Non-negative kernel sparse coding for semantic representation of motion data

Babak Hosseini, Felix Hülsmann, Mario Botsch, and Barbara Hammer
CITEC centre of excellence, Bielefeld University

Universität Bielefeld

CITEC

The Key Question

How can we represent motion data (in general, multi-dimension time-series) more interpretable to make the application of high level approaches more efficient?

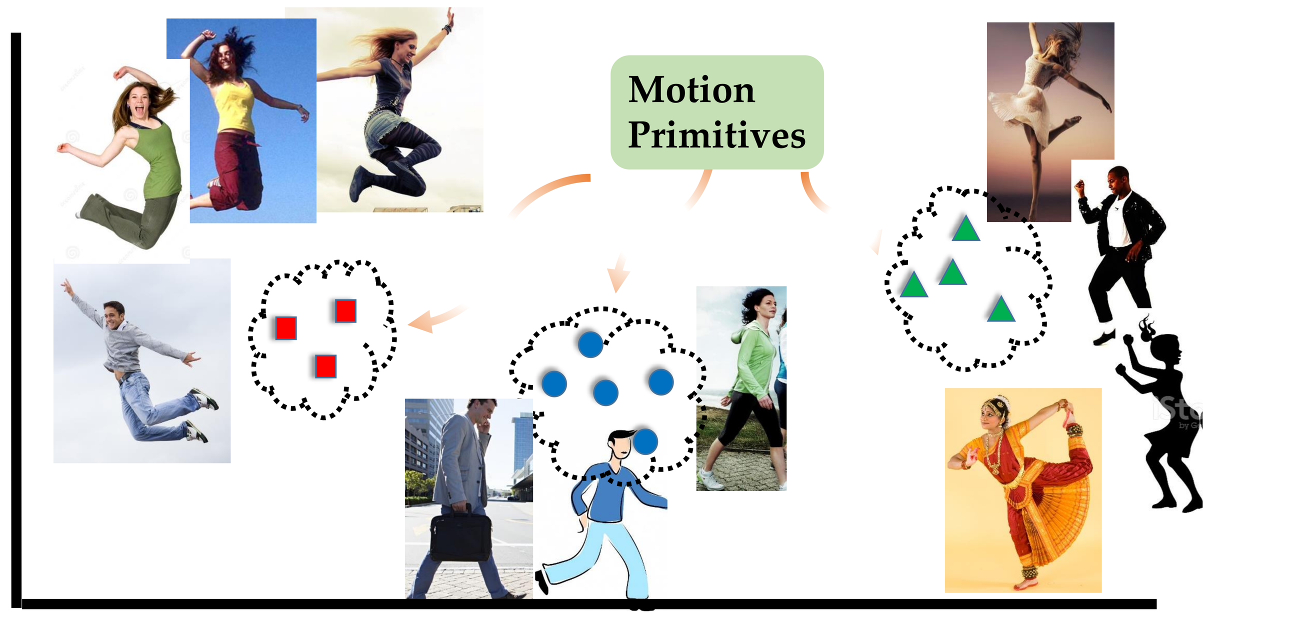
Motivation

A *semantic model* for the motion data can improve the efficiency and interpretability of high-level processing algorithms.

- e.g.: classification, clustering, search and etc.

