A Wasserstein Subsequence Kernel for Time Series

Christian Bock, Matteo Togninalli, Elisabetta Ghisu, Thomas Gumbsch, Bastian Rieck, Karsten Borgwardt

Department of Biosystems Science and Engineering
Machine Learning and Computational Biology Group

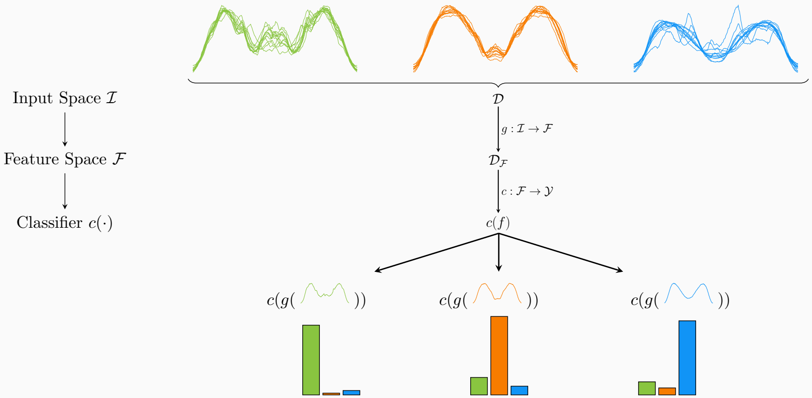
ETH Colors

ETHa ETHb ETHc ETHd ETHe ETHf ETHg ETHh ETHi

Table of contents (internal)

