Feature selection using simple and efficient machine learning models.

Case studies and software tools.

Federico Amato, Fabian Guignard, and Mikhail Kanevski

Institute of Earth Surface Dynamics, Faculty of Geosciences and Environment, University of Lausanne, Switzerland. Corresponding Author: federico.amato@unil.ch

Introduction

The selection of relevant variables from the input space, known as Feature selection (FS), is crucial to reduce the dimensionality, remove irrelevant and redundant information, increase learning accuracy and improve the interpretability of the results. It is also useful to optimize data collection, as it identifies which kind of data are more important to gather.

FS shall be considered as a dynamic process integrated into the modelling procedure, which helps to reduce the prediction error and uncertainty.

GRNN Algorithm

A FS method based on General Regression Neural Network has been developed [1]. The general regression of a scalar y on a vector independent variable $\mathbf{x}$ is:

$$\hat{y}(\mathbf{x}) = \frac{\sum_{i=1}^{n} y_i \, exp\left(-D(\mathbf{x}, \mathbf{x}_i)\right)}{\sum_{i=1}^{n} exp\left(-D(\mathbf{x}, \mathbf{x}_i)\right)} \tag{1}$$

In Equation (1), $D(\mathbf{x}, \mathbf{x}_i)$ is the distance function:

$$D(\mathbf{x}, \mathbf{x}_i) = \sum_{j=1}^{p} \left(\frac{x_j - x_{ij}}{\sigma_j} \right)^2 \tag{2}$$

where p is the dimension of $\mathbf{x}$ and σ_j is the smoothing parameter for the j_{th} dimension.

The traditional GRNN architecture is based on the use of one unique value of σ for all the dimensions. This Isotropic structure of the network (IGRNN) can be used as a wrapper for feature selection by adopting an exhaustive search strategy (or a forward-backward one). Hence, it will be possible to identify relevant features as the ones included in the subset of the input space which is minimising the CV-MSE. This is a complete non-parametric approach.

Anisotropic (or Adaptive) GRNN (AGRNN) are an evolution of GRNN in which different values are given to each σ_j. A proper calibration of the σ_j will scale the input features depending on their explanatory power. Specifically, a large smoothing parameter will give rise to a lower discriminative power of the associated feature, and vice versa. Hence, AGRNN can be considered as an embedded feature selection method.

Feature Selection

Isotropic selector: a wrapper FS criterion based on IGRNN. This approach has the advantage of giving a complete description of the input space. A subset of feature will be labelled as:

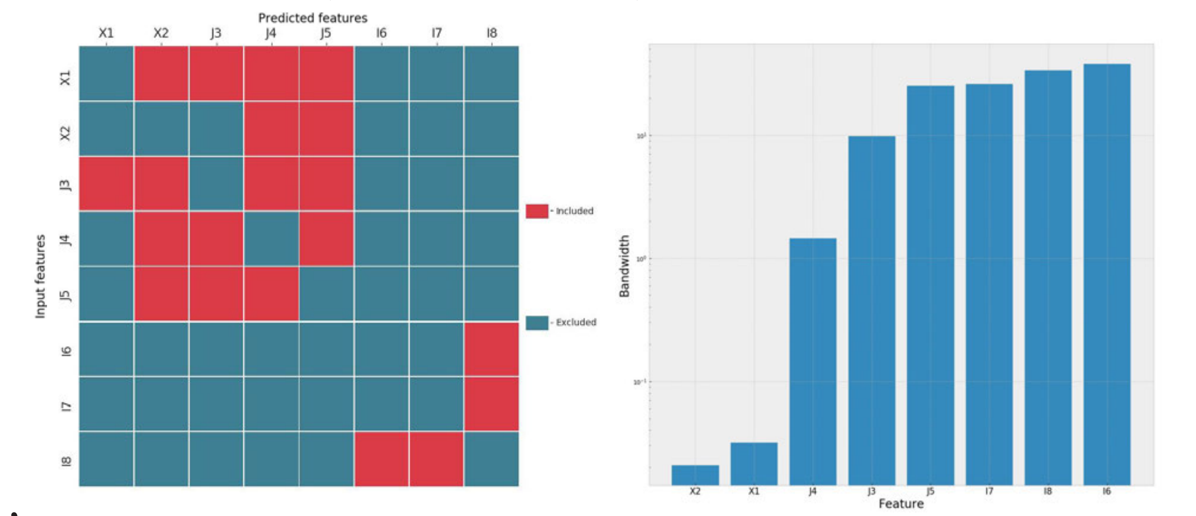
1. *relevant* if it is the one minimising the CV-MSE in an exhaustive search (or a forward-backward feature selection);
2. *redundant* if it is the one minimising the CV-MSE in an exhaustive search (or a forward-backward feature selection) when using as output the relevant features;
3. *irrelevant* if the features included in it are neither relevant nor redundant.

Redundancy and irrelevancy are associated to the identification of *relatedness*, i.e. the non-linear predictability of an input variable using the other features of the input space.

