKD in a non-negative sparse coding framework (MKD-SC) to obtain an encoding that fully/partially describes the unseen MTS categories. This achieved description is interpretable via finding possible relations between the data dimensions of the seen and unseen classes. Furthermore, by treating such dimension-based sparse encodings as compact semantic attributes, we propose a simple but effective incremental clustering algorithm to categorize the unseen data classes in an unsupervised way based on those encoded attributes. Our incremental clustering can gradually build a hierarchal tree via on-line observation of the unknown (unseen) MTS samples. Empirical evaluation on real benchmarks demonstrates the effectiveness of our MKD-SC framework in the interpretable reconstruction of the unseen data regarding their dimension-based overlapping with the seen classes. It also shows the effect of the learned attributes (encodings) on improving the clustering performance of the unseen MTS categories.

Keywords: Multivariate time-series, dictionary learning, unseen classes

1 Introduction

Zero-shot learning is the problem of recognizing novel categories of data when no prior information is available during the training phase [2, 9, 15]. One effective approach to such transfer learning is the incorporation of semantic attributes as descriptive features to map the input data to an intermediate semantic space, which can discriminate between different unseen categories [9, 15]. Another concern in this area of research is the partial/complete reconstruction of the unseen classes based on their relation to the learned semantic attributes or to the training data [12, 13].

One important application of zero-shot learning is multivariate time-series (MTS) in the general meaning such as audio data and human motions [5, 11] with a considerable

* This research was supported by the Cluster of Excellence Cognitive Interaction Technology ‘CITEC’ (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG).
number of unknown classes. Different from image and video data, MTS do not possess any general spatial dependency between its dimensions. Nevertheless, it is usually expected to find semantic attributes shared between different classes of a MTS dataset. As an example of MTS data, consider the Cricket Umpire signal Out in Figure 1, which can be described as the left hand is raised while the right hand is down. This can provide us with a semantic understanding of the data without having any prior knowledge about its class. Furthermore, such descriptions can be considered as semantic attributes in order to distinguish the unknown MTS data samples into distinct categories which reflect their unknown ground truth labels. Although the semantic descriptions are class specific, the individual attributes could be shared among different classes with partial similarities.

Sparse coding (SRC) is the idea of constructing an input data using weighted combinations (sparse codes) of sparse selected entries from a set of learned bases (dictionary). Such sparse representations can capture essential intrinsic characteristics of a dataset [14]. Furthermore, via assuming an implicit mapping of the data to a high-dimensional feature space, it is possible to formulate SRC using the kernel representation of the data [8] to also model nonlinear data structures. Consequently, a subset of the existing research has benefited from SRC methods in designing more effective attributes for dealing with unseen classes of data; however, these efforts are mainly limited to the image (spatial) and video (spatiotemporal) datasets [13, 16]

Despite the current achievements in learning unseen MTS data, the existing methods are either depended on having prior information about the novel classes (e.g. samples/labels) [11], or they cannot interpret the unseen data based on their learned attributes. Furthermore, to our knowledge, there is no research reported on the partial/complete reconstruction of unseen classes for MTS data in general (e.g. recorded motion signals).

Contributions: In this paper, we propose a novel framework for having a semantic description and categorizing the unseen classes in a MTS dataset based on dictionary learning in the feature space. The following contributions distinguish our work from the relevant literature:

- We design a novel dictionary structure which learns attributes that are effective in representing multivariate time-series based on their individual dimensions.

Fig. 1: General overview of our framework. The dictionary (MKD) learns the semantic attributes based on the seen classes. These attributes are used for a semantic description of the data from the unseen classes, which leads to categorizing and partial reconstruction of the data.
We propose an unsupervised kernel-based SRC method for partial reconstruction of unseen MTS data in the feature space along with their interpretable encoding. We design an incremental clustering algorithm based on the sparse encodings of the unseen data which gradually creates a clustering dendrogram of the unseen classes.

In the next section, we introduce and explain our proposed framework, and we evaluate it in Sec. 4, followed by the conclusion section.

