Transferring the Contamination Factor between Anomaly Detection Domains by Shape Similarity

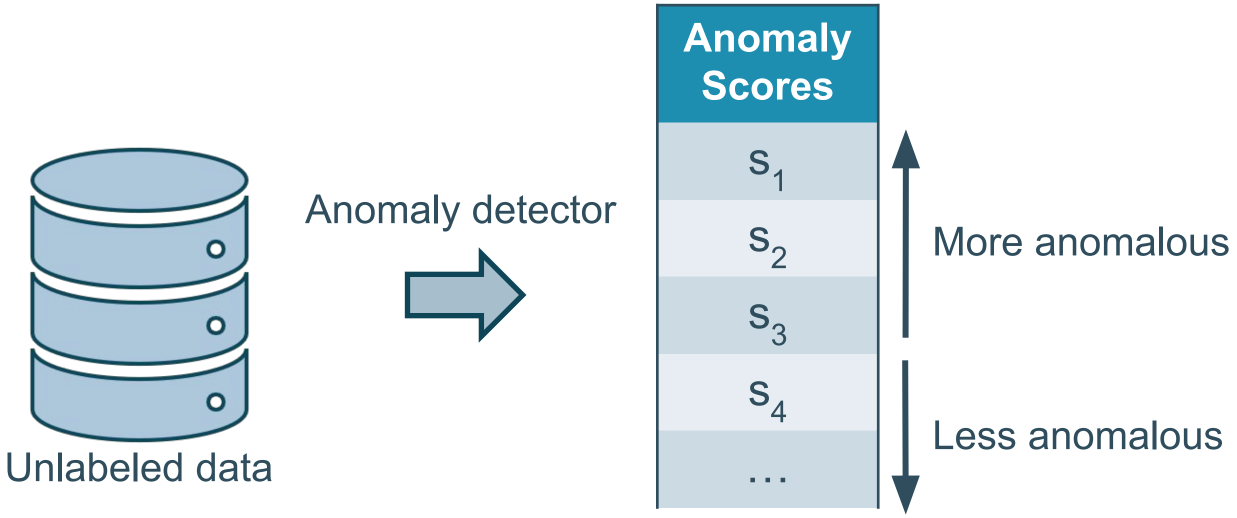
Lorenzo Perini, *Vincent Vercruyssen, Jesse Davis*

AAAI 2022

https://people.cs.kuleuven.be/~lorenzo.perini/

@LorenzoPerini95

Anomaly detectors compute anomaly scores indicating how anomalous examples are



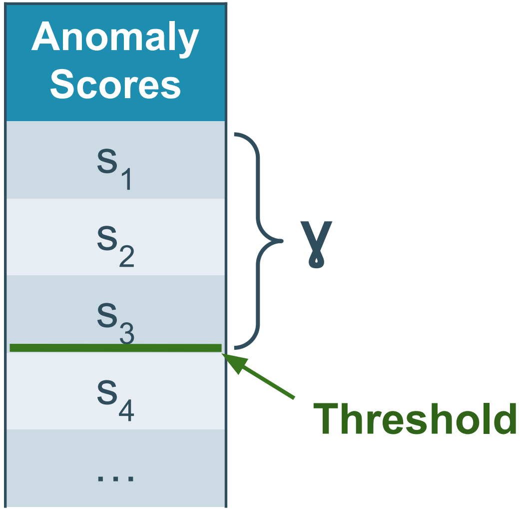
Anomaly detection is usually tackled from an unsupervised perspective, because ***labels are expensive or difficult to acquire***

Converting anomaly scores to hard predictions is challenging without labels

In practice, one needs the model’s hard predictions for decision making.

A common approach is to set the threshold based on the contamination factor ɣ

In practice, it is the user that has to provide such a contamination factor



Monitoring a fleet of assets makes the domain knowledge hard to acquire

- Many practical scenarios involve a fleet of assets
- Each asset has its own different contamination factor
- It is unfeasible that the user provides all the contamination factors

How to estimate the contamination factor with "limited" domain knowledge?


Wind Turbine 1


Wind Turbine 2


Wind Turbine 3


Wind turbine 4

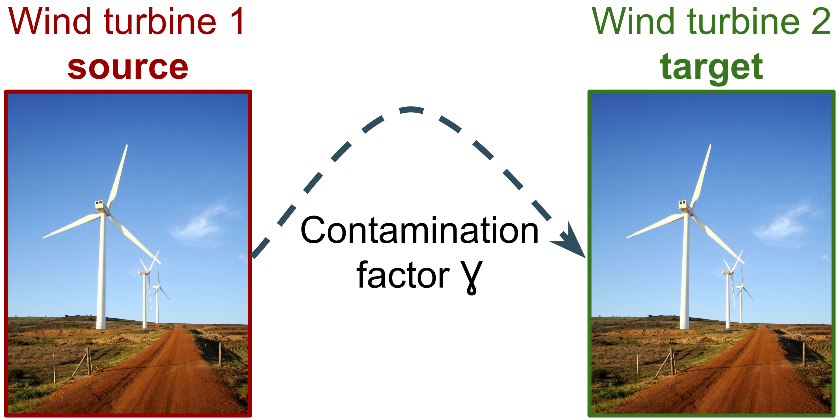
What we propose is to use Transfer Learning

GIVEN:

- ★ An anomaly detection model;
- ★ an unlabeled source domain with a *known contamination factor*;
- ★ an unlabeled target domain;

DO:

- ★ estimate the *contamination factor* of the target domain.



