UNIVERSITÄT
BIELEFELD
Technische Fakultät

Explaining incremental models

Workshop on Progressive Data Analysis at IEEE VIS 2024

Barbara Hammer
AG Machine Learning
Bielefeld University

Incremental machine learning

Batch machine learning

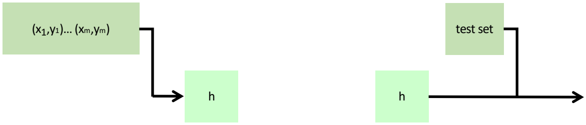
Given a **set of training data**

$\mathrm{D} = \{(x^1, y^1), \ldots, (x^m, y^m) \in X \times Y\}$

sampled w.r.t. a probability distributions P on $X \times Y$

We aim for **a model** $h : X \to Y$ such that the error on a test set $T \sim P$

$E = \sum_{(x,y) \in T} l(h(x), y)$ is minimized.



Incremental machine learning

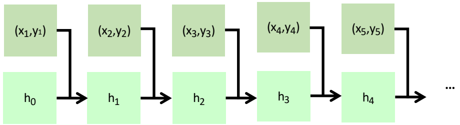
Given a **stream of training data**

$(x^1, y^1), \ldots, (x^t, y^t), \ldots \in X \times Y$

sampled w.r.t. a family of probability distributions P_t on $X \times Y$

We aim for a **learning scheme which incrementally adapts a model** $h_t : X \to Y$ based on (x^t, y^t) such that the interleaved train-test error

$E = \sum_t l(h_{t-1}(x_t), y_t)$ is minimized.

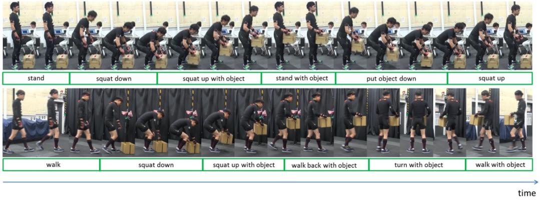


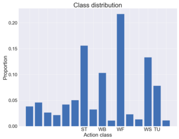
Example: Personalization of models

Data:

https://www.xsens.com

- 17 IMUs, 50 Hz, 6 interpolated sensors
- 4 subjects, 9 movements, 10-20 repetitions





Task:

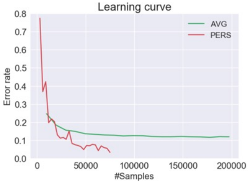
- predict based on current sensor values
- compare
  - individual online behavior
  - averaged model

Viktor Losing, Taizo Yoshikawa, Martina Hasenjäger, Barbara Hammer, Heiko Wersing: Personalized Online Learning of Whole-Body Motion Classes using Multiple Inertial Measurement Units. ICRA 2019: 9530-9536

UNIVERSITÄT BIELEFELD
Technische Fakultät

Example: Personalization of models

Average vs individual:



Error rate:

Feature set	#Dimensions AVG	#Dimensions PERS	AVG	PERS
Single frame	35	35	0.246	0.172
Stacked	1050	1050	0.190	0.148
DCT	175	175	0.179	0.142

Viktor Losing, Taizo Yoshikawa, Martina Hasenjäger, Barbara Hammer, Heiko Wersing: Personalized Online Learning of Whole-Body Motion Classes using Multiple Inertial Measurement Units. ICRA 2019: 9530-9536

UNIVERSITÄT BIELEFELD
Technische Fakultät

Summaries, models, and data

River

Losing, V., Hasenjäger, M. A Multi-Modal Gait Database of Natural Everyday-Walk in an Urban Environment. *Sci Data* **9**, 473 (2022). https://doi.org/10.1038/s41597-022-01580-3

- Jacob Montiel, Max Halford, Saulo Martiello Mastelini, Geoffrey Bolmier, Raphaël Sourty, Robin Vaysse, Adil Zouitine, Heitor Murilo Gomes, Jesse Read, Talel Abdessalem, Albert Bifet: River: machine learning for streaming data in Python. J. Mach. Learn. Res. 22: 110:1-110:8 (2021)
- Md. Mahbub Alam, Luís Torgo, Albert Bifet: A Survey on Spatio-temporal Data Analytics Systems. ACM Comput. Surv. 54(10s): 219:1-219:38 (2022)
- Viktor Losing, Barbara Hammer, Heiko Wersing: Incremental on-line learning: A review and comparison of state of the art algorithms. Neurocomputing 275: 1261-1274 (2018)
- Bartosz Krawczyk, Leandro L. Minku, João Gama, Jerzy Stefanowski, Michal Wozniak: Ensemble learning for data stream analysis: A survey. Inf. Fusion 37: 132-156 (2017)
- Gregory Ditzler, Manuel Roveri, Cesare Alippi, Robi Polikar: Learning in Nonstationary Environments: A Survey. IEEE Comput. Intell. Mag. 10(4): 12-25 (2015)
- …