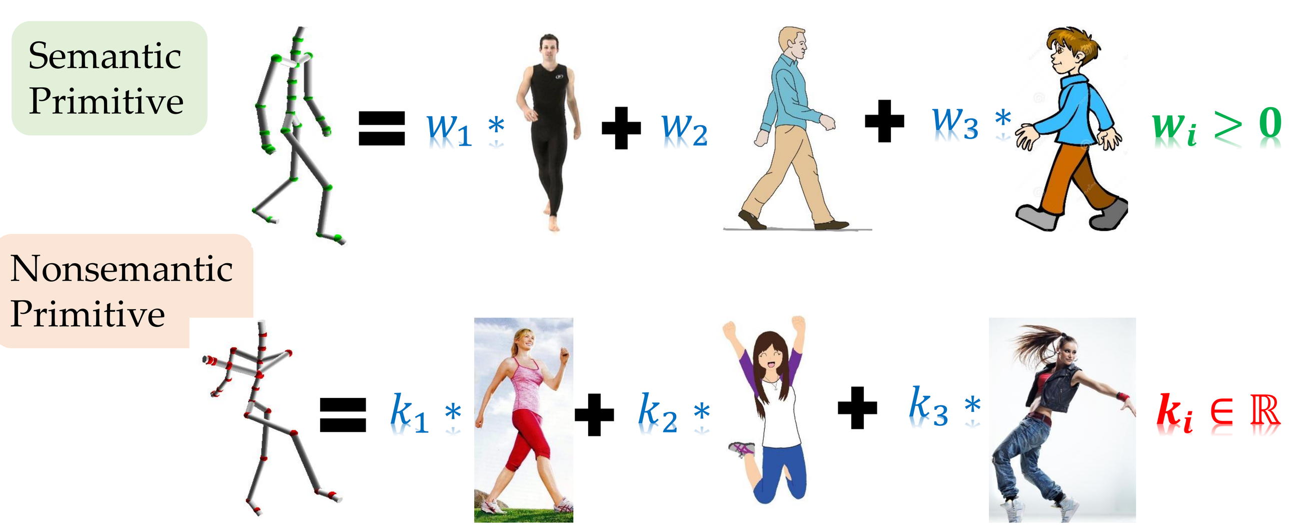
A *dictionary based model* can preserve the semantic information of motion data and make the representation more interpretable by:

- Being *Invariant* to temporal shift and scaling.
- Reconstructing data using motion primitives *semantically similar* to the data.
- Motion primitives being created from *similar data samples* to hold a semantic identity.
- Using *sparse* number of primitives for representation of data.



Hypothesis

- *Non-negative Sparse Coding* to model motion data:
  - → Positive linear combination can preserve the semantic information and provide semantic primitives.
  - → Sparse representation of data provides a compact model
- *Dynamic Time Warping (DTW)* as the distance measure:
  - → Similarities become invariant to temporal shifts and scaling.

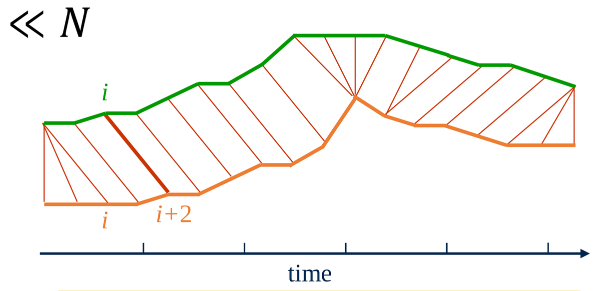


Non-negative Kernel Sparse Coding

Similarity kernel for motion data:

- Kernel function: $Y \in (\mathbb{R}^n)^* \Rightarrow \phi(Y) \in (\mathbb{R}^N)^*, n \ll N$
- Similarity(x, y) using DTW distance:

$$\mathcal{K}(DTW(x,y)) = exp(-\frac{DTW(x,y)^2}{\sigma})$$



Temporal alignment of two signals using Dynamic Time Warping (DTW)

Sparse coding optimization framework:

- A: Dictionary matrix
  → linear combination of exemplars in feature space.
- X: Sparse coding vector
  → linear combination of dictionary columns

$$\min_{X,A} \|\Phi(\mathbf{Y}) - \Phi(\mathbf{Y})\mathbf{AX}\|_F^2 + \boxed{\lambda\|A\|_1^2}$$
$$\text{s.t} \quad \|\boldsymbol{X}_i\|_0 \le T, \quad \boxed{a_{ij} \ge 0,} \quad \boxed{x_{ij} \ge 0} \, \forall i,j$$