Task: transfer the contamination factor between related unlabeled anomaly detection domains

How can we transfer the contamination factor?

This Paper Makes Four Contributions:

1. Introduce the problem of transferring the contamination factor γ;

2. Propose a 4-step approach TrADe that:
   2.1. computes the anomaly scores in each domain separately using the same anomaly detection algorithm;
   2.2. sets the threshold on the source domain based on the given γ;
   2.3. estimates the target threshold by exploiting the domains similarity;
   2.4. derives the target contamination factor as the proportion of scores above the target threshold.

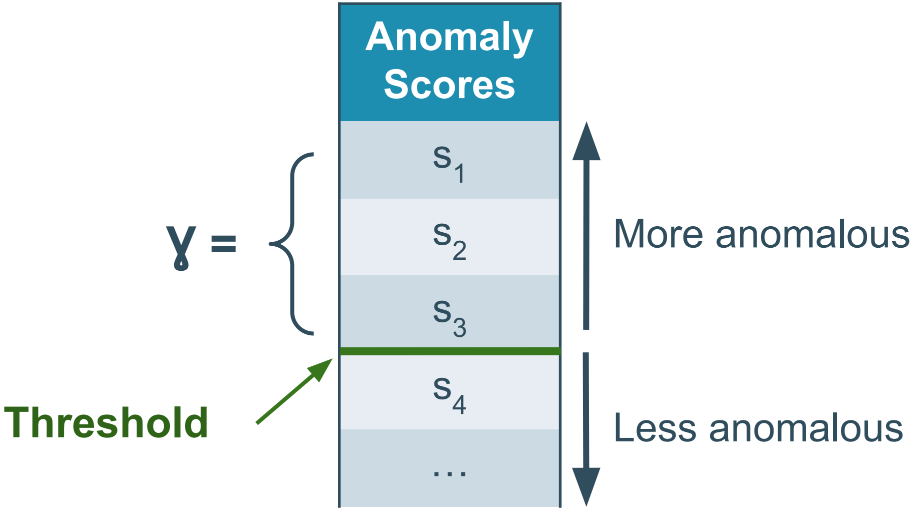
3. Analyze the convergence behaviour of TrADe when increasing the target size;

4. Perform an extensive empirical evaluation on benchmark and real world data.

Why is the contamination factor relevant in anomaly detection?

The contamination factor ɣ allows to set a threshold

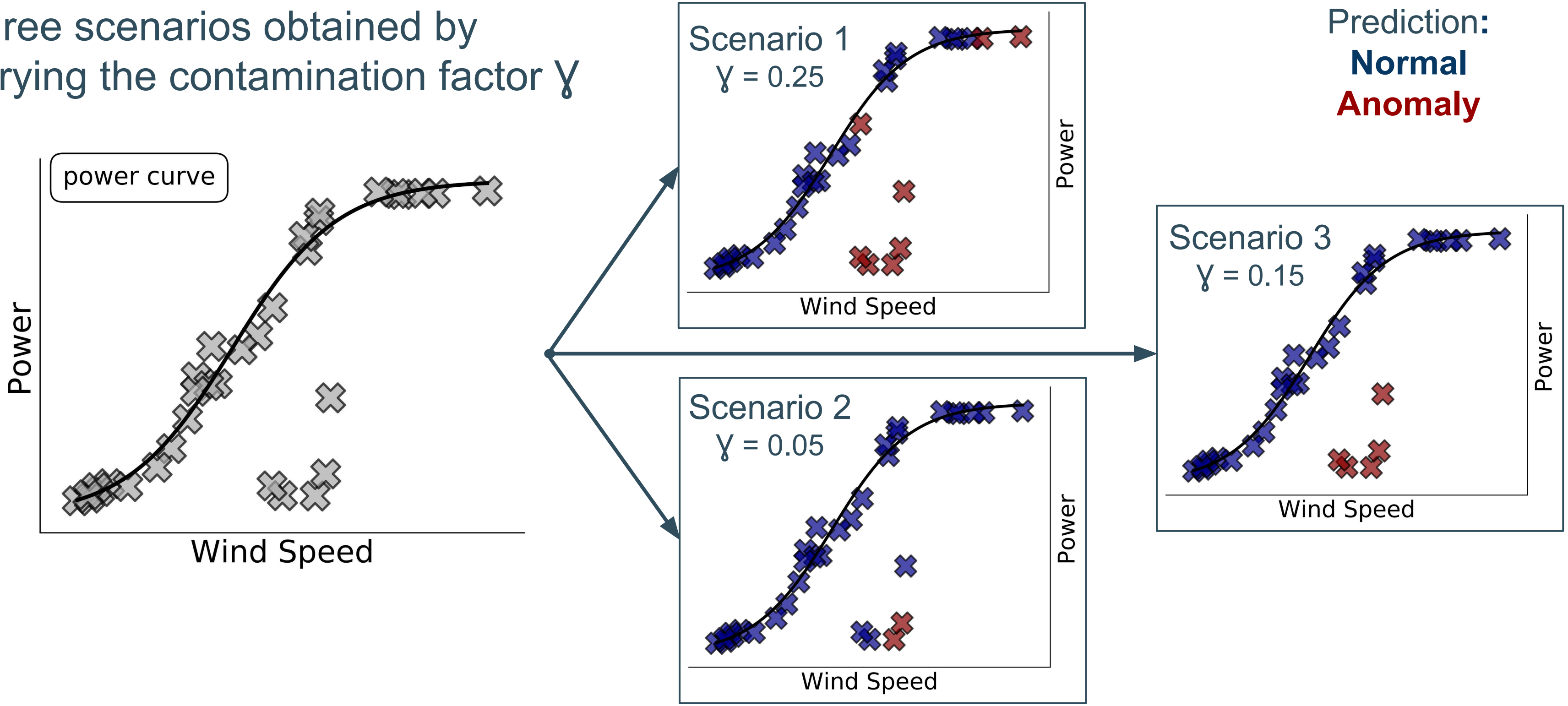
The contamination factor **ɣ** is the *expected proportion* of anomalies in the data.