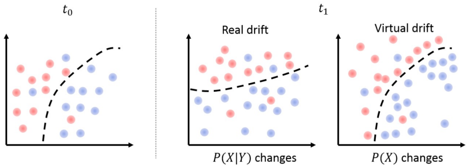
UNIVERSITÄT BIELEFELD
Technische Fakultät

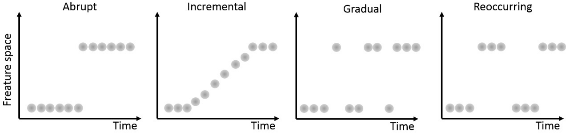
Drift

Drift

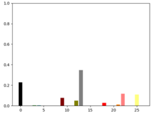
Drift is present if there exist time points $t_0 \neq t_1$ such that

$$P_{t_0} \neq P_{t_1}$$





Drift



Rialto data set, taken from
V. Losing, B. Hammer and H. Wersing, "KNN Classifier with Self Adjusting Memory for Heterogeneous Concept Drift," 2016 IEEE 16th International Conference on Data Mining (ICDM), Barcelona, Spain, 2016, pp. 291-300, doi: 10.1109/ICDM.2016.0040.

Challenges

- **Drift detection**: identify points in time where the underlying distribution changes (**when**?)
- **Drift localization**: identify regions in space where the difference of the distribution manifests itself (**where**?)
- **Drift explanation**: provide intuitive insight about the drift characteristics (**why**?)

→ **XAI for drifting scenarios**

Incremental feature importance

UNIVERSITÄT BIELEFELD
Technische Fakultät

Feature importance

Feature importance values:

Given an input space $X = X_1 \times \cdots \times X_n$,
given a model $f: X \to Y$,
given data $D \subseteq (X \times Y)^m$.

Find values $(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$
which represent the relevance of the features for the model f and data D

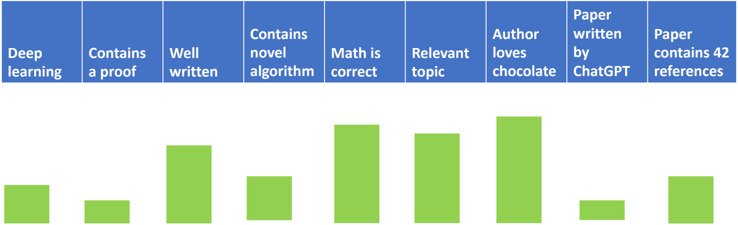
Surveys on properties:

[Degeest, Alexandra ; Frénay, Benoît ; Verleysen, Michel. *Reading grid for feature selection relevance criteria in regression.* In: *Pattern Recognition Letters*, Vol. 148, p. 92-99 (2021) http://hdl.handle.net/2078.1/250255 -- DOI : 10.1016/j.patrec.2021.04.031]
[A.Bommert, X.Sun, B.Bischl, J.Rahnenführer, M.Lang, Benchmark for filter methods for feature selection in high-dimensional classification data. Comput. Stat. Data Anal. 143 (2020)]

Which papers get accepted at a conference?

Deep learning	Contains a proof	Well written	Contains novel algorithm	Math is correct	Relevant topic	Author loves chocolate	Paper written by ChatGPT	Paper contains 42 references	
x				x	x			x	−
	x	x		x	x	x			+
x					x		x		−
	x	x	x	x	x	x			+
x	x	x			x		x	x	−
x		x							−
x		x		x	x	x			+
	x	x	x	x					−
		x		x	x	x			+
x		x		x	x	x		x	+
									−

Feature importance values



UNIVERSITÄT BIELEFELD
Technische Fakultät

Permutation feature importance (PFI)