2 Problem Statement

Presenting a multivariate time-series in the vectorial space, each data sequence \( i \) is denoted by \( X_i = [x_i(1) \ldots x_i(T)] \in \mathbb{R}^{f \times T} \), where \((f, T)\) represents the number of dimensions and different lengths for \( T \) respectively in the MTS sequences. Assuming an implicit data tensor \( X = \{X_i\}_{i=1}^N \in \mathbb{R}^{N \times f \times T} \) contains the observed data classes, its corresponding label vector \( l \in \mathbb{R}^N \) maps \( X \) to \( c \) distinct known data classes. On the other hand, there exists a set of unseen MTS data \( Z \in \mathbb{R}^{N_z \times f \times T} \) with a corresponding label vector \( q \in \mathbb{R}^{N_z} \) such that \( q \cap l = \emptyset \). Based on the above description, we are interested in solving the following problems:

- Obtaining semantic descriptions (Figure 1) for each sequence \( Z_i \in Z \) which can explain/reconstruct the dimensions of \( Z_i \) based on the seen classes \( X \).
- Using the achieved semantic descriptions to distinguish the sequences of \( Z \) into distinct categories which correspond to the unknown label vector \( q \).

3 Multiple-Kernel Dictionary Learning Framework

Similar to Figure 1, it is often expected in the real-life problems (e.g. human motions) to find partial similarities between the seen and unseen classes of MTS data for a subset of their dimensions. Therefore, these similarities can lead to an interpretable description of the unknown classes. Such a description can also lead to better categorizing of the novel data without having any prior information on its class labels. To achieve the above, we design a specific multiple-kernel dictionary (MKD) structure which is trained based on \( X \) and learns semantic attributes similar to Figure 1. This is done via combining dimensions of similar MTS samples in the feature space under non-negativity constraints. These attributes can encode each unseen \( Z_i \in Z \) leading to an interpretable description of its dimensions and to better separate it from previous (unknown) classes in \( Z \) (Figure 1).

To be more specific, we assume there exists \( f \) individual non-linear implicit functions \( \{\Phi_i(X)\}_{i=1}^f \) for mapping the respective dimensions of the sequence \( X \) individually into RKH-spaces corresponding to kernels \( K_i \) [8]. A weighted combination of these kernels with coefficients \( \beta_i \geq 0 \) induces an embedding of the data \( \Phi(X, \beta) := [\sqrt{\beta_1}\Phi_1(X)^\top \ldots \sqrt{\beta_f}\Phi_f(X)^\top]^\top \) in the feature space, which can be obtained via the vector \( \beta \in \mathbb{R}^f \) and applies different weights to the individual dimensions.

We assume that cluster structures are complemented by different weighting schemes of the individual kernels. Thus, we consider \( k \) different weighting schemes and denote \( B = [\beta_1 \ldots \beta_k] \in \mathbb{R}^{f \times k} \) as the weighting matrix. Kernel sparse coding selects dictionary functions as weighted sums of elements of this embedding. We define the novel multiple kernel dictionary (MKD) matrix \( \Phi_B(U) \) as

\[
\Phi_B(U) := [\Phi(X, \beta_1)u_1 \ldots \Phi(X, \beta_k)u_k]
\]

where \( U = [u_1 \ldots u_k] \in \mathbb{R}^{N \times k} \) (1)
Each dictionary column $\Phi(\mathcal{X}, \beta_i)u_i$ is a weighted combination of dimensions and samples from $\mathcal{X}$ which can represent semantic attributes similar to those of Figure 1. The coefficient matrix $U$ of the dictionary is subject of an optimization, to obtain the best representational power for the given data.

Sparse coding aims for a reconstruction of the data $\mathcal{X}$ in the kernel space via the MKD: $\Phi(\mathcal{X}) \approx \Phi_B(U)\Gamma$ with a matrix of sparse codes $\Gamma = [\gamma_1 \ldots \gamma_N] \in \mathbb{R}^{k \times N}$. We propose the following MKD sparse coding framework (MKD-SC) for training the weighting scheme of the dimensions which enhance these similarities, i.e. cluster structure.

