Functional Kernel Hypothesis Testing for Channel Selection in Time Series Classification

Qi Huang, **Thomas Bäck** and Niki van Stein

LIACS, Leiden University, The Netherlands

Results

Four symbolic results obtained on four datasets from the UCR multivariate suite [1] are shown in Table 1,2,3 and 4.

	Range of channel relevance		
	(0.4,0.6]	(0.6,0.8]	(0.8,1.0]
F1	0.80	0.82	**0.83**
	0.85		0.83
	0.80	**0.88**	
	0.86		
N	2	2	2

Table 1. Results obtained on RacketSports

	Range of channel relevance			
	[0,0.1]	(0.1,0.2]	(0.2,0.3]	(0.3,0.4]
F1	0.39	0.45	0.46	0.48
	0.43		**0.55**	
	0.47			0.48
	0.39	0.52		
	0.52			
N	4	3	2	1

Table 2. Results obtained on HandMovementDirection

	Range of channel relevance						
	[0,0.15]	(0.15,0.3]	[0.3,0.45]	[0.45,0.6]	(0.6,0.75]	(0.75,0.9]	(0.9,1.0]
F1	0.70	0.71	0.70	0.73	0.76	0.69	0.76
	0.70	**0.77**					
	0.71		**0.77**				
	0.70			0.76			
	0.72				**0.77**		
	0.74					0.76	
	0.74						0.76
	0.77						
N	1	2	2	5	10	10	31

Table 3. Results obtained on Heartbeat

	Range of channel relevance						
	(0.3,0.4]	(0.4,0.5]	(0.5,0.6]	(0.6,0.7]	(0.7,0.8]	(0.8,0.9]	(0.9,1.0]
F1	0.69	0.67	0.81	0.78	0.75	0.60	0.88
	0.69	**0.91**					
	0.75		**0.91**				
	0.82			**0.91**			
	0.87				0.89		
	0.89					0.87	
	0.91						0.88
	0.91						
N	1	4	4	3	2	2	8

Table 4. Results obtained on NATOPS

"F1" means the (weighted) $f1$ scores (averaged over 10 independent runs) and N means the number of channels that falls into the relevance range (denoted at the top of each column). The darker the color of a cell is, the better the performance. The best results per data set are shown in **bold** font.

In Table 1, leaving out the two channels with *lower relevance*, the classifier yields better performance than using all channels.

In Table 2, only using the top 3 relevant channels outperforms using all 10 channels.

In Table 3, leaving out 10 channels the classifier is still performing as good as including all channels.

In Table 4, leaving out the 8 most relevant channels leads to the same $f1$ score as leaving out the 9 least relevant channels. Looking at the channel relevance scores of the NATOPS benchmark, we can observe that actually all channels show some relevance and therefore this does make sense.

Time Series Channel Selection

Consider multivariate or multi-channel time series classification as the problem. This research aims to handle the following scenario:
Given a TSC task on time series T with M channels (sensors), find the most $P \leq M$ relevant channels for predicting the segment-wise labels L.

Methodology

A functional kernel two-sample test aims to test the distributional equivalence between two functions supported by the same vector space. The core idea is to use a reproducing kernel defined on functional space to construct and test the metric discrepancies between the mean embedding of two functions [2]. It is beneficial to use multiple kernels since each kernel is limited by its form. Using this functional kernel-based two-sample test we can identify relevant channels.

Different Kernels

	Standard	Cos	Sqr	Cov	FPCA	Total
HMD	1	**7**	0	0	2	10
HB	0	0	**61**	0	0	61
RS	0	0	**4**	2	0	6
NATOPS	1	1	**19**	3	0	24

Table 5. Number of times that a kernel obtains the highest relevance

Table 5 shows the number of times the relevance of a channel is determined by a certain type of kernel. The squaring function expansion is the most confident kernel on three out of four problems.

Conclusions

- A novel functional kernel-based two-sample test approach is proposed to determine the channel relevance in multivariate time series classification tasks.
- The two-sample test methodology, combining different static kernels, works very well in order to reduce the number of channels (and maintain or increase accuracy).
- Choosing the best static functional kernel as well as selecting hyper-parameters can however still pose a challenge.

References

[1] Hoang Anh Dau, Eamonn Keogh, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Ann Ratanamahatana, Yanping, Bing Hu, Nurjahan Begum, Anthony Bagnall, Abdullah Mueen, Gustavo Batista, and Hexagon-ML. The ucr time series classification archive, October 2018. https://www.cs.ucr.edu/~eamonn/time_series_data_2018/.

[2] George Wynne and Andrew B. Duncan. A Kernel Two-Sample Test for Functional Data. *Journal of Machine Learning Research*, 23(73):1–51, 2022.