ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0.
Available from http://www.i6doc.com/en/.

Multiple-Kernel Dictionary Learning for Reconstruction and Clustering of Unseen Multivariate Time-series

Babak Hosseini[1] and Barbara Hammer[1] *

CITEC cluster of excellence, Bielefeld University
Bielefeld, Germany

Abstract. There exist many approaches for description and recognition of unseen classes in datasets. Nevertheless, it becomes a challenging problem when we deal with multivariate time-series (MTS) (e.g., motion data), where we cannot apply the vectorial algorithms directly to the inputs. In this work, we propose a novel multiple-kernel dictionary learning (MKD) which learns semantic attributes based on specific combinations of MTS dimensions in the feature space. Hence, MKD can fully/partially reconstructs the unseen classes based on the training data (seen classes). Furthermore, we obtain sparse encodings for unseen classes based on the learned MKD attributes, and upon which we propose a simple but effective incremental clustering algorithm to categorize the unseen MTS classes in an unsupervised way. According to the empirical evaluation of our MKD framework on real benchmarks, it provides an interpretable reconstruction of unseen MTS data as well as a high performance regarding their online clustering.

1 Introduction

Zero-shot learning is the problem of recognizing novel categories of data when no prior information is available during the training phase [1, 2, 3]. One practical 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 [2, 3]. 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 the training data [4, 5].

An important application of zero-shot learning is multivariate time-series (MTS) in the general meaning such as audio data and human motions [6, 7] with a considerable number of unknown classes. Different from images and video, 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 an MTS dataset. As an example of MTS data, consider the Cricket Umpire signal *Out* in Fig. 1 which can be described as the *left hand is raised* while *the right hand is down*. Such encoding provides us with a semantic understanding of the data without having any prior knowledge about its class label. We can also consider such descriptions as semantic attributes in order to distinguish the unknown MTS data samples into distinct categories that reflect their unknown labels. Although the semantic descriptions are class specific, we can share the individual attributes among classes which have between-class 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 [8].

*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).

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0.
Available from http://www.i6doc.com/en/.

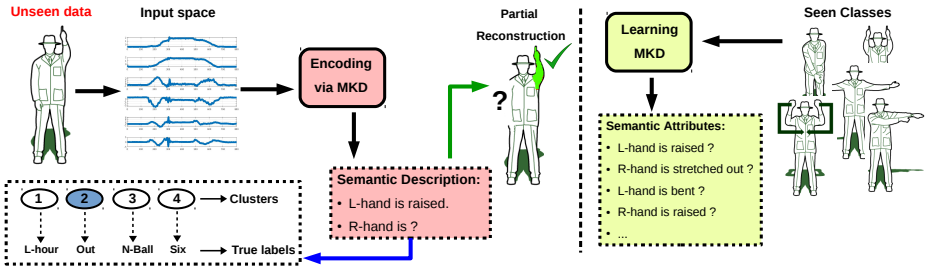


Fig. 1: General overview of our framework. The dictionary (MKD) learns the semantic attributes based on the seen classes. These attributes are used for a semantic description of the data from the unseen classes, which leads to categorizing and partial reconstruction of the data.

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 [9] to model also 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 [5, 10]

Despite the current achievements in learning unseen MTS data, either the existing methods are depended on having prior information about the novel classes (e.g., samples/labels) [7], 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).

To address the above concerns, we provide the following contributions:
1- We design a novel dictionary structure which learns attributes that can represent MTS based on the dimension level.
2- We propose an unsupervised kernel-based SRC method for partial reconstruction of unseen MTS data in the feature space along with their interpretable encoding.
3- We design an incremental clustering based on the sparse encodings of the unseen data which gradually creates a clustering dendrogram of the unseen classes.

After formulating the problem in Sec. 2, we introduce and explain our proposed framework in Sec. 4, and we evaluate it in Sec. 4 followed by the conclusion section.