Deep learning	Contains a proof	Well written	Contains novel algorithm	Math is correct	Relevant topic	Author loves chocolate	Paper written by ChatGPT	Paper contains 42 references	
x				x	x			x	−
	x	x		x	x	x			+
x					x		x		−
	x	x	x	x	x	x			+
x	x	x			x		x	x	−
x		x							−
x		x		x	x	x			+
	x	x	x	x					−
		x		x	x	x			+
x		x		x	x	x		x	+
									−

Permutation feature importance (PFI)

Deep learning	Contains a proof	Well written	Contains novel algorithm	Math is correct	Relevant topic	Author loves chocolate	Paper written by ChatGPT	Paper contains 42 references	
x		x		x	x			x	−
	x			x	x	x			+
x		x			x		x		−
	x		x	x	x	x			+
x	x	x			x		x	x	−
x									−
x		x		x	x	x			+
	x	x	x	x					−
				x	x	x			+
x		x		x	x	x		x	+
									−

Permutation feature importance (PFI)

Permutation feature importance:
Given an input space $X = X_1 \times \cdots \times X_n$,
given a model $h: X \to Y$,
given data $D \subseteq (X \times Y)^m$

Denote a permutation $\varphi: \{1, \dots, m\} \to \{1, \dots, m\}$.
Permutation feature importance of feature i is given as average over the change of loss when permuting the feature

$$\hat{\phi}_\varphi(i) \coloneqq \frac{1}{m} \sum_j \left| h\left(x_1^j, \dots, x_i^{\varphi(j)}, \dots, x_n^j\right) - y^j \right| - \left| h\left(x^j\right) - y^j \right|$$

[L. Breiman, “Random Forests”, Machine Learning, 45(1), 5-32, 2001]

UNIVERSITÄT BIELEFELD
Technische Fakultät

Incremental feature importance values

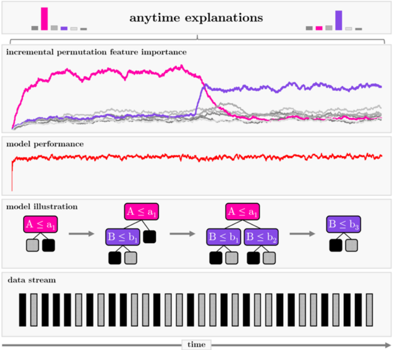
Incremental feature importance:

Given a stream of training data

$(x^1, y^1), \dots, (x^t, y^t), \dots \in X \times Y$

sampled w.r.t. P_t and incremental models $h_t : X \to Y$

For **every point in time *t***, find values $(\lambda_1^t, \dots, \lambda_n^t) \in \mathbb{R}^n$ which represent the relevance of the features for the model h_t and data sample (x^t, y^t)



Maximilian Muschalik, Fabian Fumagalli, Barbara Hammer, Eyke Hüllermeier:
Agnostic Explanation of Model Change based on Feature Importance. Künstliche Intell. 36(3): 211-224 (2022)

UNIVERSITÄT BIELEFELD
Technische Fakultät

Incremental PFI?

time	Deep learning	Contains a proof	Well written	Contains novel algorithm	Math is correct	Relevant topic	Author loves chocolate	Paper written by ChatGPT	Paper contains 42 references	
	x				x	x			x	−
		x	x		x	x	x			+
IEEE VIS	x					x		x		−
		x	x	x	x	x	x			+
	x	x	x			x		x	x	−
NeurIPS	x		x							−
	x		x		x	x	x			+
		x	x	x	x					−
ICML			x		x	x	x			+
	x		x		x	x	x		x	+
										−

Permutation feature importance (PFI)

PFI targets

$$\hat{\phi}_\varphi(i) \coloneqq \frac{1}{m}\sum_j \left| h\left(x_1^j, \dots, x_i^{\varphi(j)}, \dots, x_n^j\right) - y^j \right| - \left|h\left(x^j\right) - y^j\right|$$

Reliance is the increase in error, averaged over all instantiations of feature i

$$\phi(i) \coloneqq \sum_j \sum_{j' \neq j} \frac{\left|h\left(x_1^j, \dots, x_i^{j'}, \dots, x_n^j\right) - y^j\right|}{m(m-1)} - \sum_j \frac{\left|h\left(x^j\right) - y^j\right|}{m}$$

