Transferring the Contamination Factor between Anomaly Detection Domains by Shape Similarity

Lorenzo Perini, Vincent Vercruyssen, Jesse Davis

Supplementary Material

This supplement contains the full proofs for our theoretical results and additional details about our experiments:

- First, we provide the proofs for the two proposition listed in the methodology section.
- Second, we give the proof of our Theorem 3 presented in the theoretical analysis section. To do so, we provide a number of subtheorems that are needed to prove the main result.
- Finally, we further supply additional details about our experiments and results.

Proofs of methodology section propositions

Proof of Proposition 1. By definition, p^λ is a distribution if two properties hold:

1. *Non-Negativity.* Since $p(x)$ is always non-negative, it follows that $p^\lambda(x) \geq 0$ for any $x \in [0,1]$;
2. *Integral equal to* 1. Integrating on the support,

$$\int_0^1 p^\lambda(x)\,dx = \int_0^1 \frac{\lambda p(\lambda x)}{\int_0^\lambda p(y)\,dy}\,dx = \int_0^\lambda \frac{\lambda p(z)}{\lambda \int_0^\lambda p(y)\,dy}\,dz = 1,$$

because the same integral is on both sides of the ratio. □

Proof of Proposition 2. By hypothesis, $\gamma_m^T = \mathbb{P}\left(T_m \geq \lambda_m^T\right) = \mathbb{P}\left(Y_m^T = 1\right)$. Then,

$$\begin{aligned}\mathbb{E}\left[\hat{\gamma}_m^T\right] &= \mathbb{E}\left[\frac{\sum_{i=1}^m \mathbb{1}_{\{h(x)\geq\lambda_m^T\}}(x_i)}{m}\right] \\ &= \sum_{i=1}^m \frac{\mathbb{E}\left[\mathbb{1}_{\{h(x)\geq\lambda_m^T\}}(x_i)\right]}{m} \\ &= \sum_{i=1}^m \frac{\mathbb{P}\left(\{h(x_i)\geq\lambda_m^T\}\right)}{m} \\ &= \sum_{i=1}^m \frac{\mathbb{P}\left(T_m \geq \lambda_m^T\right)}{m} = \sum_{i=1}^m \frac{\gamma_m^T}{m} = \gamma_m^T,\end{aligned}$$

where the equality $\mathbb{E}[\mathbb{1}_{\{h(x)\geq\lambda_m^T\}}(x)] = \mathbb{P}(\{h(x) \geq \lambda_m^T\})$ holds by the definition of indicator function.

Theoretical Convergence Analysis

Our **main theoretical result** can be summarized as follows. For $m \to +\infty$,

- *Our estimate of the **target threshold** λ_m^T converges to the real target value λ^T (Theorem 3 first part);*
- *Our estimate of the **target contamination factor** $\hat{\gamma}_m^T$ converges on average to the real target value γ^T (Theorem 3 second part).*

Because the contamination factor is derived from the predictive threshold, we first focus on proving that the target predictive threshold of T_m converges to the actual value of the ground-truth T. Essentially, the equalities in Eq. 3 hold if the limit symbol is allowed to pass out through the functions. We motivate the four steps of Eq. 3 by answering the four following questions:

Q1. Does the set of values minimizing $KL(S^{\lambda^S} \,||\, T^\lambda)$ contains only the real source threshold λ^T?

Q2. Given that $t_m \to t$ uniformly, does it hold for any λcut distribution, i.e. $t_m^\lambda \to t^\lambda$ for any $\lambda \in [\delta, 1]$?

Q3. Does the KL divergence $KL(S^{\lambda^S} \,||\, T_m^\lambda)$ converge to the KL divergence $KL(S^{\lambda^S}||T^\lambda)$?

Q4. Are the function $\arg\min$ and the limit interchangeable?

Finally, we move to the contamination factor by answering to:

Q5. Does the target contamination factor's estimate $\hat{\gamma}_m^T$ converge to the true target contamination factor γ^T?

Q1. Uniqueness of the Limit Given that the optimization problem in Eq. 1 may return a set of solutions, we first investigate the uniqueness of the theoretical target threshold λ^T, which is obtained as a solution of the optimization problem with the real T instead of T_m. For this task, we take advantage of the following theorem stating that if two λcut distributions are equal *almost surely*, then the two thresholds must be equal.