Motivation

Time Series Classification in a Nutshell

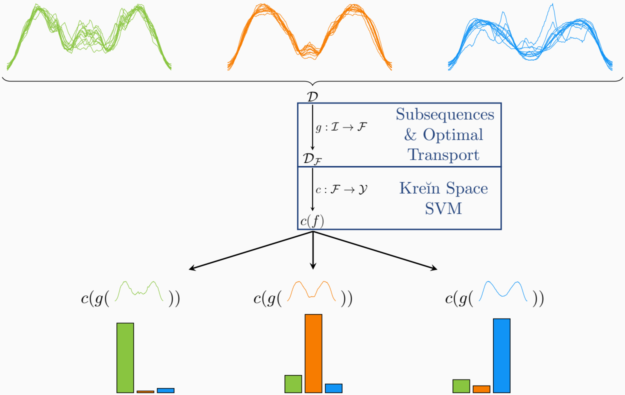


Time Series Classification in a Nutshell

Input Space $\mathcal{I}$

Feature Space $\mathcal{F}$

Classifier $c(\cdot)$

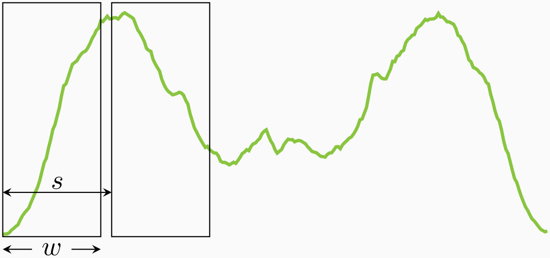


Feature Space: Subsequences and Optimal Transport

Subsequence Extraction

The Sliding Window Approach

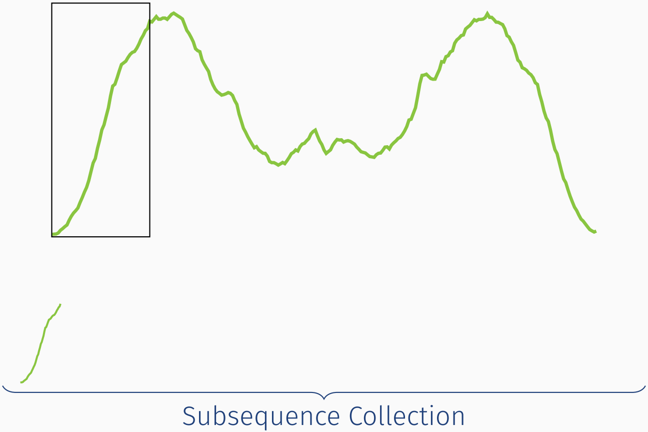
- Window Size w
- Stride s



Subsequence Extraction

The Sliding Window Approach

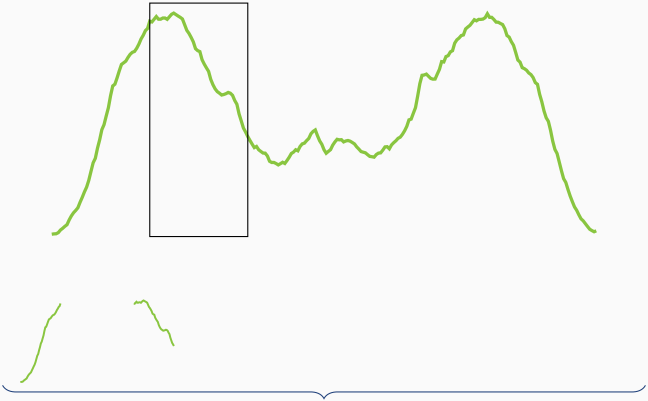
- Window Size w
- Stride s



Subsequence Extraction

The Sliding Window Approach

- Window Size w
- Stride s

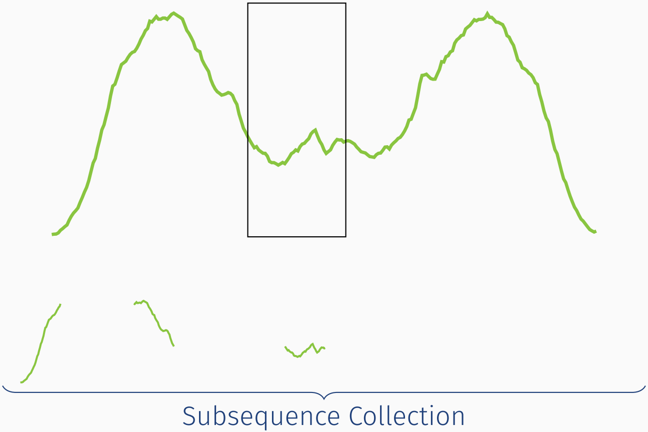


Subsequence Collection

Subsequence Extraction

The Sliding Window Approach

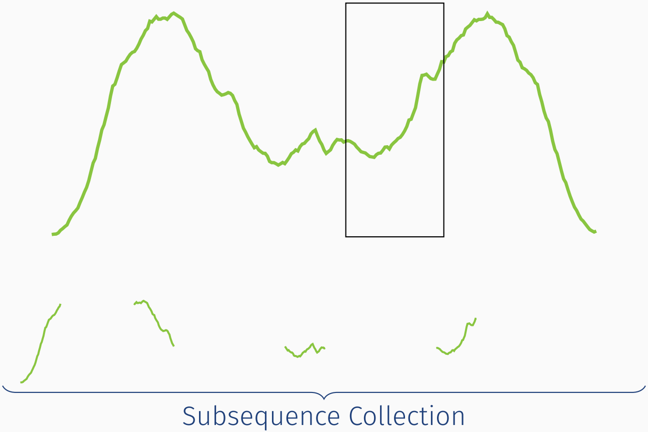
- Window Size w
- Stride s



Subsequence Extraction

The Sliding Window Approach

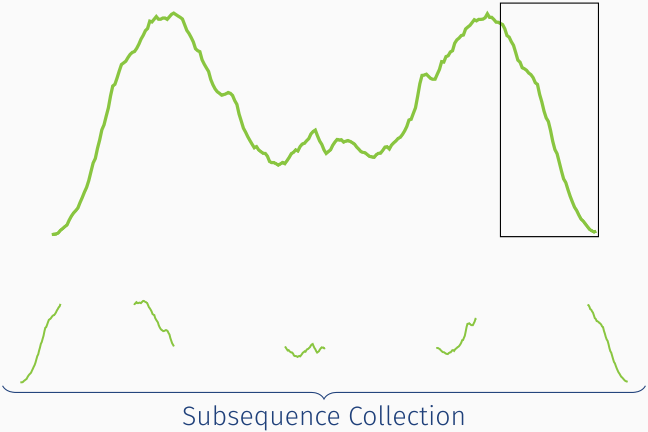