The problem can be optimized via the non-negative quadratic pursuit (NQP) algorithm. The term $\lambda$ is the trade-off constant between different parts of the objective function, and $u_{ij}, \beta_{ij}, \gamma_{ij}$ address the $j$-th entry of the $i$-th column of $U, B$, and $\Gamma$ respectively.

The rationale behind this objective is as follows: The first part of the loss term measures the reconstruction error of the sparse coding based on the Frobenius norm $\|.\|_F$. The term $\|.\|_0$ denotes the $l_0$-norm which employs sparsity constraints for elements of $\Gamma$ via the constant $T_0$ which results in having each $\mathbf{X}$ constructed by sparse contributions from $\mathcal{X}$. The $l_2$-norm constraint on $\Phi(\mathcal{X}, \beta_i)u_i$ prevents the optimization solutions from becoming degenerated [14].

Inspired from [6], the second loss term in the objective of Eq. 2 enforces each dictionary vector $u_i$, to use data samples from $\mathcal{X}$, such that their weighted features $\Phi(x_i, \beta_i)$ are local neighbors in the feature space. Hence this objective has a tendency to focus on dictionary elements which group similar time series together, and which chooses weighting scheme of the dimensions which enhance these similarities, i.e. cluster structures are mirrored in the dictionary elements and coefficient weighting schemes. Hence the dictionary $\Phi_B(U)$, which results from the optimization problem in Eq. 2, contains attributes, which are weighted combinations of different exemplars and dimensions from $\mathcal{X}$. The non-negativity constraints result in having similar resources combined together which leads to learning semantic attributes for $\Phi_B(U)$ and an interpretable sparse description based on each $\gamma_i$ [7].

**Estimating the Sparse Codes of Unseen data:** Encoding an unseen data $z$ based on the trained $\{U, B\}$ is possible via optimizing the reconstruction error of this problem, i.e. optimizing Eq. 2 with $\lambda = 0$. In the Sec. 3.2 and 3.3, we benefit from this framework to describe and categorize unseen multivariate time-series samples.

### 3.1 Optimization Scheme

We optimize the parameters $U, \Gamma$ and $B$ in alternating steps, such that at each update step, we optimize Eq. 2 with respect to one parameter while fixing the others. Similar to [6], via using the associated kernel matrices $\{\mathcal{K}_i(\mathcal{X}, \mathcal{X}) = \Phi_i(\mathcal{X})^\top \Phi_i(\mathcal{X})\}_{i=1}^f$ it is possible to formulate Eq. 2 in terms of each $\gamma_i$, $u_i$, and $\beta_i$ individually and without an explicit reference to the embeddings $\Phi_i$, which leads to $f$ constrained non-negative quadratic programings with the general form of

$$\min_{x} \frac{1}{2} x^\top H x + c^\top x \quad \text{s.t.} \|x\|_0 < T_0, \ x_i \in \mathbb{R}^+ \ \forall i$$

These problems can be optimized via the non-negative quadratic pursuit (NQP) algorithm from [4]. Thus, We take the following steps for each iteration of the training phase:
1- Updating $\gamma_i$, $\forall i = 1, \ldots, N$
2- Updating $u_i$ and $\beta_i$, in a subsequent order $\forall i = 1, \ldots, k$

Due to the page limit, we put the detail regarding the reformulation of Eq. 2 and the optimization steps in the online extended version of the paper.

The optimization framework of Eq. 2 is non-convex when considering $\{U, \Gamma, B\}$ together. However, each of the 3 sub-problems related to the update of $\gamma_i$, $u_i$, so $\beta_i$ are convex, and the main optimization loop converges in a limited number of steps.