What's the impact of the contamination factor on the model predictions?

ɣ has a relevant impact on the model performance

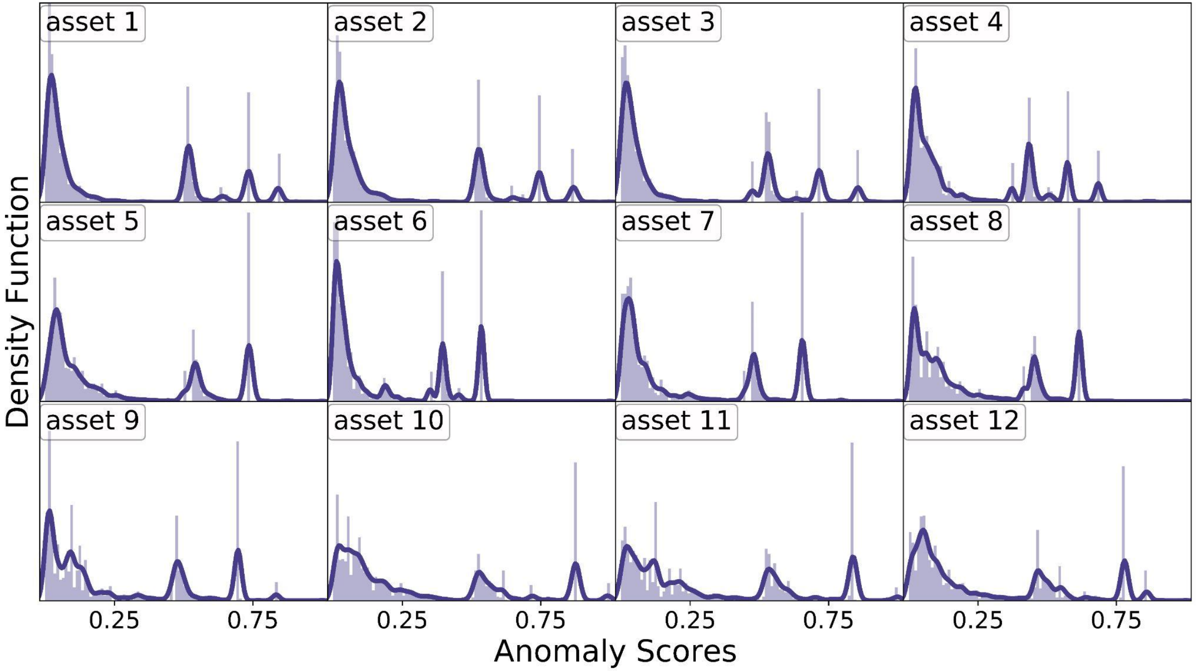
Three scenarios obtained by varying the contamination factor ɣ



Prediction:
Normal
Anomaly

TrADe: Transferring the contamination factor between Anomaly Detection domains by shape similarity

Assumption: normal anomaly scores produced by the same algorithm are similarly distributed