- Window Size w
- Stride s



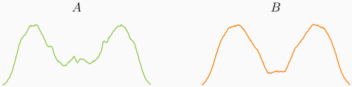
Subsequence Extraction

The Sliding Window Approach

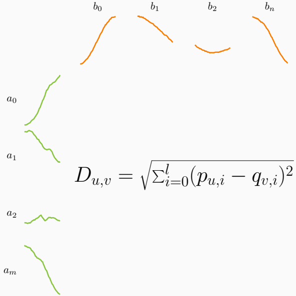
- Window Size w
- Stride s



Given Two Time Series



we Construct the Subsequence Distance Matrix D



$$D_{u,v} = \sqrt{\Sigma_{i=0}^{l} (p_{u,i} - q_{v,i})^2}$$

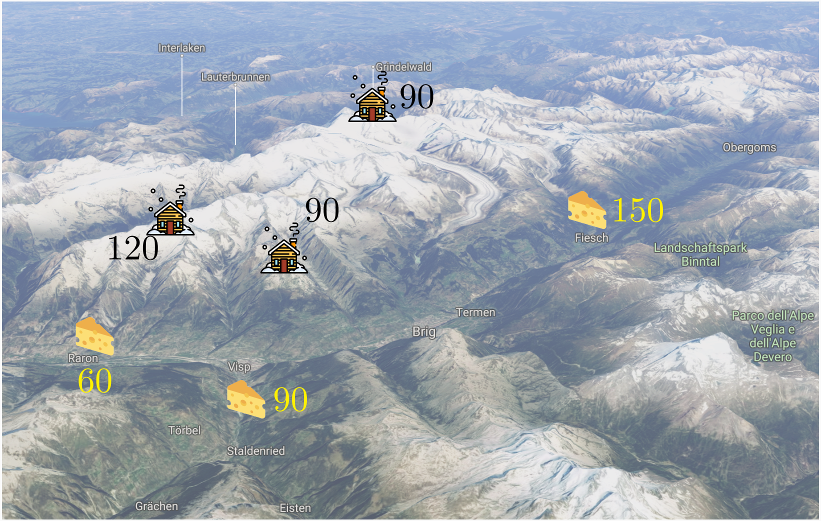
and obtain $\quad W_1(A, B) = \min_{P \in \Gamma(A,B)} \langle D, P \rangle_{\mathrm{F}}$

Optimal Transport and the Wasserstein Distance

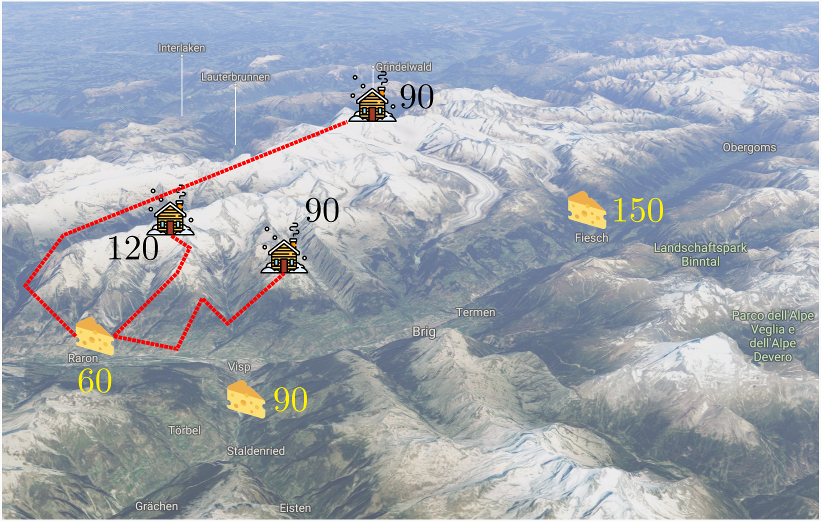
Transportation Theory

- Study of optimal transportation and allocation of resources
- Gaspard Monge (1746–1818)
  Morph one probability distribution into a second one with minmal cost.
- Frank Lauren Hitchcock (1875 – 1957)
  Transport a commodity from different sources to different “sinks”.
- Leonid Kantorovich (1912 - 1986)
  Distribute soldiers to the frontline.