Anisotropic selector: an embedded FS criterion based on AGRNN. The bandwidth values of the kernel included in the algorithm express a measure of the relevance of the features.

Simulated Dataset

The *Butterfly* dataset was introduced in [2]. It is constituted by one target variable (Y) and eight features, of which two are relevant (namely X_1 and X_2), three are redundant (J_3, J_4 and J_5) and three are irrelevant (I_6, I_7 and I_8).



,

Figure 1: Relatidness matrix (left) and σ_j values found with AGRNN (right).

IGRNN			AGRNN
Relevant	**Irrelevant**	**MSE**	**MSE**
(X1, X2)	(I6, I7, I8)	0.001	0.001

Table 1: Relevant and redundant features after the Isotropic feature selection and MSE using the best model for both IGRNN and AGRNN.

Wind Speed Dataset

Monthly wind speed averages in Switzerland for the year 2008 have been collected from 118 stations belonging to the MeteoSwiss network.

A 21-d input space has been designed including the geographical space (latitude, longitude and elevation) and features derived by applying filters and derivatives on the digital elevation model.

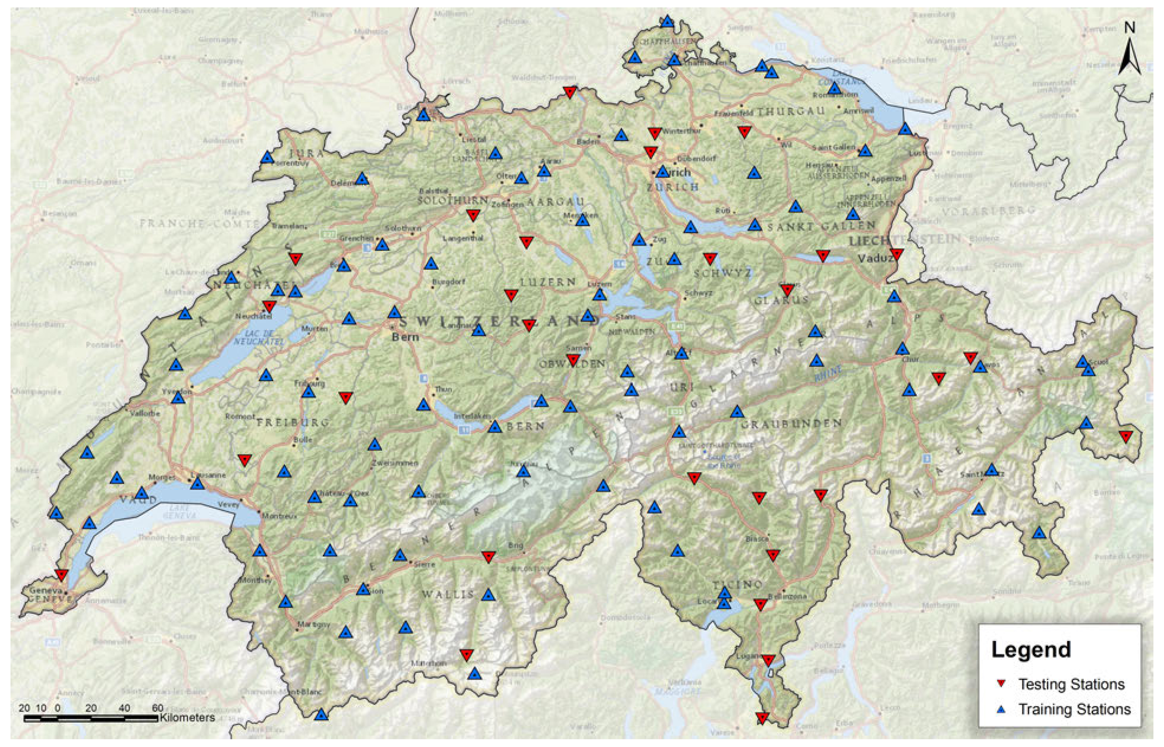


Figure 2: The 118 MeteoSwiss stations, divided into training and testing.

Month	IGRNN			AGRNN
	Relevant	**Irrelevant**	**MSE**	**MSE**
Jan	(4, 5, 6, 13, 15, 18, 19, 20, 21)	(10, 11, 12)	0.386	0.665
Feb	(3, 5, 6, 13, 15, 18, 21)	(10, 11)	0.372	1.343
Mar	(2, 3, 4, 5, 6, 13, 15, 16, 18, 20)	(10)	0.486	1.660
Apr	(3, 5, 14, 15, 17, 20, 21)	(10, 11)	0.343	0.775
May	(3, 4, 11, 15, 17, 20, 21)	(-)	0.225	0.775
Jun	(1, 3, 14, 15, 17, 21)	(10, 11)	0.270	0.447
Jul	(2, 4, 5, 9, 15, 16, 17)	(6, 10, 13, 14, 18 21)	0.280	0.463
Aug	(3, 5, 6, 14, 15, 17, 19, 21)	(10, 11)	0.254	0.531
Sep	(2, 3, 5, 6, 13, 15, 17, 18, 19, 21)	(10)	0.226	0.871
Oct	(5, 6, 13, 14, 15, 18, 19, 20, 21)	(10, 11, 12)	0.255	0.765
No	(5, 6, 13, 15, 18, 19, 21)	(10, 11, 12)	0.293	1.129
Dec	(5, 6, 13, 15, 18, 19, 21)	(10, 11, 12)	0.309	1.556

Table 2: Relevant and redundant features after the Isotropic feature selection and MSE for IGRNN and AGRNN (5-fold CV, 20 runs) for each month of 2008.