3.2 Partial Reconstruction of Unseen Time-series

In realistic MTS datasets such as human actions, it is expected to observe partial similarities between the dimensions of different classes. Therefore, we define the following error measure for reconstruction of a selected set of dimensions $S$ related to $Z$:

$$J_{\text{rec}}^S(Z, B, U) = \|1^S \Phi(Z) - 1^S \Phi_B(U) \Gamma\|_2^2 / \|1^S \Phi(Z)\|_2^2$$

(4)

where $B^S$, and $I^S$ are modified versions of $B$ and the identity matrix respectively via making all the entries zero except the rows corresponding to $S$. Consequently, the learned dictionary $\Phi_B(U)$ can partially reconstruct the unseen time-series $Z$ for the subset $S$ of its dimensions, if $J_{\text{rec}}^S(Z, B, U)$ is relatively small. Provided a small reconstruction error $J_{\text{rec}}^i(Z, B, U)$ is present, we can look at component-wise class assignments: the $i$-th dimension of $Z$ can be assigned to the class $q$ from $X$ which has the highest contribution to the reconstruction of this dimension of $Z$, as $q = \arg \max_q \sum_{j: \Phi_j = q} \sum_{r=1}^h \beta_{ir} \gamma_{jr}$.

3.3 Incremental Clustering of Unseen Time-series

Based on the discussion of section 3.2 and relying on the partial similarity of different MTS classes and descriptive quality of the learned attributes of the dictionary, we propose Algorithm 1. This algorithm incrementally clusters the unseen sequences from $Z$ into a dendrogram $\mathcal{H}$ as they are observed in an online fashion, as well as finding potential sub-clusters among them. To that aim, for each unknown MTS sequence $Z$ we prepare an encoding matrix $R \in \mathbb{R}^{N \times f}$, $i$-th column of which represents the weights of contribution from $X$ in the reconstruction of the $i$-th dimension of $Z$. Therefore,

<table>
<thead>
<tr>
<th>Algorithm 1: Incremental Clustering of an Encoded MTS data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input:</strong> $R$: Calculated encoding matrix of the unseen data $Z$, $\mathcal{H}$: The hierarchical tree.</td>
</tr>
<tr>
<td><strong>Output:</strong> Place of $Z$ in the hierarchy $\mathcal{H}$.</td>
</tr>
<tr>
<td>1. If $\exists C_n$ such that $d(Z, C_n) \leq d(C_n)$ then</td>
</tr>
<tr>
<td>2. If $C_n$ is a leaf node then add $Z$ to $C_n$;</td>
</tr>
<tr>
<td>3. If $(d(C_{n_1}) + d(C_{n_2})) / 2 \leq k_{\text{cluster}}$ then</td>
</tr>
<tr>
<td>4. split $C_n$ into $C_{n_1}$ and $C_{n_2}$ using $k$-means;</td>
</tr>
<tr>
<td>5. If $(d(C_{n_1}) + d(C_{n_2})) / 2 \leq k_{\text{merge}}$ then</td>
</tr>
<tr>
<td>6. Replace $C_n$ with $C_{n_1}$ and $C_{n_2}$;</td>
</tr>
<tr>
<td>7. Re-cluster ${C_{n_1}, C_{n_2}}$ with same branch leaves;</td>
</tr>
<tr>
<td>8. else add ${C_{n_1}, C_{n_2}}$ as the children of $C_n$;</td>
</tr>
<tr>
<td>9. else Create a new child for $C_n$ as $C_{n_1}$ and add $z$ to it;</td>
</tr>
<tr>
<td>10. else Create a new leaf at the top level containing $Z$;</td>
</tr>
</tbody>
</table>

---

1 https://github.com/bab-git/MKD_U/seen_MTS
Table 1: Information regarding the selected datasets for the experiments.