Theorem 4. *Let T be a real continuous random variable with distribution $t\colon [0,1] \to (0, +\infty)$. Let's assume that there exists $\lambda_1,\ \lambda_2 \in [\delta, 1]$, for any fixed $\delta > 0$, such that the equality*

$$t^{\lambda_1}(x) = t^{\lambda_2}(x)$$

holds for almost every $x \in [0,1]$, where t^{λ_1} and t^{λ_2} are, respectively, the λ_1 and λ_2 cut distributions of t. Then,

$$\lambda_1 = \lambda_2.$$

Proof. Since $\lambda_1,\ \lambda_2 \in [\delta, 1]$, there exists a value $a \in [\max\{\delta - 1, -\lambda_1, -\lambda_2\}, 1 - \delta]$ such that $\lambda_1 = \lambda_2 + a$. Then, for almost every $x \in [0,1]$,

$$t^{\lambda_1}(x) = t(\lambda_1 x)\cdot\frac{\lambda_1}{\int_0^{\lambda_1} t(y)\,dy} = t((\lambda_2+a)x)\cdot\frac{\lambda_2 + a}{\int_0^{\lambda_2+a} t(y)\,dy}$$

and

$$1 = \frac{t^{\lambda_1}(x)}{t^{\lambda_2}(x)} = \frac{t((\lambda_2+a)x)}{t(\lambda_2 x)}\cdot\frac{(\lambda_2+a)\int_0^{\lambda_2} t(y)\,dy}{\lambda_2\int_0^{\lambda_2+a} t(y)\,dy} = \frac{t((\lambda_2+a)x)}{t(\lambda_2 x)}\cdot\frac{1}{c},$$

where $\frac{1}{c} > 0$ refers to the second factor. Hence, we derive that

$$t((\lambda_2 + a)x) = c\cdot t(\lambda_2 x) \implies t\,(by) = c\cdot t(y),$$

where $b = \left(1+\frac{a}{\lambda_2}\right) \in \left(\max\left\{0, 1-\frac{1-\delta}{\lambda_1}, 1-\frac{\lambda_2}{\lambda_1}\right\}, 1+\frac{1-\delta}{\lambda_1}\right)$ and $y = \lambda_2 x \in [0, \lambda_2]$. By recurrence, for any $n \in \mathbb{N}$,

$$t(y) = c \cdot t\left(\frac{y}{b}\right) = \cdots = c^n \cdot t\left(\frac{y}{b^n}\right).$$

If $a > 0$, then it is easy to check that

$$c = \frac{\lambda_2}{\lambda_2 + a} \cdot \frac{\int_0^{\lambda_2+a} s(y)\, dy}{\int_0^{\lambda_2} t(y)\, dy} < 1.$$

Then, for $n \to +\infty$

$$t(y) = c^n \cdot t\left(\frac{y}{b^n}\right) \to 0 \text{ for all } y \in [0, \lambda_2],$$

which means that $t(y) = 0$ for all $y \in [0, \lambda_2]$ and, consequently, $t(x)$ is constant and equal to 0 for all $x \in [0, 1]$. This is a contradiction.

On the other hand, if $a < 0$ then $c > 1$. However, because t is positive and, especially, $t(0) > 0$, we can conclude that

$$t(y) = c^n \cdot t\left(\frac{y}{b^n}\right) \to +\infty \implies t(y) = +\infty \quad \forall\, y \in [0, \lambda_2],$$

which is again a contradiction. Thus, $a = 0$ and $\lambda_1 = \lambda_2$. □

We exploit this result to show the uniqueness of the limit solution.

Theorem 5 (Uniqueness of the solution)**.** *Let S and T be two real continuous random variables with distribution, respectively, $s, t\colon [0,1] \to (0, +\infty)$. Let $\lambda^S, \lambda^T \in [\delta, 1]$ be two fixed thresholds, for any $\delta > 0$, such that*

$$KL\left(S^{\lambda^S} \,\middle\|\, T^{\lambda^T}\right) = 0, \tag{4}$$

where S^{λ^S}, T^{λ^T} are, respectively, the λ^S and λ^T cut distributions. Then,

$$\arg\min_{\lambda \in [\delta,1]} \left\{KL\left(S^{\lambda^S} \middle\| T^{\lambda}\right)\right\} = \lambda^T.$$