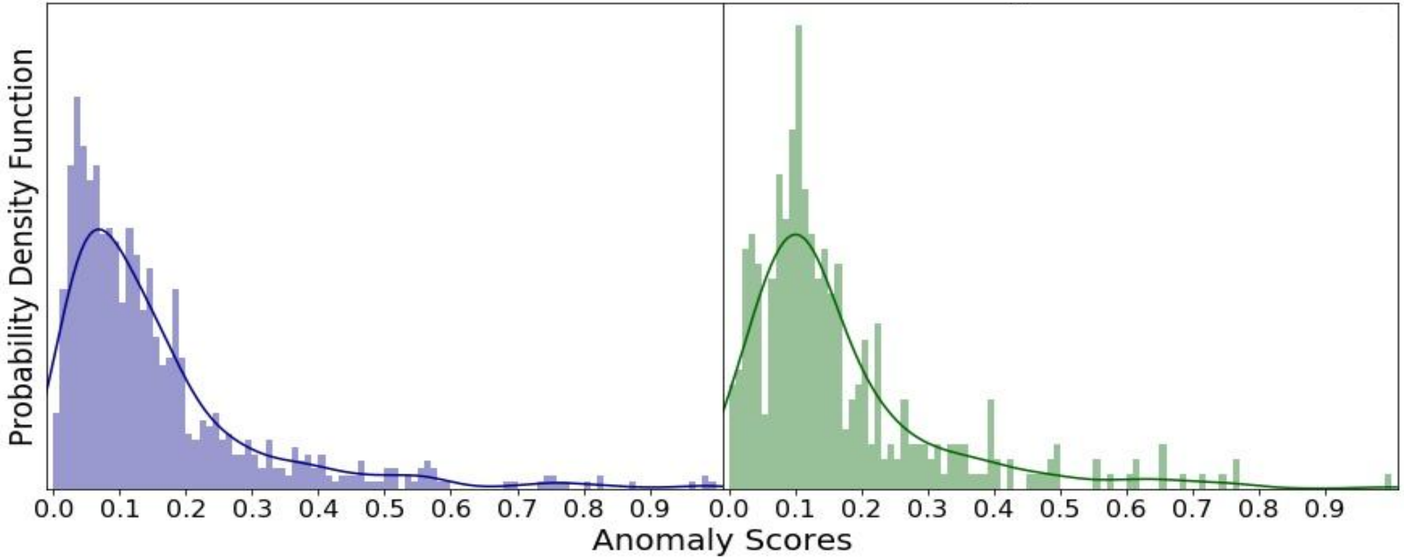
- 12 proprietary ds
- Same anomaly detection algorithm

Step 1. Use the same algorithm to compute the anomaly scores of the two domains

Compute the **distribution of the anomaly scores** of the two separate domains:

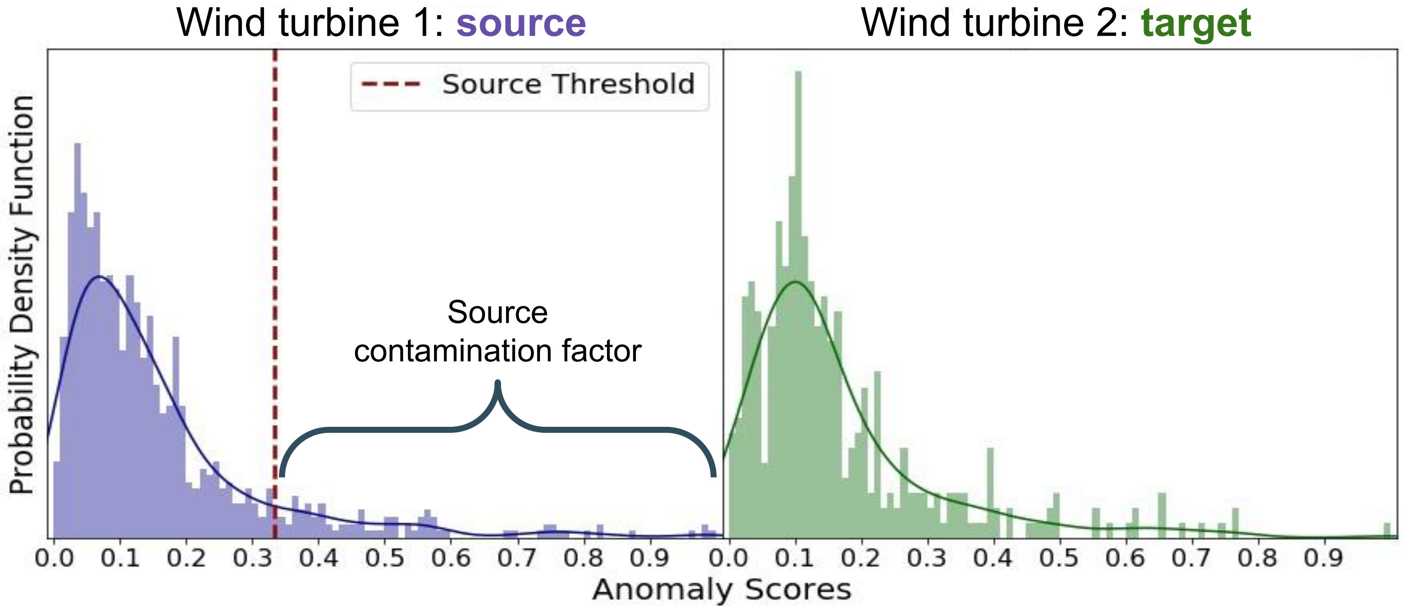
Wind turbine 1: **source**

Wind turbine 2: **target**



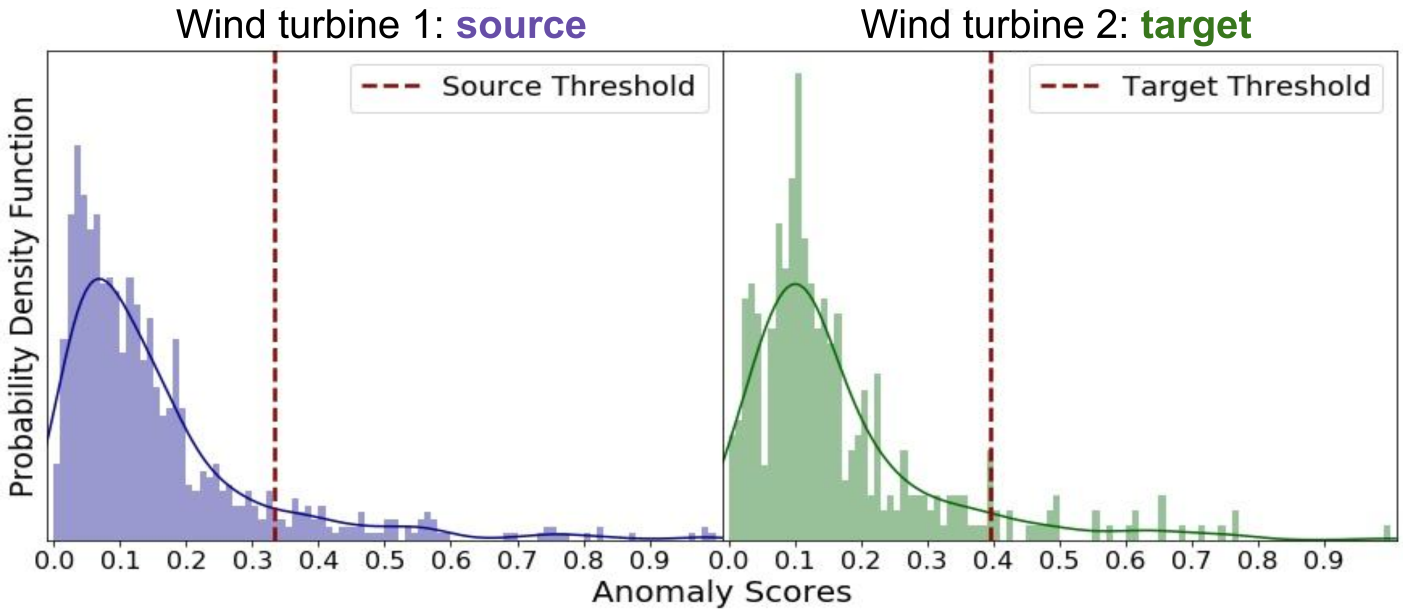
Step 2. Estimate the source predictive threshold

Set a **threshold** on the source anomaly scores such that the percentage of scores greater than the threshold **equals the source contamination factor**:



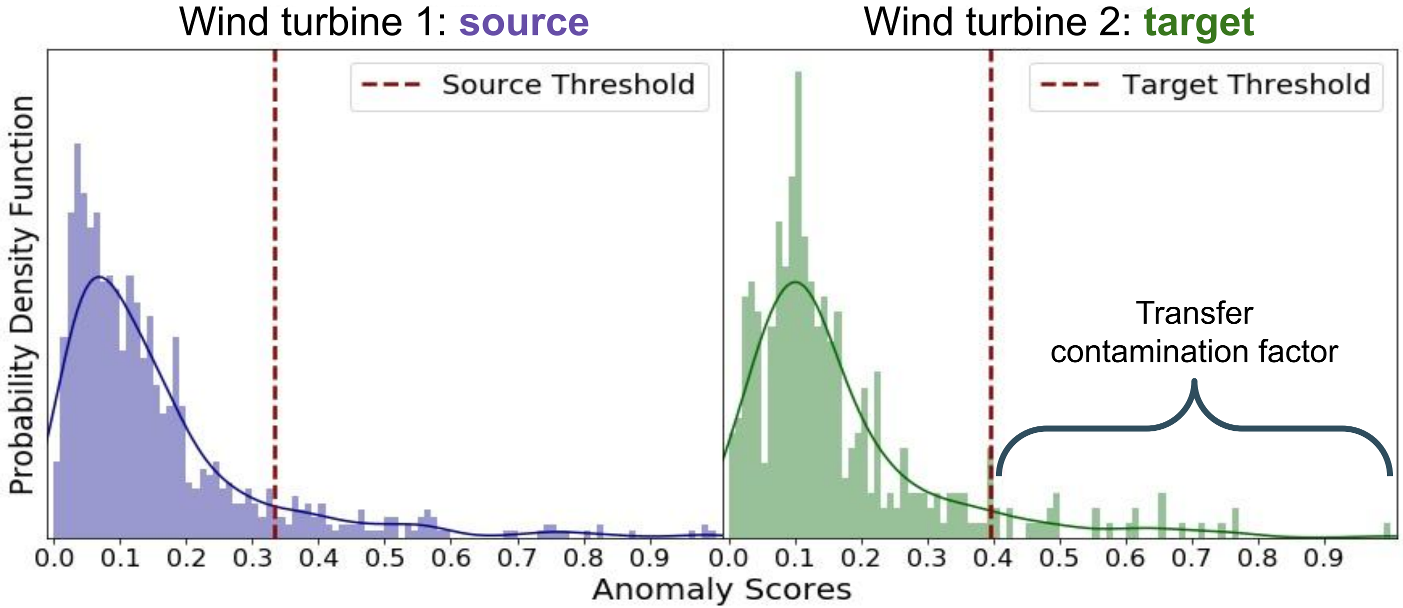
Step 3. Transfer the source predictive threshold

