



Multi-variate time series classification with specialized kernels

Imola Fodor SM3500474

21/09/2023



Introduction

- Unstructured data such as graphs, time-series, networks require specific systems to be modelled, if not summarized by expert's feature vectors.
- One technology that enables this are kernels, i.e. projecting the input into a space, where classification/regression can be performed, indirectly modelling raw data.
- In this work approaches to Multi-variate Time-series Classification(MTS) with template-extraction are presented. MTS can usually be viewed as consisting of multiple Uni-variate Time-series (UTS)



Agenda

- Overview of methods from research
- Approach 1. Extended Frobenius norm as Gram matrix
- Approach 2. Feature vector derived from Gram matrix
- Approach 3. Fisher Kernel as Gram Matrix
- Results



Overview of methods from research

1. Linear Time Warping Kernel - LTW
2. Dynamic Time Warping Kernel - DTW
3. SAX Kernels
4. Global Alignment Kernel -GA
 1. Better than above, but for long series better the LTW
5. Fisher Kernels (*Approach 3.*)
6. Extended Frobenius norm – Keros (*Approach 1.*)
7. Feature vector with Kernel + Manifold (*Approach 2.*)



Approach 1. Extended Frobenius norm as Gram matrix

The Extended Frobenius norm (Eros) computes the similarity between two matrices using the principal components (PCs).

- With PCs, directions where orthogonal projections of data points are maximized are identified.

It has been asserted that the Gram matrix is symmetric positive definite.

Using the dual form of an SVM Classifier, we can use the similarity matrix $N \times N$, where N is the number of series, as the Gram matrix, i.e. kernel matrix in the decision function.

In inference, we use the support vectors to construct the Kernel with the new test point.



Approach 1. Extended Frobenius norm as Gram matrix

Each cell of the Gram matrix, $K(x, x')$ is the Eros similarity for MTS series A and B with n sensors:

$$Eros(V_A, V_B, w) = \sum_{i=1}^n w_i | \langle a_i, b_i \rangle |_1 = \sum_{i=1}^n w_i |\cos \theta_i|$$

, where V_A and V_B are the two right eigenvector matrices obtained by applying SVD to the covariance matrices of A and B respectively. $V_A = [a_1, a_2, \dots, a_n]$, where a_i is a column orthonormal vector, idem for V_B .

The weight vector of D entries is unique for all $K(x, x')$ and it's calculated by:

1. Create S , an $n \times N$ matrix, where each column is the vector of eigenvalues for the given series.
2. Aggregate values in each row of S by averaging and construct the w vector from the means. Successively normalize the vector.



Approach 2. Feature vector derived from Gram matrix

1. Compute Gram matrix for each MTS (compare UTS-s)
2. Map matrix into tangent space of Riemannian manifold
3. Use vector as feature vector in SVM primal form Classification

For linear kernel:

$$K(x, x') = \sigma_b^2 + \sigma^2(x - c)(x' - c)$$

Or, e.g., for Gaussian kernel:

$$K(x, x') = \sigma^2 \exp\left(-\frac{\|x - x'\|^2}{2l^2}\right)$$

where σ , σ_b , l , c are the associated parameters.

Finally, the feature vector is:

$$z = \log(K) = M^{1/2} \log\left(M^{\frac{1}{2}} K M^{\frac{1}{2}}\right) M^{1/2}$$

, where M is the mean of the kernel matrices on the manifold M



Approach 3. Fisher Kernel as Gram Matrix

1. Generate the probabilistic model, the Gaussian Mixture Model (GMM), that explains the training data $P(X|\theta)$. Training data being all the n inputs of all the MTS of the set $\{X_1, X_2, \dots, X_N\}$.

Then the likelihood that a specific sequence X of length n has been generated by class c is $P(X|Class = c) = \prod_{i=1}^n p(x_i|Class = c)$.