2 Problem Statement

Presenting a multivariate time-series in the vectorial space, $\mathbf{X}_i = [\vec{x}_i(1) \dots \vec{x}_i(T)] \in \mathbb{R}^{f \times T}$ denotes sequence i, where (f, T) represents the number of dimensions and the sequence lengths respectively. The training set $\mathcal{X} = \{\mathbf{X}_i\}_{i=1}^{N}$ belongs to c distinct data classes with the label set $l = \{1, \cdots, c\}$. Accordingly, the set of unseen MTS $\mathcal{Z}$ belongs to the label set q, such that $q \cap l = \emptyset$. Based on the above description, we are interested in: **1-**Obtaining semantic attributes which create interpretable relations between sequences $\mathbf{Z}_i \in \mathcal{Z}$ and the seen classes $\mathcal{X}$ (Fig. 1). **2-**Using the obtained semantic attributes for efficient clustering of the unseen set $\mathcal{Z}$.

3 Multiple-Kernel Dictionary Learning Framework

Similar to Fig. 1, it is a common observation for real-world MTS data (e.g., human motions) to find partial similarities between different data classes when considering a

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

subset of their dimensions. Therefore, these similarities can lead to an interpretable description for a novel data sample (from $\mathcal{Z}$) via its relation to the seen classes (from $\mathcal{X}$). Furthermore, such a description leads to a better clustering of novel data points $\mathbf{Z}_i$ without having any prior information on their class labels. To achieve the above, we design a specific multiple-kernel dictionary (MKD) structure which is trained based on $\mathcal{X}$ and learns semantic attributes similar to Fig. 1-left. To be more specific, MKD combines dimensions of similar MTS samples in the feature space under non-negativity constraints. These attributes can encode each unseen $\mathbf{Z}_i \in \mathcal{Z}$ as an interpretable description of its dimensions and to better separate it from previous (unknown) classes in $\mathcal{Z}$ (Fig. 1-right).

To be more specific, we assume there exist f non-linear implicit kernel functions $\{\Phi_i(\mathbf{X})\}_{i=1}^{f}$ to map each dimension of $\mathbf{X}$ into an individual RKH-spaces [9]. A weighted combination of these kernels with individual coefficients $\beta_i \geq 0$ (entries of $\vec{\beta}$) induces an embedding of the data in the feature space as $\Phi(\mathbf{X}, \vec{\beta}) := [\sqrt{\beta_1}\Phi_1(\mathbf{X})^\top \cdots \sqrt{\beta_f}\Phi_f(\mathbf{X})^\top]^\top$. We can apply this embedding to the whole training data via $\Phi(\mathcal{X}, \vec{\beta}) := [\Phi(\mathbf{X}_1, \vec{\beta}) \cdots \Phi(\mathbf{X}_N, \vec{\beta})]$, and additionally we consider k different weighting schemes of the individual kernels as $\mathbf{B} = [\vec{\beta}_1 \cdots \vec{\beta}_k] \in \mathbb{R}^{f \times k}$ to complement different existing classes in the data. Now, We define our novel multiple kernel dictionary (MKD) matrix $\Phi_{\mathbf{B}}(\mathbf{U})$ as

$$\Phi_{\mathbf{B}}(\mathbf{U}) := [\Phi(\mathcal{X}, \vec{\beta}_1)\vec{u}_1 \cdots \Phi(\mathcal{X}, \vec{\beta}_k)\vec{u}_k] \qquad \text{where } \mathbf{U} = [\vec{u}_1 \dots \vec{u}_k] \in \mathbb{R}^{N \times k}.$$

Each dictionary column $\Phi(\mathcal{X}, \vec{\beta}_i)\vec{u}_i$ is a weighted combination of selected dimensions and selected samples from $\mathcal{X}$ based on the value of $\vec{\beta}_i$ and $\vec{u}_i$ respectively. Due to the relation of $\Phi(\mathcal{X}, \vec{\beta}_i)\vec{u}_i$ to different dimensions of $\mathcal{X}$, its columns can learn semantic attributes similar to those of Fig. 1.