Find the threshold on the target scores such that the **shape of the score distributions** below the thresholds are as **similar** as possible:



Step 4. Derive the target contamination factor

Derive the **target contamination factor** from threshold as in step 2:



Weighing an ensemble of detectors yields better results

TrADe is model agnostic → Detectors may produce different scores distributions resulting in inaccurate estimates

Ensemble:

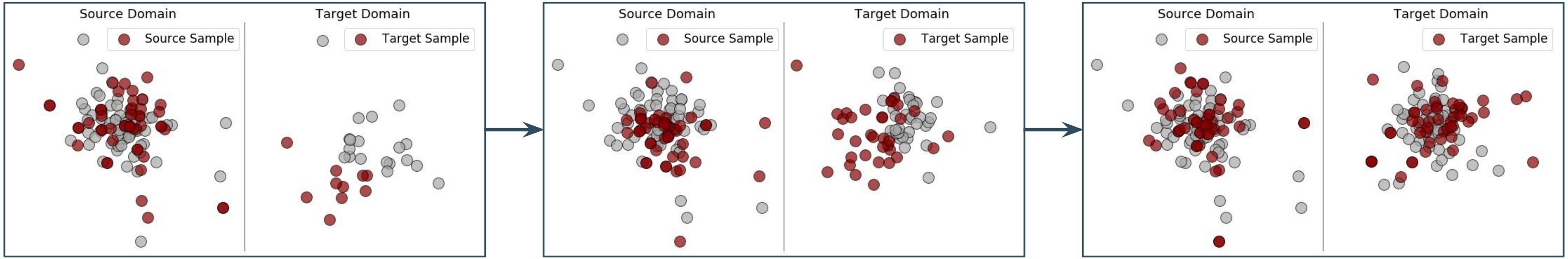
1. Estimate the target contamination using multiple detectors;
2. Compute the KL divergence between the source and target anomaly scores distribution;
3. Take a weighted average, where the weights depend on the KL values.

Theoretical Convergence Analysis

Key assumptions: target gets unbiased & same shape to the limit

1. The sample from the source domain is representative of its distribution, while the target sample reduces its potential bias when collecting new data;

sample size **n** sample size **k** sample size **n** sample size **2k** sample size **n** sample size **5k**



i.i.d sample *biased* sample i.i.d sample *less biased* sample i.i.d sample *almost unbiased* sample

2. The source and target normal distributions share the same shape to the limit.

Two theoretical results

1. The estimated target predictive threshold converges to its real value;
   ➔ The transfer is theoretically well defined and sound (Step 3);

2. The target contamination factor estimator is unbiased;
   ➔ Deriving the contamination factor from the predictive threshold returns the expected value (Step 2 and Step 4).

Extensive empirical evaluation

- ★ Q1. Does TrADe accurately estimate the target contamination factor?
- ★ Q2. Does a more accurate estimate improve the model performance?
- ★ Q3. Is an ensemble of detectors more accurate than single detectors?
- ★ Q4. How does TrADe perform when varying the source contamination factor?

Experimental setup

Three groups of datasets:

- 12 proprietary water consumption, 2 wind turbines, 9 public IoT datasets;

Setup:

- Pick a source-target pair (in total 12x11 + 2x1 + 9x8 = 206 pairs);
- Estimate the target contamination factor and use it to make predictions;