This can be seen as a **sampling strategy for the marginal distribution** of feature i .
It holds

$$\phi(i) = \frac{m}{m-1} \cdot E_{\varphi \sim unif(\mathfrak{S}(m))}\, \hat{\phi}_\varphi(i)$$

[Fisher, A.; Rudin, C.; and Dominici, F. 2019. All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *Journal of Machine Learning Research*, 20(177): 1–81.]
[Fabian Fumagalli, Maximilian Muschalik, Eyke Hüllermeier, Barbara Hammer: Incremental Permutation Feature Importance (iPFI): Towards Online Explanations on Data Streams. CoRR abs/2209.01939 (2022), Mach. Learn. 112(12): 4863-4903 (2023)]

Incremental PFI (iPFI)

Algorithm 1: iPFI explanation at time t for feature j

Require: : $\alpha \in (0, 1)$, sampling strategy φ_t, and $\hat{\phi}_{t-1}^{(S_j)}$.

procedure ExplainOne(h_t, x_t, y_t, j)
    $x_s \leftarrow \boxed{\text{Sample}(\varphi_t)}$
    $\hat{\lambda}_t^{(S_j)} \leftarrow \|h_t(x_t^{(\bar{S}_j)}, x_s^{(S_j)}) - y_t\| - \|h_t(x_t) - y_t\|$
    $\hat{\phi}_t^{(S_j)} \leftarrow (1-\alpha) \cdot \hat{\phi}_{t-1}^{(S_j)} + \alpha \cdot \hat{\lambda}_t^{(S_j)}$
    $\varphi_{t+1} \leftarrow \boxed{\text{UpdateSampler}(\varphi_t, x_t)}$
end procedure

run several estimations of feature relevance in parallel

draw one estimation for x_s from marginal distribution

take moving average of relevance estimation

[Fabian Fumagalli, Maximilian Muschalik, Eyke Hüllermeier, Barbara Hammer: Incremental Permutation Feature Importance (iPFI): Towards Online Explanations on Data Streams. CoRR abs/2209.01939 (2022), Mach. Learn. 112(12): 4863-4903 (2023)]
[Maximilian Muschalik, Fabian Fumagalli, Barbara Hammer, Eyke Hüllermeier: Agnostic Explanation of Model Change based on Feature Importance. Künstliche Intell. 36(3): 211-224 (2022)]

Incremental PFI (iPFI)

Algorithm 1: iPFI explanation at time t for feature j

Require: : $\alpha \in (0,1)$, sampling strategy φ_t, and $\hat{\phi}_{t-1}^{(S_j)}$.

procedure EXPLAINONE(h_t, x_t, y_t, j)

$\quad x_s \leftarrow \boxed{\text{Sample}(\varphi_t)}$

$\quad \hat{\lambda}_t^{(S_j)} \leftarrow \|h_t(x_t^{(\bar{S}_j)}, x_s^{(S_j)}) - y_t\| - \|h_t(x_t) - y_t\|$

$\quad \hat{\phi}_t^{(S_j)} \leftarrow (1-\alpha) \cdot \hat{\phi}_{t-1}^{(S_j)} + \alpha \cdot \hat{\lambda}_t^{(S_j)}$

$\quad \varphi_{t+1} \leftarrow \boxed{\text{UpdateSampler}(\varphi_t, x_t)}$

end procedure

Complete history / uniform sampling:

store all data: $x^1, x^2, \ldots$

store histogram

t

Recent history / geometric sampling:

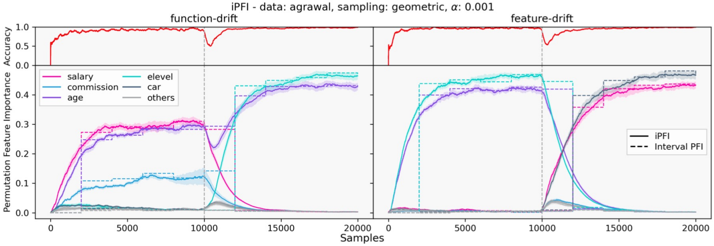
store L data: $x^{i_1}, x^{i_2}, \ldots, x^{i_L}$

substitutes one point, $p = \frac{1}{L}$

x^{t+1}

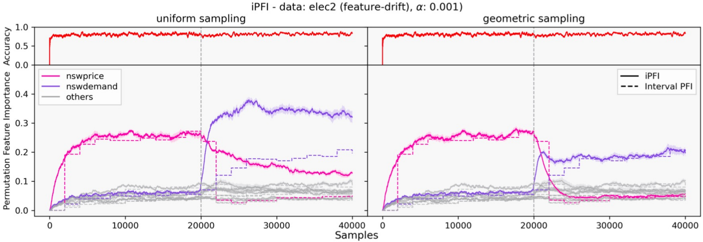
t

Incremental PFI (iPFI)