To fit $(\mathbf{U}, \mathbf{B})$ to the data efficiently, we aim for the sparse reconstruction $\Phi(\mathcal{X}) \approx \Phi_{\mathbf{B}}(\mathbf{U})\mathbf{\Gamma}$ in the feature space based on a sparse matrix of codings $\mathbf{\Gamma} = [\vec{\gamma}_1 \dots \vec{\gamma}_N] \in \mathbb{R}^{k \times N}$. To that aim, We propose the following MKD sparse coding framework (MKD-SC) for training the dictionary parameters $(\mathbf{B}, \mathbf{U})$ and sparse codes $\mathbf{\Gamma}$:

$$\begin{aligned} \min_{\mathbf{B},\mathbf{\Gamma},\mathbf{U}} \quad & \|\Phi(\mathcal{X}) - \Phi_{\mathbf{B}}(\mathbf{U})\mathbf{\Gamma}\|_F^2 \\ \text{s.t.} \quad & \|\vec{\gamma}_i\|_0 < T_0, \;\; \|\Phi(\mathcal{X}, \vec{\beta}_i)\vec{u}_i\|_2^2 = 1, \;\; u_{ij}, \beta_{ij}, \gamma_{ij} \in \mathbb{R}^+, \;\; \forall ij, \end{aligned} \tag{1}$$

where $u_{ij}, \beta_{ij},$ and γ_{ij} denote the j-th entry of the i-th column of $\mathbf{U}, \mathbf{B},$ and $\mathbf{\Gamma}$ respectively. The loss term in Eq. 1 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 $\mathbf{\Gamma}$ via the constant T_0 which results in having each $\mathbf{X}_i$ constructed with sparse contributions from $\mathcal{X}$. The l_2-norm constraint on $\Phi(\mathcal{X}, \vec{\beta}_i)\vec{u}_i$ prevents the optimization solutions from becoming degenerated [8].

Hence the dictionary $\Phi_{\mathbf{B}}(\mathbf{U})$, which results from the optimization problem in Eq. 1, contains attributes (columns), which are weighted combinations of different exemplars and dimensions from $\mathcal{X}$. The non-negativity constraints result in having similar resources become combined which leads to learning semantic attributes for $\Phi_{\mathbf{B}}(\mathbf{U})$ and an interpretable sparse description based on each $\vec{\gamma}_i$ [11]. In the Sec. 3.2 and 3.3, we benefit from this framework to describe and categorize unseen MTS samples.

3.1 Optimization Scheme

We optimize the parameters $\mathbf{U}$, $\mathbf{\Gamma}$, and $\mathbf{B}$ in alternating steps, such that at each update step, we optimize Eq. 1 with respect to one parameter while fixing the others. Based

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

on the dot-product relations $\{\mathcal{K}_i(\mathcal{X},\mathcal{X}) = \Phi_i(\mathcal{X})^\top \Phi_i(\mathcal{X})\}_{i=1}^{f}$, it is possible to rewrite Eq. 1 in terms of each of $(\vec{\gamma}_i, \vec{u}_i, \vec{\beta}_i)$ individually to obtain a general convex form of

$$\min_{\vec{x}} \tfrac{1}{2}\vec{x}^\top \mathbf{H}\vec{x} + \vec{c}^\top \vec{x} \qquad \text{s.t. } \|\vec{x}\|_0 < T_0, \;\; x_i \in \mathbb{R}^+ \;\; \forall i, \tag{2}$$

in which $(\mathbf{H}, \vec{c})$ are computed without any explicit reference to the embeddings Φ_i. Such problems can be optimized via the non-negative quadratic pursuit (NQP[1]) algorithm from [12]. Due to the page limit, we will put the detail regarding the reformulation of Eq. 1 and the optimization steps in the online extended version of the paper [2].