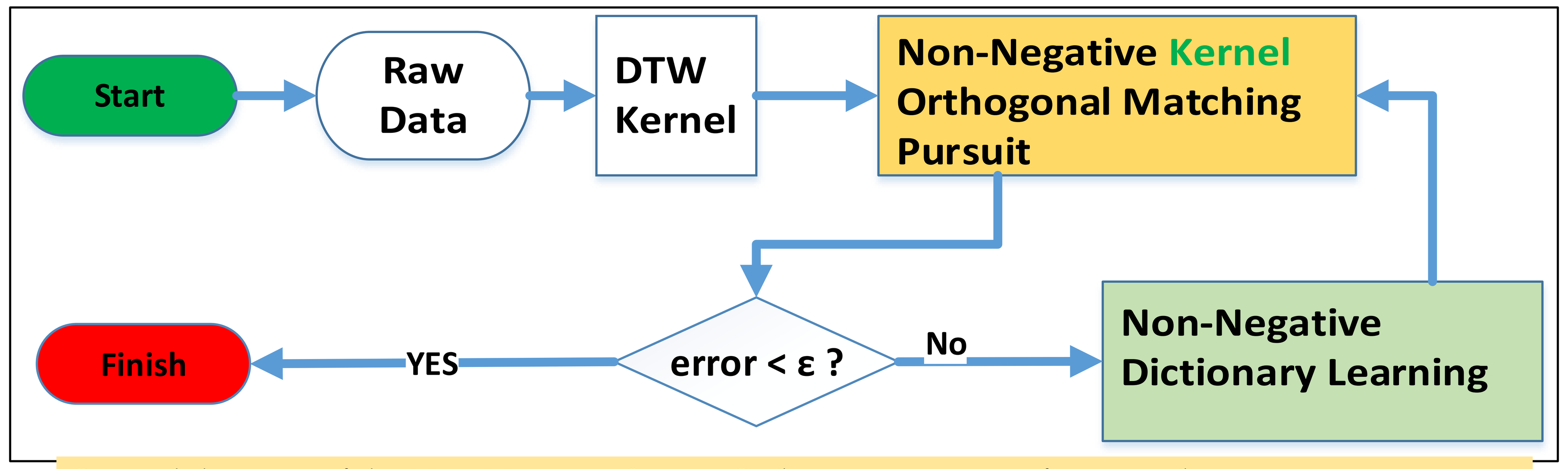
Alternating Optimization:

a) Finding best non-negative sparse x vector

$$\min_{X} \|\Phi(\mathbf{Y}) - \Phi(\mathbf{Y})\mathbf{A}\boxed{\mathbf{X}}\|_F^2$$
$$\text{s.t} \quad \|\boldsymbol{X}_i\|_0 \le T, \quad x_{ij} \ge 0 \, \forall i,j$$

$$\min_{A} \|\Phi(\mathbf{Y}) - \Phi(\mathbf{Y})\mathbf{AX}\|_F^2 + \boxed{\lambda\|A\|_1^2}$$
$$\text{s.t} \quad a_{ij} \ge 0, \quad \forall i,j$$

b) Finding best non-negative sparse A matrix



General diagram of the Non-Negative Sparse Coding optimization framework

Extension: Classification Framework

- Label-consistent sparse coding optimization framework:

$$\min_{X,A} \|\Phi(\mathbf{Y}) - \Phi(\mathbf{Y})\mathbf{AX}\|_F^2 + \boxed{\alpha\|Q - Q\mathbf{AX}\|_F^2 + \beta\|H - H\mathbf{AX}\|_F^2} + \lambda\|A\|_1^2$$
$$\text{s.t} \quad \|\boldsymbol{X}_i\|_0 \le T, \quad a_{ij} \ge 0, \quad x_{ij} \ge 0 \, \forall i,j.,$$

- H and Q are constructed from classification labels for training phase
- α and β are weights for classification accuracy and sparsity
- *Same optimization* algorithm, only using a new Kernel matrix:

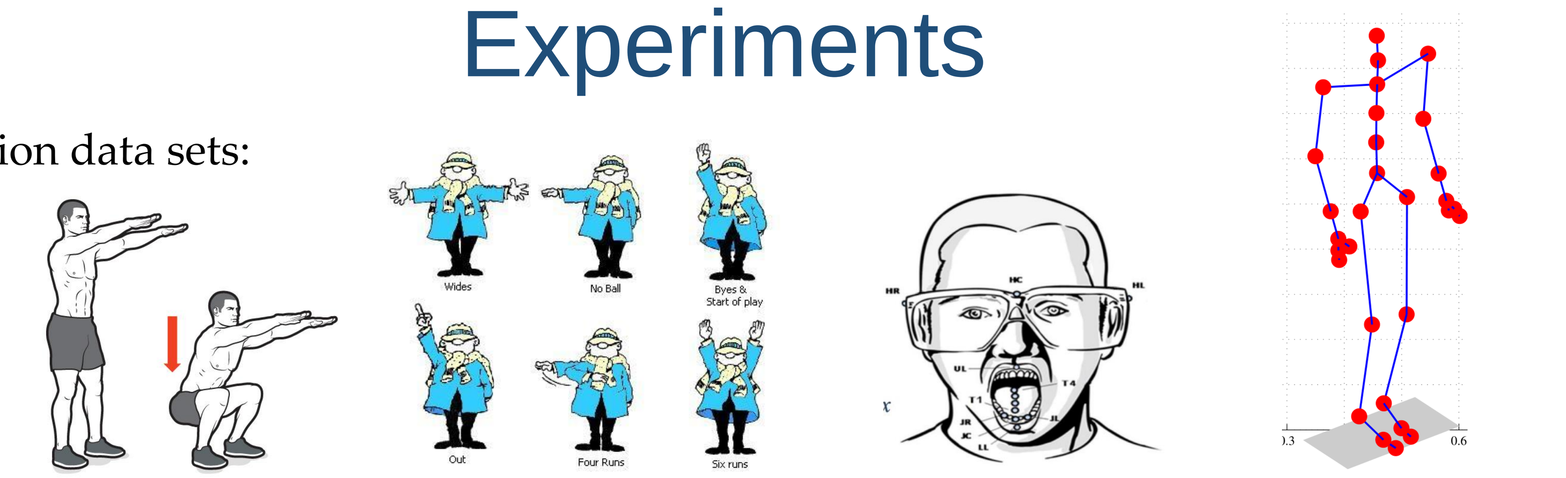
$$\boxed{\widetilde{\mathcal{K}}(\mathbf{Y}_i, \mathbf{Y}_i) = \mathcal{K}(\mathbf{Y}_i, \mathbf{Y}_j) + \alpha\langle \mathbf{Q}_i, \mathbf{Q}_j\rangle + \beta\langle \mathbf{H}_i, \mathbf{H}_j\rangle}$$

Acknowledgement

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).