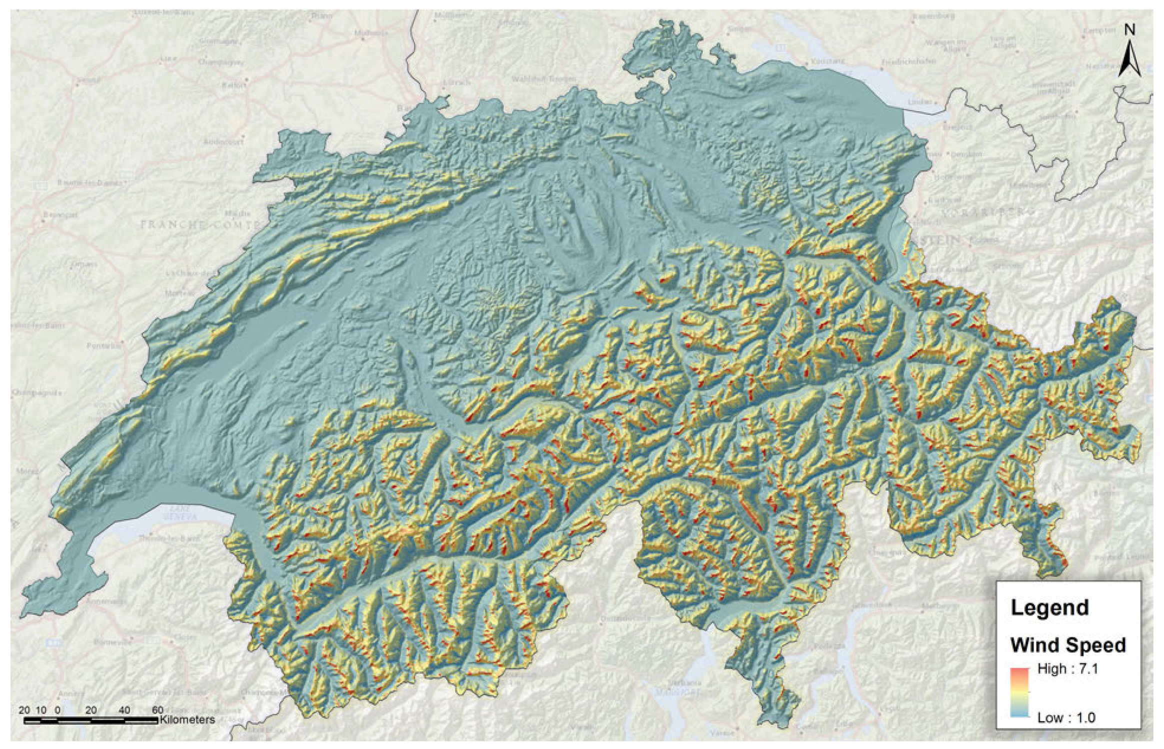


Figure 3: Wind speed prediction for April using the best subset of features obtained with the Isotropic selector.

Conclusions

The necessity of changing the way feature selection is traditionally intended within the Machine Learning workflow has been discussed. Despite being generally considered as part of pre-processing, several time-depending environmental phenomena require a dynamic FS. To support this statement two different GRNN-based approaches to FS, namely a wrapper with exhaustive search strategy and an embedded approach, have been applied. Two case studies were discussed, one based on a simulated dataset and one based on monthly means of wind speed in Switzerland. The results on the latter dataset highlighted the necessity of using different features in different time periods to obtain reliable prediction. The wrapper approach is not advisable in extremely high dimensional spaces. Anisotropic GRNN can therefore be extremely useful. However, further studies must be conducted to improve the interpretability of the bandwidth values of the AGRNN kernel as a measure of the relevance of the features, to understand the behaviour of the algorithm in high dimensional spaces and to scale it for application on big data.

Software availability

The GRNN model has been implemented in Python. It is fully integrated with Scikit-learn, the most used Python library for machine learning. It is included in the package *pyGRNN*, freely available on GitHub.

Acknowledgements

This research was partly supported by the National Research Programme 75 (PNR75) "Big Data" of the Swiss National Science Foundation.

References

[1] M. Kanevski, A. Pozdnoukhov, V. Timonin. Machine Learning for Spatial Environmental Data: theory, applications and software. EPFL Press, 2009

[2] J. Golay, M. Leuenberger, M. Kanevski, Feature selection for regression problems based on the Morisita estimator of intrinsic dimension, Pattern Recognition 70, 2017