<table>
<thead>
<tr>
<th>Dataset</th>
<th># sequences</th>
<th># classes</th>
<th># dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cricket Umpire</td>
<td>180</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>CMU mocap</td>
<td>200</td>
<td>20</td>
<td>62</td>
</tr>
<tr>
<td>Articulatory Words</td>
<td>575</td>
<td>25</td>
<td>9</td>
</tr>
<tr>
<td>Squat</td>
<td>81</td>
<td>3</td>
<td>142</td>
</tr>
</tbody>
</table>

Table 2: DRA measure (%) for unseen Cricket classes, and their partial descriptions based on the training classes. Based on Table 2 and Figure 2 report the reconstruction quality and the interpretation of the unseen classes respectively. Based on Table 2: DRA measure (%) for unseen Cricket classes, and their partial descriptions based on the training classes.

\[ r_{ji} = \sum_{l=1}^{k} \beta_{i\ell} u_{j\ell} \tau_{l} \] where \( r_{ji} \) denotes the \( j \)-th entry of the \( i \)-th column of \( R \). This matrix is considered as a rich encoded descriptor for dimensions of \( Z \) based on \( X \) and is used in Algorithm 1 to compute \( Z \) to the previously categorized unseen data in \( H \) to find the best place for \( Z \) in the dendrogram. Line 1 of the algorithm finds \( C_{n} \) as the most similar node to \( Z \) based on the distance term \( d(Z, C_{n}) = \| R_{Z} - R_{C_{n}} \|_{F}^{2} \), and the intra-cluster distance for each node \( C_{n} \) as \( d(C_{n}) = E_{Z_{i} \in C_{n}} \|d(R_{Z_{i}}, R_{C_{n}})\| \), where \( R_{C_{n}} = E_{Z_{i} \in C_{n}}[R_{Z_{i}}] \).

In line 7, each of \( \{C_{n1}, C_{n2}\} \) is merged with the best \( C_{m} \) from the same branch to form \( C_{m+n} \) which satisfies \( 2d(C_{m+n})/[d(C_{m}) + d(C_{n})] \leq k_{\text{clust}} \) for \( i = 1, 2 \).

Although, one might want to tune the parameters \( \{k_{\text{rmv}}, k_{\text{clust}}\} \) (e.g. via doing a cross-validation on \( X \)), in practice, choosing \( k_{\text{rmv}} = 0.3 \), \( k_{\text{clust}} = 0.5 \) gives an acceptable clustering result.

## 4 Experiments

To evaluate the performance of our sparse coding framework for representation and discrimination of unseen data, we choose the MTS datasets Cricket Umpire, CMU mocap, Articulatory Words, and Squat with the descriptions provided by [7] as well as in Table 1. For all the datasets, to compute the kernel matrices for individual dimensions of the data, we use the Gaussian kernel \( K_{l}(X_{i}, X_{j}) = \exp(-D_{l}(X_{i}, X_{j})/\delta_{l}) \forall l = 1 \ldots f \), where \( D_{l}(X_{i}, X_{j}) \) is the computed pairwise DTW-distance [1] (can be substituted with any other data distance) between the \( l \)-th dimension of \( X_{i} \) and \( X_{j} \). Parameter \( \delta_{l} \) is set equal to the average of \( D_{l}(x_{i}, x_{j}) \) distances over all the data samples. To make sure that \( K_{l}(X, X) \) remains PSD, we use the common clipping operation by setting its negative eigenvalues to zero. For the tuning of \( (\lambda, T_{0}) \) in Eq. 2, we use 5-fold cross-validation based on its objective value. Also, the dictionary size is determined as \( k = \{\# \text{seen classes}\} \times T_{0} \).

### 4.1 Partial Reconstruction of Multivariate Time-series

In order to evaluate the reconstruction quality for each unseen data \( Z \), we define the dimension-reconstruction accuracy measure as \( DRA := \# \text{dimensions that } \beta_{ji} < 0.1 \) using Eq. 4. In addition, each reconstructed dimension of \( Z \) which satisfies the above threshold is interpreted via the class of data with the most contribution as in Sec. 3.2.

We select Cricket dataset to demonstrate how the proposed MKD-SC framework reconstructs the unseen classes in the dimension-level, and Table 2 and Figure 2 report the reconstruction quality and the interpretation of the unseen classes respectively. Based on Table 2: DRA measure (%) for unseen Cricket classes, and their partial descriptions based on the training classes.
on the results, the *No ball* class is fully reconstructed via its partial relation to the movement of the left hand in the *Short* class and to the right hand in the *Wide* class.