Experiments

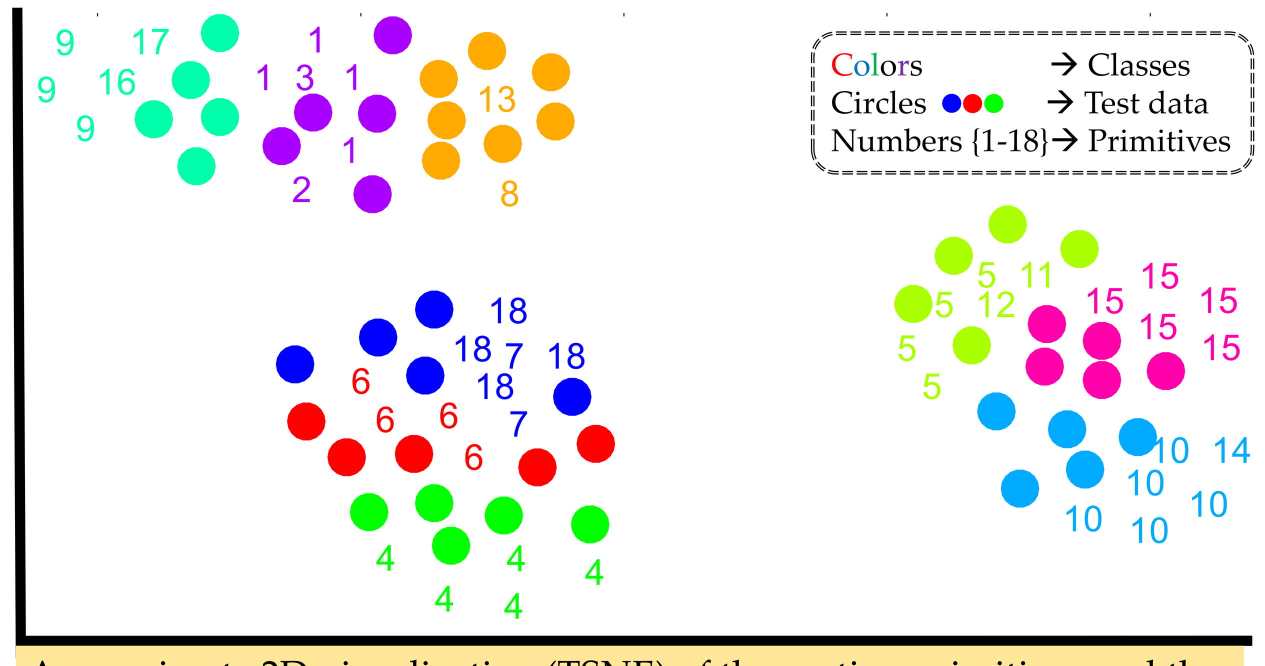
Motion data sets:



Squat Phases Cricket Signals Articulatory Words CMU locomotion

Notable results of proposed algorithm:

- Primitives: positive linear combination of *similar exemplars* in the feature space



Approximate 2D visualization (TSNE) of the motion primitives and the test data for Squat dataset

- *Motion Primitives* are almost using the same class of data (%).
- Data is reconstructed choosing *small number of primitives*.

	CMU				Cricket Signals				Articulatory Words				Squat dataset			
	bSP	wSP	bDS	wDS	bSP	wSP	bDS	wDS	bSP	wSP	bDS	wDS	bSP	wSP	bDS	wDS
LC-NNKSC	1	2	100	100	1	4	100	100	1	3	100	98.1	1	1	100	100
LC-KKSVD	5	9	100	76	5	13	100	44	5	16	100	56	3	8	100	87
Affinity P.	4	6	–	–	6	4	–	–	5	11	–	–	4	5	–	–
K-Means	4	17	100	50	5	27	100	16	5	50	100	50	4	12	100	60

Evaluation of the sparseness for non-negative sparse coding

- *Higher accuracy* when the classification is based on the proposed representation
- *Reconstruction error* is still in an acceptable range

	CMU		Cricket Signals		Articulatory Words		Squat	
	Acc	Rec. Err	Acc	Rec. Err	Acc	Rec. Err	Acc	Rec. Err
LC-NNKSC	90.91	4.17	83.33	11.07	97.33	14.52	100	0.14
LC-KKSVD	86.36	7.44	83.33	10.1	97.33	7.8	85	3.4
K-Means+SVM	68	–	56.25	–	90	–	81	–
Affinity P.	90.1	–	68.75	–	92	–	100	–
K-PCA+SVM	50	–	56.25	–	60.66	–	37	–
kNN	86.36	–	79.16	–	96.66	–	100	–

Performance of the classifier extension from the proposed framework

- Representations are semantically meaningful and easy to interpret.
- The outcome facilitates the application of higher level algorithms on the data.