3.2 Partial Reconstruction of Unseen MTS

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 the reconstruction of a selected set of dimensions $\mathcal{S}$ related to data $\mathbf{Z}$:

$$\mathcal{J}_{rec}^{\mathcal{S}}(\mathbf{Z},\mathbf{B},\mathbf{U}) = \|\mathbf{I}^{\mathcal{S}}\Phi(\mathbf{Z}) - \mathbf{I}^{\mathcal{S}}\Phi_{\mathbf{B}}(\mathbf{U})\mathbf{\Gamma}\|_2^2 / \|\mathbf{I}^{\mathcal{S}}\Phi(\mathbf{Z})\|_2^2 \tag{3}$$

where $\mathbf{B}^{\mathcal{S}}$, and $\mathbf{I}^{\mathcal{S}}$ are modified versions of $\mathbf{B}$ and the identity matrix respectively via making all the entries zero except the rows corresponding to $\mathcal{S}$. Consequently, the learned dictionary $\Phi_{\mathbf{B}}(\mathbf{U})$ can partially reconstruct the unseen time-series $\mathbf{Z}$ for the subset $\mathcal{S}$ of its dimensions, if $\mathcal{J}_{rec}^{\mathcal{S}}(\mathbf{Z},\mathbf{B},\mathbf{U})$ is relatively small.

3.3 Incremental Clustering of Unseen MTS

We propose Algorithm 1 relying on the partial similarity of different MTS classes and the descriptive quality of the learned attributes of MKD. This algorithm incrementally clusters the unseen sequences of $\mathcal{Z}$ into a dendrogram $\mathcal{H}$ in an online fashion, and also finds the potential sub-clusters among them. To that aim, for each unknown MTS sequence $\mathbf{Z}$, we prepare an encoding matrix $\mathbf{R} \in \mathcal{R}^{N\times f}$, i-th column of which represents the weights of contribution from $\mathbf{X}$ in the reconstruction of the i-th dimension of $\mathbf{Z}$. Therefore, $r_{ji} = \sum_{t=1}^{k} \beta_{it} u_{jt} \gamma_t$ where r_{ji} denotes the j-th entry of the i-th column of $\mathbf{R}$. This matrix is considered as a rich encoded descriptor for dimensions of $\mathbf{Z}$ based on $\mathbf{X}$ and is used in Algorithm 1 to compare $\mathbf{Z}$ to the previously categorized unseen data in $\mathcal{H}$ to find the best place for $\mathbf{Z}$ in the dendrogram. Line 1 of the algorithm finds C_n as the most similar node to $\mathbf{Z}$ based on the distance term $d(\mathbf{Z}, C_n) = \|\mathbf{R}_{\mathbf{Z}} - \overline{\mathbf{R}}_{C_n}\|_F^2$, and

[1] https://github.com/bab-git/NQP

[2] https://github.com/bab-git/MKD_Unseen_MTS

Algorithm 1: Incremental Clustering of an Encoded MTS data

Input: $\mathbf{R}$: Encoding of the new unseen data $\mathbf{Z}$, $\mathcal{H}$: The current hierarchical tree.
Output: Place of $\mathbf{Z}$ in the hierarchy $\mathcal{H}$.

1 **If** $\exists C_n$ such that $d(\mathbf{Z}, C_n) \leq \bar{d}(C_n)$ **then**
2 **If** C_n is a leaf node **then** add $\mathbf{Z}$ to C_n;
3 **If** $(\bar{d}(C_{n1}) + \bar{d}(C_{n2}))/2\bar{d}(C_n) \leq k_{clust}$ **then**
4 split C_n into C_{n1} and C_{n2} using k-means;
5 **If** $(\bar{d}(C_{n1}) + \bar{d}(C_{n2}))/2\bar{d}(C_n) \leq k_{rmv}$ **then**
6 Replace C_n with C_{n1} and C_{n2};
7 **else** add $\{C_{n1}, C_{n2}\}$ as the children of C_n;
8 **else** Create a new child for C_n as C_{n_t} and add z to it;
9 **else** Create a new leaf at the top level containing $\mathbf{Z}$;

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

Table 1: Average of DRA measure (%) for reconstruction of the unseen classes.