### 4.2 Incremental Clustering of the Unseen Time-series

The proposed incremental clustering is implemented on the unseen categories of Table 1 after computing the $R$ matrix based on the sparse code of each unseen data (Sec. 3.3). We use the average clustering error (CE) [10], and normalized mutual information (NMI) [3] for quantitative evaluation of the clustering results. To that aim, we cut each dendrogram from where it has the equal number of clusters to the ground truth.

**Baselines:** To evaluate the quality of attributes which are learned by MKD-SC framework, as the most relevant method we choose self-learning algorithm [11] without its novelty detection part. In addition, we apply the spectral clustering algorithm on the original kernel matrix $K(Z, X')$ to compare our framework to the normal clustering of $Z$. As another baseline, we also use the NNKSC algorithm [7] as the single-kernel predecessor of MKD-SC, for which the $R$ matrix becomes an $N$-dimensional vector.

Based on clustering dendrograms in Figure 3, the unseen Squat and Cricket classes are well categorized, which shows the effectiveness of the learned attributes for distinct representation of the unknown classes. The incremental clustering also categorized these unseen classes into a few sub-clusters. For Squat (Fig. 3-a), the 3 sub-clusters for the *Go down* class are related to different performance styles of the dataset’s 3 participants regarding this specific phase of the squat. Similarly, for each unseen category of the Cricket dataset (Fig. 3-b), there are sub-clusters recognized for each of the distinct main clusters which shows the existing structured variation within each of this classes.

According to the clustering results in Table 3, the proposed MKD-SC method provides encodings which lead to better clustering of the unseen data compared to the baselines. The superiority of the original kernels over NNKSC and self-learning methods depends on the discriminative quality of the original kernels, for instance for the Cricket dataset. Self-learning method can have a better performance than NNKSC and Kernel-based baselines if its descriptor-based features can better discriminate between the different categories of the unseen classes.
Table 3: Clustering error (CE) (%) and NMI the unseen categories.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Words</th>
<th>Squat CE</th>
<th>CMU CE</th>
<th>Cricket CE</th>
<th>NMI</th>
<th>Words</th>
<th>Squat CE</th>
<th>CMU CE</th>
<th>Cricket CE</th>
<th>NMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>MKD-SC (Proposed)</td>
<td>12.31</td>
<td>0.89</td>
<td>9.28</td>
<td>0.92</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-learning [11]</td>
<td>18.75</td>
<td>0.84</td>
<td>14.25</td>
<td>0.87</td>
<td>16.63</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NNKSC[7]</td>
<td>21.61</td>
<td>0.78</td>
<td>18.88</td>
<td>0.85</td>
<td>12.45</td>
<td>0.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spectral Clustering</td>
<td>27.51</td>
<td>0.76</td>
<td>23.45</td>
<td>0.76</td>
<td>8.04</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5 Conclusion

In this research, we proposed an unsupervised framework which provides interpretable analysis of unseen classes in MTS datasets. We proposed a novel multiple-kernel dictionary structure based on individual mappings of the MTS dimensions into the feature space represented by dimension-based kernel functions. The MKD uses weighted combinations of these kernels along with selections from training data in order to learn its compact descriptive attributes (bases). Based on the learned attributes, our unsupervised MKD-SC framework reconstructs the unseen classes (partially/fully) in the feature space according to overlapping of their dimensions with those of the seen categories and provides an interpretable description of the novel data. Based on the obtained sparse encodings, we proposed an incremental clustering to gradually categorize novel MTS, and to form a dendrogram. Experiments on MTS benchmarks showed that our MKD-SC framework can describe the dimensions of the unseen classes via their relation to those of the seen ones. In addition, the incremental clustering provides better clustering accuracy comparing to the baselines and also reveals the existing sub-clusters in the data.

References