Agrawal data and ARF model: switch concept1 → concept2 (left), switch of features (right)

Incremental PFI (iPFI)



Electricity data: uniform (right) versus geometric sampling (right)

Consistency of iPFI

Theorem 2 (Bias for static Model). *If $h \equiv h_t$, then*

$$\phi^{(S_j)}(h) - \bar{\phi}_t^{(S_j)} = (1-\alpha)^{t-t_0+1}\phi^{(S_j)}(h).$$

Theorem 3 (Variance for static Model). *If $h_t \equiv h$ and $\mathbb{V}[\|h(X_s^{(\bar{S_j})}, X_r^{(S_j)}) - Y_s\| - \|h(X_s) - Y_s\|] < \infty$, then*

$$\textit{Uniform: } \mathbb{V}\left[\lim_{t\to\infty} \bar{\phi}_t^{(S_j)}\right] = \mathcal{O}(-\alpha\log(\alpha)).$$

$$\textit{Geometric: } \mathbb{V}\left[\lim_{t\to\infty} \bar{\phi}_t^{(S_j)}\right] = \mathcal{O}(\alpha) + \mathcal{O}(p).$$

[Fabian Fumagalli, Maximilian Muschalik, Eyke Hüllermeier, Barbara Hammer: Incremental Permutation Feature Importance (iPFI): Towards Online Explanations on Data Streams. CoRR abs/2209.01939 (2022), Mach. Learn. 112(12): 4863-4903 (2023)]

Incremental XAI toolbox

- compatible with RIVER
- contains
  - iPFI
  - incremental SAGE (Shapley values)

https://github.com/mmschlk/iXAI

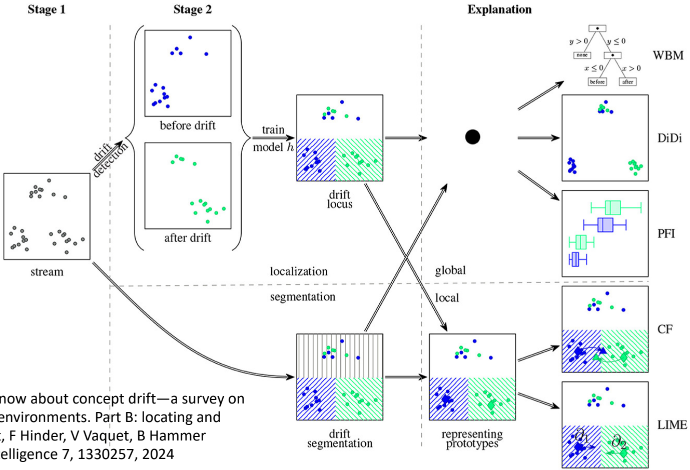
[Fabian Fumagalli, Maximilian Muschalik, Eyke Hüllermeier, Barbara Hammer: Incremental Permutation Feature Importance (iPFI): Towards Online Explanations on Data Streams. CoRR abs/2209.01939 (2022), Mach. Learn. 112(12): 4863-4903 (2023)]

[Maximilian Muschalik, Fabian Fumagalli, Barbara Hammer, Eyke Hüllermeier:
iSAGE: An Incremental Version of SAGE for Online Explanation on Data Streams. ECML/PKDD (3) 2023: 428-445]

UNIVERSITÄT
BIELEFELD
Technische Fakultät

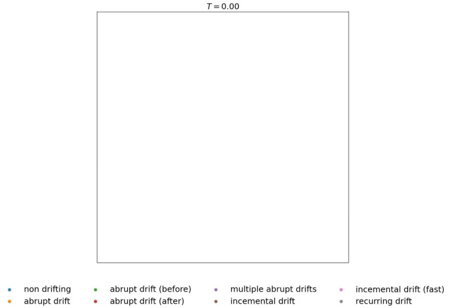
Model-based drift explanation

UNIVERSITÄT BIELEFELD
Technische Fakultät



One or two things we know about concept drift—a survey on monitoring in evolving environments. Part B: locating and explaining concept drift, F Hinder, V Vaquet, B Hammer Frontiers in Artificial Intelligence 7, 1330257, 2024