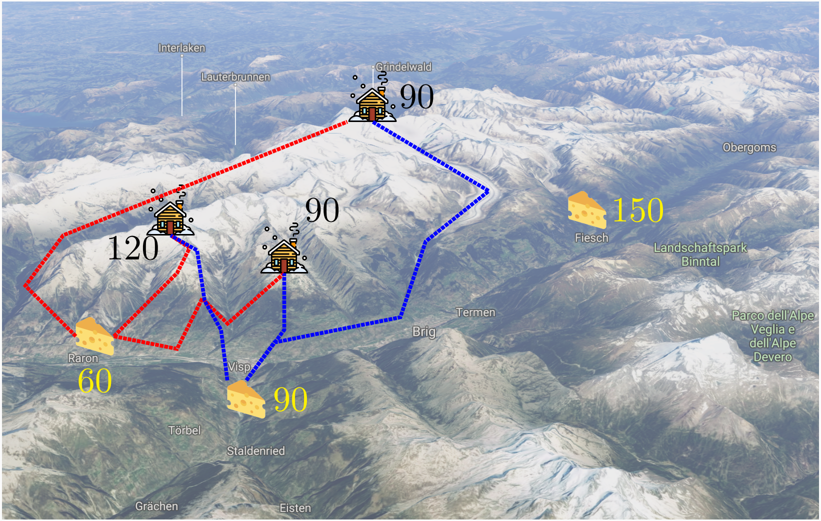
The Swiss Kantorovich Problem



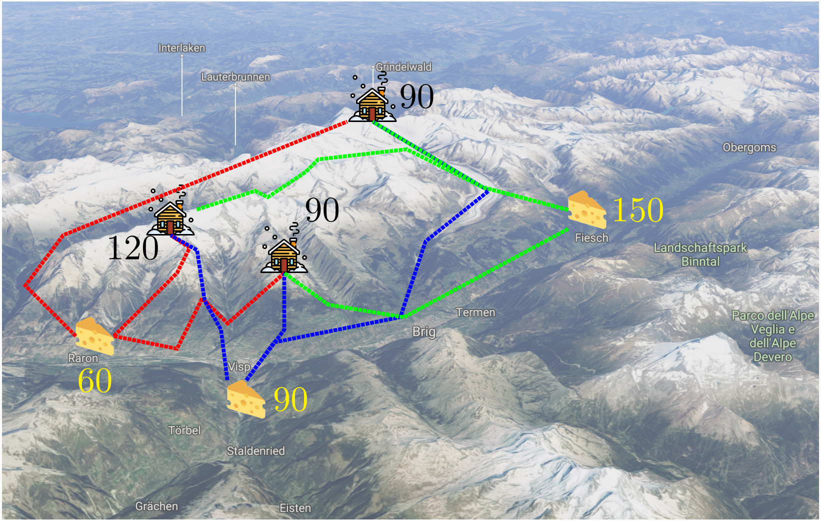
The Swiss Kantorovich Problem



The Swiss Kantorovich Problem



The Swiss Kantorovich Problem



The Swiss Kantorovich Problem

Distance Matrix D

	A	B	C
1	$D_{1,A}$	$D_{1,B}$	$D_{1,C}$
2	$D_{2,A}$	$D_{2,B}$	$D_{2,C}$
3	$D_{3,A}$	$D_{3,B}$	$D_{3,C}$

Transportation Matrix P

	120	90	90
60	?	?	?
90	?	?	?
150	?	?	?