	Cricket	CMU	Words	Squat
DRA (%)	76.4	84.5	80.2	62.6

the intra-cluster distance for each node C_n as $\bar{d}(C_n) = E_{\mathbf{Z}_i \in C_n}[d(\mathbf{R}_{\mathbf{Z}_i}, \overline{\mathbf{R}}_{C_n})]$, where $\overline{\mathbf{R}}_{C_n} = E_{\mathbf{Z}_i \in C_n}[\mathbf{R}_{\mathbf{Z}_i}]$. Regarding line 5, We choose $k_{rmv} = 0.3$ in our experiments which results in an acceptable clustering outcome.

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 [11]. For all the datasets, the Gaussian kernel matrices are computed as $\{\mathcal{K}_l(\mathbf{X}_i, \mathbf{X}_j) = exp(-\mathcal{D}_l(\mathbf{X}_i, \mathbf{X}_j)/\delta_l)\}_{l=1}^{f}$, where $\mathcal{D}_l(\mathbf{X}_i, \mathbf{X}_j)$ is the computed pairwise DTW-distance between the l-th dimension of $\mathbf{X}_i$ and $\mathbf{X}_j$ [11] (but can be substituted with any other preferred distance). For tuning T_0 and the dictionary size in Eq. 1, we use 5-fold cross-validation.

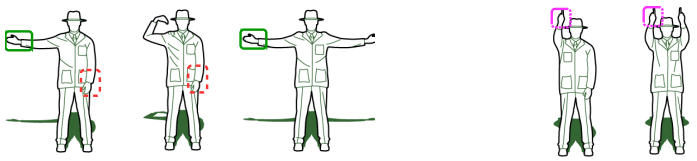
4.1 Partial Reconstruction Results

In order to evaluate the reconstruction quality for each unseen data $\mathbf{Z}$, we define the dimension-reconstruction accuracy measure as $DRA := \frac{\#\text{ dimensions that } \{\mathcal{J}^i_{rec}(\mathbf{Z},\mathbf{B},\mathbf{U}) \leq 0.1\}}{\#\text{ total dimensions}}$ using Eq. 3. Furthermore, each reconstructed dimension of $\mathbf{Z}$ which satisfies the above threshold is interpreted via the class of data with the most contribution as in Sec. 3.2. Table 1 reports the DRA values for the selected MTS datasets, where the CMU and Words datasets have higher DRA values due to their diverse set of training classes which increases the dimension-level similarity between seen and unseen classes. As an example, We illustrate the dimension-level reconstruction of 2 unseen categories from the Cricket dataset in Fig. 2, in which the *No ball* class is fully reconstructed via its relation to the movement of the left hand in the *Short* class and to that of the right hand in the *Wide* class.

4.2 Incremental Clustering Results

To evaluate the incremental clustering of Sec. 3.3 we use the average clustering error (CE) and normalized mutual information (NMI) [13]. As the most relevant baseline, we choose the self-learning algorithm [7] without its novelty detection part. Besides, we implement the spectral clustering algorithm on the original kernel matrix $\mathcal{K}(\mathbf{Z}, \mathcal{X})$ to compare our framework to the regular clustering of $\mathcal{Z}$. As another baseline, we also use the NNKSC algorithm [11] as the single-kernel predecessor of MKD-SC, for which the $\mathbf{R}$ matrix becomes an N-dimensional vector.

According to the clustering results in Table 2, the proposed MKD-SC method provides encodings which lead to better clustering of the unseen data compared to the



(a) *no ball* $\rightarrow \{short + \ wide\}$ (b) *out* $\rightarrow \{six\}$

Fig. 2: Dimension-level interpretation of *no-ball* and *out* (Cricket) based on the training classes. Related dimensions are specified via using same-color rectangles.

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

Table 2: Clustering error (CE) (%) and NMI the unseen categories.