2. Calculate $U_X = \Delta \log(P(X|\theta))$ w.r.t. to each parameter in θ .

Get a feature vector 6x1 parameters if the generative model is a GMM for two classes, $K = 2$. The three params in θ being prior probability for each mixture Gaussian, each component of the mean vector μ_l and the diagonal covariance matrix $\Sigma_{l,c}$.

3. Fisher Kernel

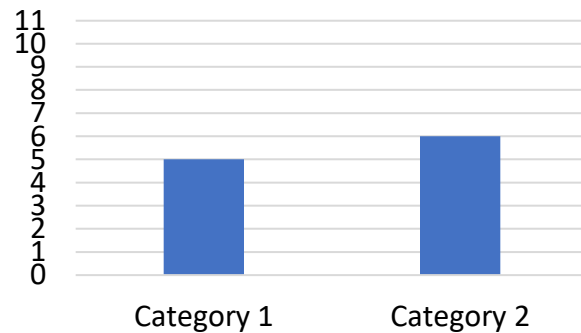
$$K(x, x') = U_X^T I U_{X'}$$

,where I is the Fisher information matrix.

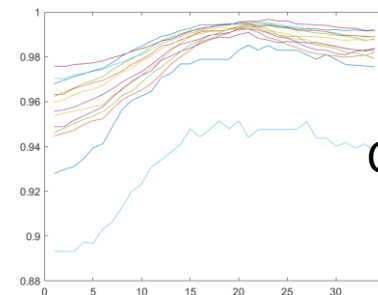
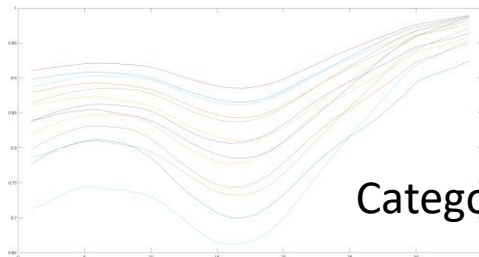


Data

Each datapoint (experiment) is a Multivariate Time-series acquired from $n = 13$ sensors. Each experiment is performed in one of the two conditions.



The raw data has been normalized by dividing each univariate series by its own max value. Further, a moving mean was performed.

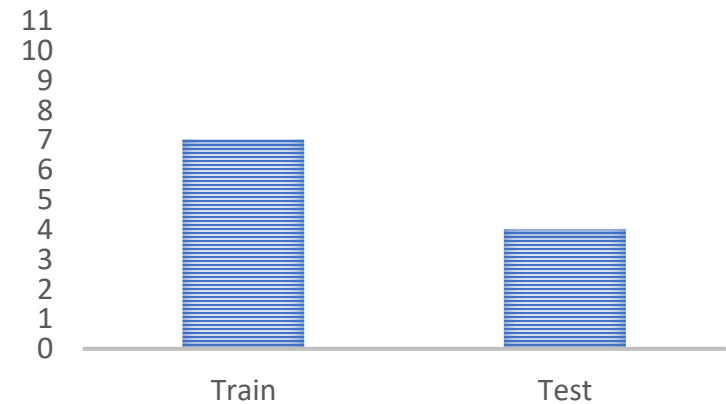




Results

Implementation can be found at <https://github.com/ImolaFodor/MTS-Classification-Kernels> , written in Matlab.

	1. Keros Kernel	2. Feature vector in M	3. Fisher Kernel
Correctly classified	0/4	4/4	0/4





Bibliography

Approach 1. A PCA-based Similarity Measure for Multivariate Time Series by Yang and Shahabi, 2004

Approach 2. Multivariate time series classification using kernel matrix by Jiancheng Sun, Huimin Niu, 2022

Approach 3. "USING THE FISHER KERNEL METHOD FOR WEB AUDIO CLASSIFICATION" by Pedro J. Moreno and Ryan Rifkin, 2000

Kernel Methods for Time Series Classification and Regression by Mourtadha Badiane et al.

Video lectures of Julien Mairal and collaborators.



Thank you for the attention!