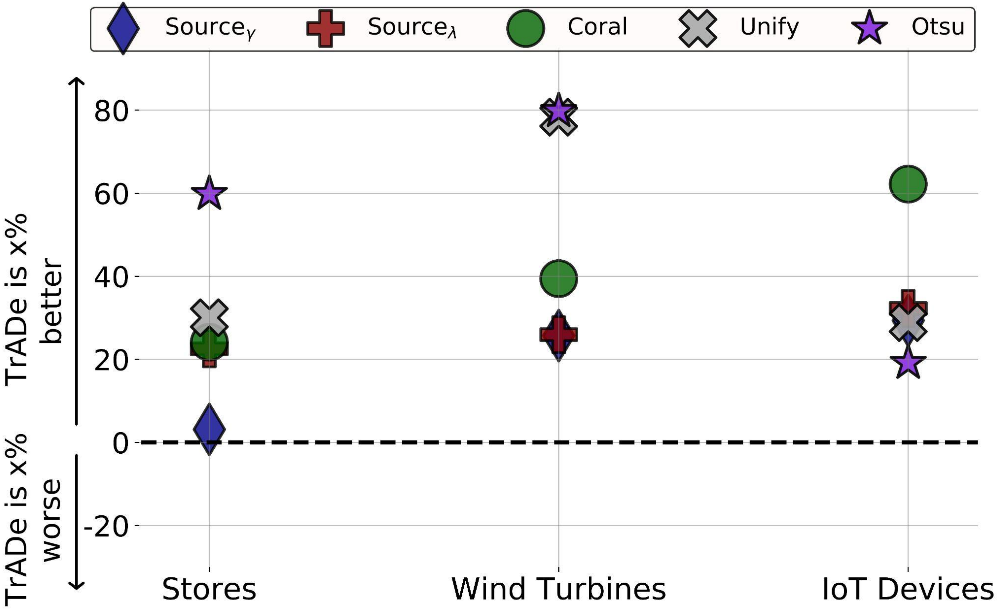
Baselines:

- **Source**$_{\gamma}$ and **Source**$_{\lambda}$ use the source values as target estimates;
- **Coral** aligns the two domains and uses Source$_{\lambda}$;
- **Unify** and **Otsu** estimate the contamination factor using only the target domain.

Q1. Does TrADe accurately estimate the target contamination factor?

Average Improvement of the ***MAE*** using **TrADe** against the baselines:

MAE



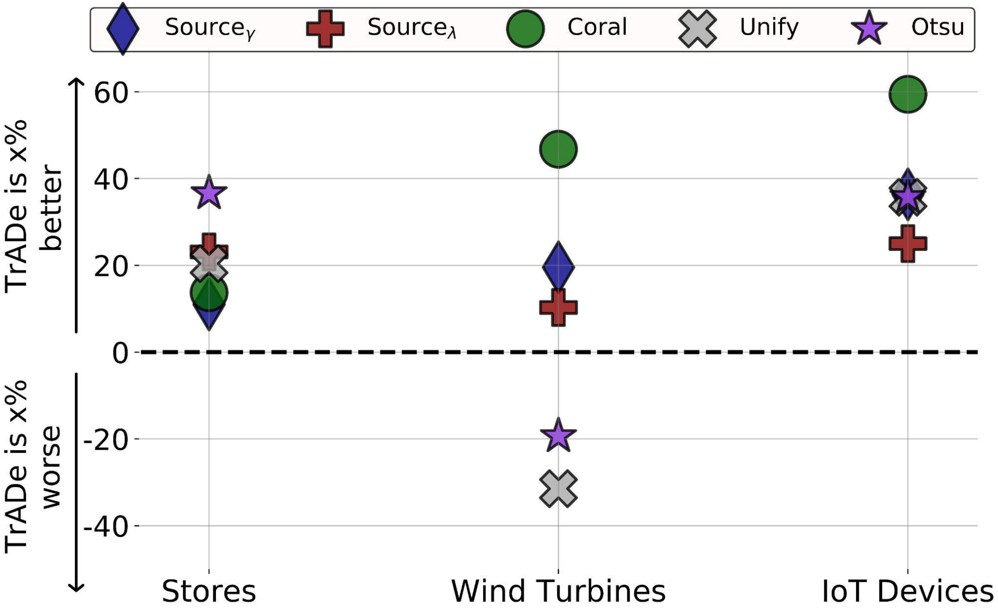
Aggregated results by

1. target domain;
2. dataset group.

Q2. Does a more accurate contamination factor improve the model performance?

Average Improvement of the ***F_1 score*** using **TrADe** against the baselines:

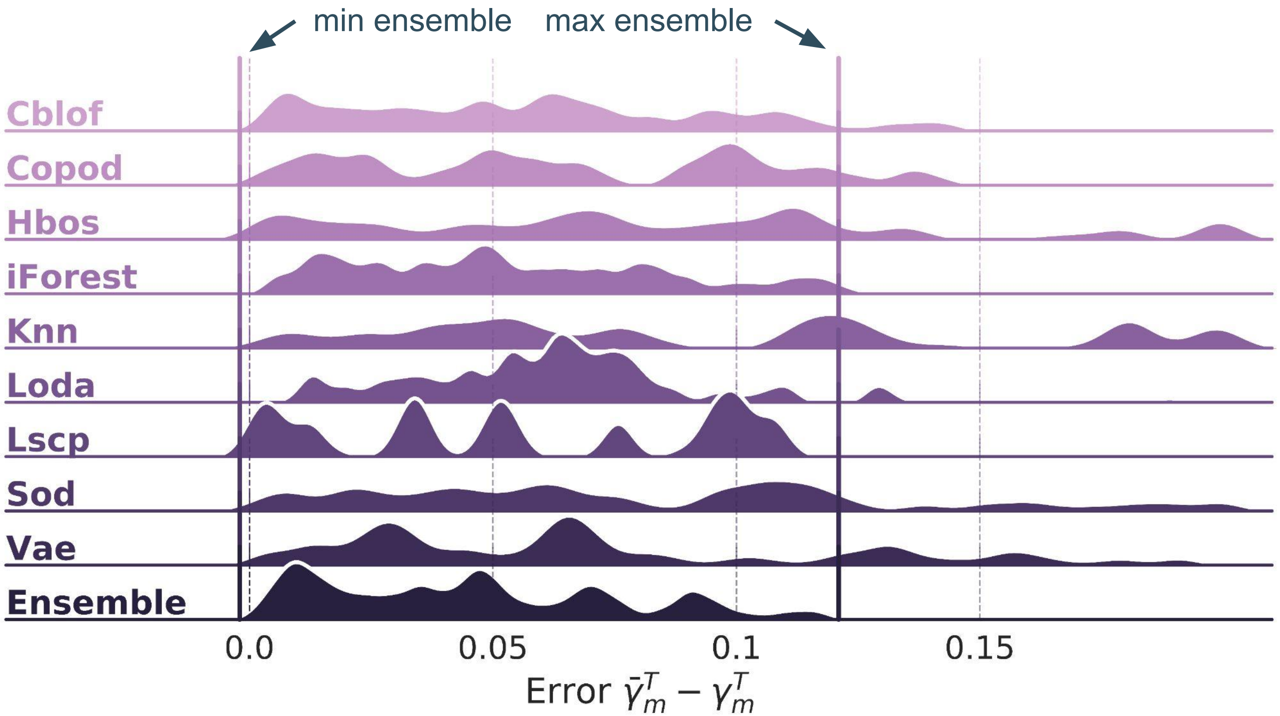
F_1 score



Aggregated results by

1. target domain;
2. dataset group.

Q3. How does the ensemble perform with respect to each individual variant?

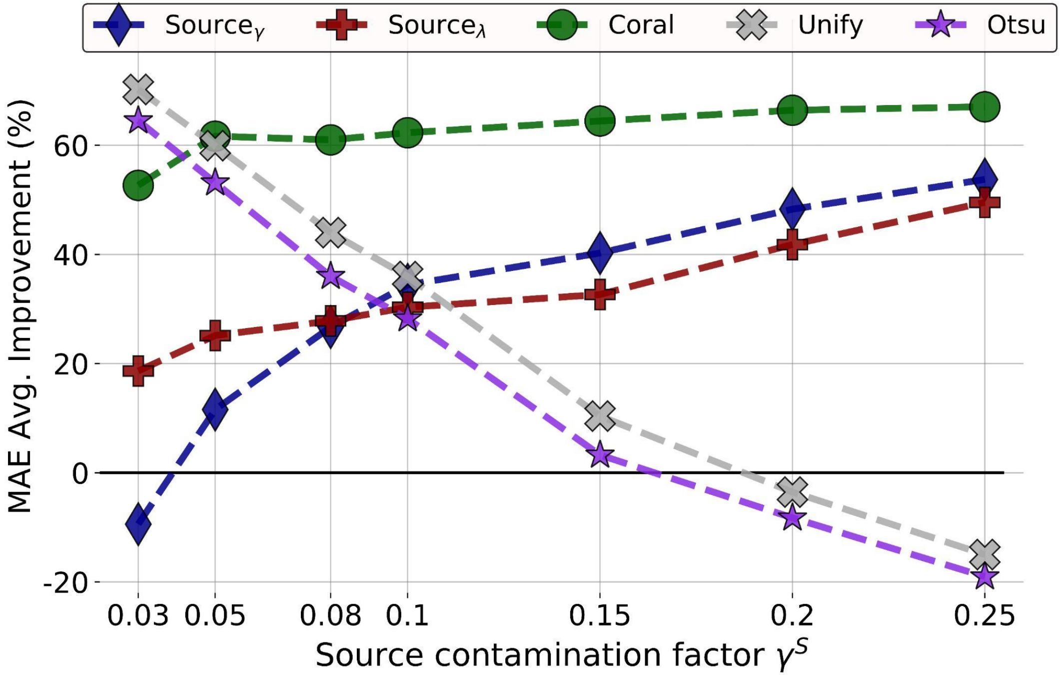


Each TrADe's variant is named with its anomaly detector

The ensemble shows little variance and limited errors

Q4. How does TrADe perform when varying the source contamination factor?

MAE



We aggregate results by target domain and we use only IoT data.

Target contamination fixed to 0.01.

In conclusion, we

- ➢ Introduce the problem of transferring the contamination factor;
- ➢ Propose TrADe, a novel 4-step approach for such a transfer;
- ➢ Analyze theoretically TrADe’s behaviour to the limit;
- ➢ Perform an extensive empirical analysis using benchmark and real world data.

Code and experiments are available online: *https://github.com/Lorenzo-Perini/TransferContamination*

Contact us
name.surname@kuleuven.be

Lorenzo Perini*, Vincent Vercruyssen, Jesse Davis*

AAAI 2022

https://people.cs.kuleuven.be/~lorenzo.perini/

@LorenzoPerini95