Methods	Words		Squat		CMU		Cricket	
	CE	NMI	CE	NMI	CE	NMI	CE	NMI
MKD-SC(Proposed)	12.31	0.89	0	1	9.28	0.92	0	1
Self-learning [7]	18.75	0.84	0	1	14.25	0.87	16.63	0.85
NNKSC[11]	21.61	0.78	15.74	0.88	18.88	0.85	12.45	0.87
Spectral Clustering	27.51	0.76	13.04	0.90	23.45	0.76	8.04	0.89

baselines. The superiority of the spectral-clustering over NNKSC and self-learning methods (e.g., for Cricket dataset) depends on the discriminative quality of the original kernels. Self-learning method can have a better performance than NNKSC and spectral-clustering when its descriptor-based features can better discriminate between the different categories of the unseen classes.

5 Conclusion

In this research, we proposed an unsupervised framework which provides interpretable analysis of unseen classes in MTS datasets. It is constructed based on a novel MKD structure which uses the kernel representations of MTS dimensions to learn semantic attributes. Based on these attributes, our unsupervised MKD-SC framework reconstructs the unseen classes (partially/entirely) in the feature space according to the relation of their dimensions to those of the seen categories which provides an interpretable description of the novel data. Based on the obtained sparse encodings, we proposed an incremental clustering to categorize novel MTS into distinct clusters gradually. Experiments on real MTS benchmarks show the effectiveness of our MKD-SC framework in obtaining interpretable descriptions for unseen MTS classes. Additionally, the incremental clustering provides better clustering accuracy comparing to the baselines.

References

[1] Ibrahim Alabdulmohsin, Moustapha Cisse, and Xiangliang Zhang. Is attribute-based zero-shot learning an ill-posed strategy? In *ECML/PKDD'16*, pages 749–760. Springer, 2016.

[2] Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In *CVPR'09*, pages 951–958. IEEE, 2009.

[3] Richard Socher, Milind Ganjoo, Christopher D Manning, and Andrew Ng. Zero-shot learning through cross-modal transfer. In *Advances in neural information processing systems*, pages 935–943, 2013.

[4] Peixi Peng, Yonghong Tian, Tao Xiang, Yaowei Wang, Massimiliano Pontil, and Tiejun Huang. Joint semantic and latent attribute modelling for cross-class transfer learning. *TPAMI*, 40(7):1625–1638, 2018.

[5] Qiang Qiu, Zhuolin Jiang, and Rama Chellappa. Sparse dictionary-based representation and recognition of action attributes. In *ICCV'11*, pages 707–714. IEEE, 2011.

[6] Heng-Tze Cheng, Feng-Tso Sun, Martin Griss, Paul Davis, and Jianguo Li. Nuactiv: Recognizing unseen new activities using semantic attribute-based learning. In *MobiSys'13*, pages 361–374. ACM, 2013.

[7] Di Lu, Junqi Guo, and Xi Zhou. Self-learning based motion recognition using sensors embedded in a smartphone for mobile healthcare. In *WASA'16*, pages 343–355. Springer, 2016.

[8] Ron Rubinstein, Michael Zibulevsky, and Michael Elad. Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit. *Cs Technion*, 40(8):1–15, 2008.

[9] Ling Jian, Zhonghang Xia, Xijun Liang, and Chuanhou Gao. Design of a multiple kernel learning algorithm for ls-svm by convex programming. *Neural Networks*, 24(5):476–483, 2011.

[10] Ziming Zhang and Venkatesh Saligrama. Zero-shot learning via semantic similarity embedding. In *Proceedings of the IEEE international conference on computer vision*, pages 4166–4174, 2015.

[11] B. Hosseini, F. Hülsmann, M. Botsch, and B. Hammer. Non-negative kernel sparse coding for the analysis of motion data. In *ICANN 2016*, volume 9887, pages 506–514. Springer, 2016.

[12] B. Hosseini, F. Petitjean, Forestier G., and B. Hammer. Confident kernel dictionary learning for discriminative representation of multivariate time-series. *Journal of machine learning research (Under review)*.

[13] Wencheng Zhu, Jiwen Lu, and Jie Zhou. Nonlinear subspace clustering for image clustering. *Pattern Recognition Letters*, 107:131–136, 2018.