The Swiss Kantorovich Problem

Distance Matrix D

	A	B	C
1	$D_{1,A}$	$D_{1,B}$	$D_{1,C}$
2	$D_{2,A}$	$D_{2,B}$	$D_{2,C}$
3	$D_{3,A}$	$D_{3,B}$	$D_{3,C}$

Transportation Matrix P

	120	90	90
60	?	?	?
90	?	?	?
150	?	?	?

Find P

	b_A	b_B	b_C
a_1	$P_{1,A}$	$P_{1,B}$	$P_{1,C}$
a_2	$P_{2,A}$	$P_{2,B}$	$P_{2,C}$
a_3	$P_{3,A}$	$P_{3,B}$	$P_{3,C}$

such that

$$\sum_{j \in \{A,B,C\}} P_{i,j} = a_i, \forall i \in \{1,2,3\}$$

$$\sum_{i \in \{1,2,3\}} P_{i,j} = b_j, \forall j \in \{A,B,C\}$$

The Swiss Kantorovich Problem

Distance Matrix D

	A	B	C
1	$D_{1,A}$	$D_{1,B}$	$D_{1,C}$
2	$D_{2,A}$	$D_{2,B}$	$D_{2,C}$
3	$D_{3,A}$	$D_{3,B}$	$D_{3,C}$

Transportation Matrix P

	120	90	90
60	?	?	?
90	?	?	?
150	?	?	?

Find P

	b_A	b_B	b_C
a_1	$P_{1,A}$	$P_{1,B}$	$P_{1,C}$
a_2	$P_{2,A}$	$P_{2,B}$	$P_{2,C}$
a_3	$P_{3,A}$	$P_{3,B}$	$P_{3,C}$

such that

$$\sum_{j \in \{A,B,C\}} P_{i,j} = a_i, \forall i \in \{1,2,3\}$$

$$\sum_{i \in \{1,2,3\}} P_{i,j} = b_j, \forall j \in \{A,B,C\}$$

that minimises the cost function

$$C(P) = \sum_i \sum_j D_{i,j} P_{i,j} = \langle D, P \rangle_{\mathbf{F}}$$

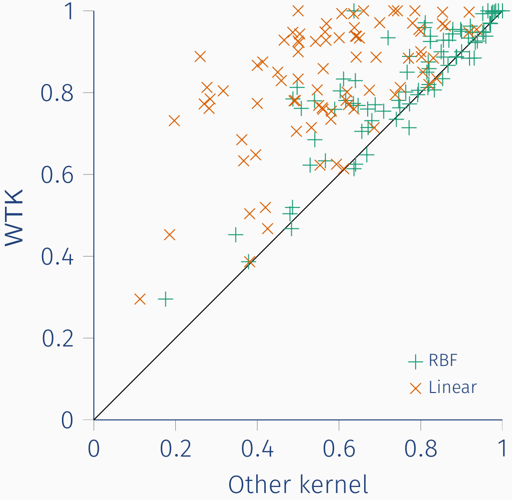
A Wasserstein Kernel for Time Series

Results

Experiments

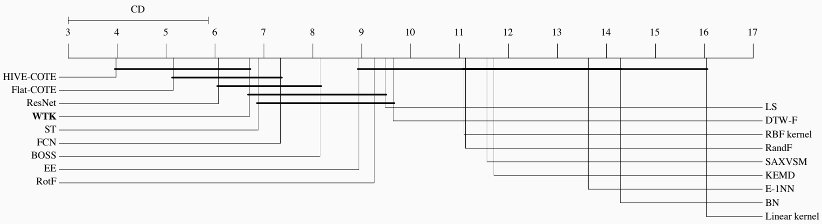
Datasets	*UCR Time Series Archive* 85 datasets predetermined train/test splits
Hyperparameters	Selected via a 5-fold cross-validation on the training set
Evaluation metric	Classification accuracy
Comparison partners	– Other kernels – DTW-1NN – State-of-the-art methods

Comparison with Other Kernels



Critical Difference Plot

The classification performances of methods sharing horizontal bars are not significantly different.



Take aways

Questions?

References