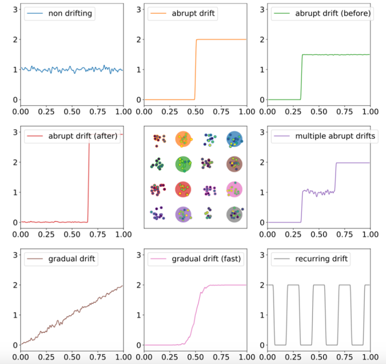
Drift segmentation



Hinder F., Hammer B.
Concept drift segmentation via Kolmogorov-trees
Verleysen M. (Ed.), ESANN (2021)

Drift segmentation

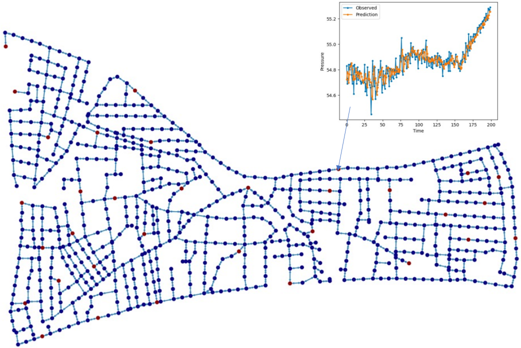
Algorithm:
(ensemble of) decision trees
splits into subsets l_1 and l_2
s.t. difference of $P(T|l_1)$ and $P(T|l_2)$ is maximum
e.g. using Kolmogorov Smirnov statistics



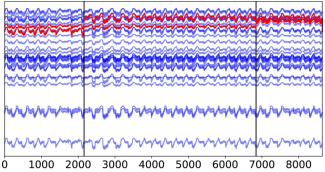
Hinder F., Hammer B.
Concept drift segmentation via Kolmogorov-trees
Verleysen M. (Ed.), ESANN (2021)

UNIVERSITÄT BIELEFELD
Technische Fakultät

Identification of sensor faults in WDS

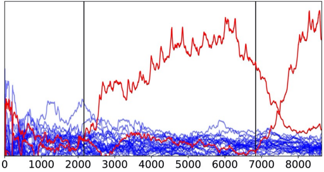


Identification of sensor faults in WDS



(a) Raw data

pressure values at
29 locations in the network

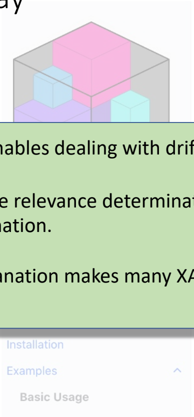


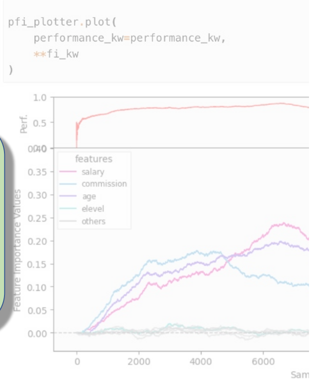
(b) Incremental permutation feature importance

Adaptive random forest for
drift segmentation
and iPFI for feature relevance determination

One or two things we know about concept drift—a survey on monitoring in evolving environments. Part B: locating and explaining concept drift, F Hinder, V Vaquet, B Hammer Frontiers in Artificial Intelligence 7, 1330257, 2024

Take away





Incremental learning enables dealing with drift.

Fast incremental feature relevance determination enables anytime explanation.

Model based drift explanation makes many XAI technologies available

UNIVERSITÄT BIELEFELD
Technische Fakultät

Thanks to ...

Fabian Fumagalli, Martina Hasenjäger, Fabian Hinder, Eyke Hüllermeier, Viktor Losing, Maximilian Muschalik, Valerie Vaquet, Heiko Wersing, Taizo Yoshikawa

SAIL SUSTAINABLE LIFE-CYCLE OF INTELLIGENT SOCIO-TECHNICAL SYSTEMS
TRR 318 CONSTRUCTING EXPLAINABILITY
KI-AKADEMIE.OWL
water futures
Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen
DFG
Bundesministerium für Bildung und Forschung
erc European Research Council Established by the European Commission