Proof. By the Gibbs inequality, for any $\lambda \in [\delta, 1]$, the inequality

$$KL\left(S^{\lambda^S} \,\middle\|\, T^{\lambda}\right) \geq 0$$

holds. Thus, the inclusion

$$\lambda^T \in \arg\min_{\lambda \in [\delta,1]} \left\{KL\left(S^{\lambda^S} \,\middle\|\, T^{\lambda}\right)\right\}$$

comes directly from the hypothesis. Let's now assume that there exists another global minimum $\lambda^* \in [\delta, 1]$ such that

$$\lambda^* \in \arg\min_{\lambda \in [\delta,1]} \left\{KL\left(S^{\lambda^S} \,\middle\|\, T^{\lambda}\right)\right\},$$

which implies

$$KL\left(S^{\lambda^S} \,\middle\|\, T^{\lambda^*}\right) = 0.$$

We now prove that, for almost every $x \in [0,1]$, $t^{\lambda^T}(x) = t^{\lambda^*}(x)$. Let's start from $t^{\lambda^*}(x)$. By definition of KL divergence,

$$\begin{aligned}
0 &= KL\left(S^{\lambda^S} \,\middle\|\, T^{\lambda^*}\right) \\
&= \int_0^1 s^{\lambda^S}(x) \log\left(\frac{s^{\lambda^S}(x)}{t^{\lambda^*}(x)}\right) dx \\
&= -\int_0^1 s^{\lambda^S}(x) \log\left(\frac{t^{\lambda^*}(x)}{s^{\lambda^S}(x)}\right) dx \\
&\geq -\int_0^1 s^{\lambda^S}(x) \left(\frac{t^{\lambda^*}(x)}{s^{\lambda^S}(x)} - 1\right) dx \\
&= -\int_0^1 \left(t^{\lambda^*}(x) - s^{\lambda^S}(x)\right) dx = 0,
\end{aligned}$$

where the only inequality comes from the property of logarithms $\log(x) \leq x - 1$ for $x \in (0, 1]$, and the final result is 0 because each integral is equal to 1. As a result, the inequality turns out to be an equality and, for almost every $x \in [0,1]$, $s^{\lambda^S}(x) = t^{\lambda^*}(x)$. By repeating the same procedure with $t^{\lambda^T}(x)$ instead of $t^{\lambda^*}(x)$, we get that, for almost every $x \in [0,1]$, $s^{\lambda^S}(x) = t^{\lambda^T}(x)$. Because the union of two sets with measure equal to 0 is still a set with measure equal to 0 and by transitivity, we conclude that $t^{\lambda^*}(x) = t^{\lambda^T}(x)$ for almost every $x \in [0,1]$. By applying Theorem 4 to t^{λ^*} and t^{λ^T}, we prove that $\lambda^* = \lambda^T$, meaning that the solution is unique. □

Theorem 5 proves that only one solution exists, so that if the sequence of estimated predictive thresholds λ_m^T converges, then its limit is exactly the theoretical predictive threshold λ^T.

Q2. Convergence of λcut distributions Given that the λcut distribution is an actual distribution (Prop. 1), the following theorem shows that assuming that the target distribution t_m converges uniformly to the theoretical distribution t, then the same relationship holds for any of their λcut distributions.

Theorem 6 (Uniform convergence of λcut distributions)**.** *Let t_m, for any $m \in \mathbb{N}$, be a sequence of continuous distributions such that $t_m \to t$ uniformly in $[0,1]$ for $m \to +\infty$. Then, for any $\lambda \in [\delta, 1]$,*

$$t_m^\lambda \xrightarrow{m \to +\infty} t^\lambda \quad \textit{uniformly in } [0,1],$$

where t_m^λ and t^λ are λcut distributions as defined in definition 1.

Proof. Let's fix $\varepsilon > 0$ and $\lambda \in [\delta, 1]$. Since t is bounded, there exists a constant $K \geq 0$ such that $|t(x)| \leq K$ for all $x \in [0,1]$. As $t_m \to t$ uniformly, there exists $M_1 \in \mathbb{N}$ such that

$$|t_m(x) - t(x)| < 1 \quad \forall\, x \in [0,1],\ m \geq M_1,$$

which is equivalent to

$$|t_m(x)| < |t(x)| + 1 \leq K + 1 \quad \forall\, x \in [0,1],\ m \geq M_1.$$

Because of the uniform convergence of t_m, which is continuous, bounded and with integral equal to 1, the sequence $\int_0^\lambda t_m(y)\,dy$ converges to $\int_0^\lambda t(y)\,dy$ and so does the reciprocal sequence.
Thus, by definition, there exists $M_2 \in \mathbb{N}$ such that

$$\left|\frac{1}{\int_0^\lambda t_m(y)\,dy} - \frac{1}{\int_0^\lambda t(y)\,dy}\right| < \frac{\varepsilon}{2\lambda(K+1)} \quad \forall\, m \geq M_2.$$

Likewise, given that t_m converges uniformly to t, there exists $M_3 \in \mathbb{N}$ such that

$$|t_m(x) - t(x)| < \frac{\varepsilon \int_0^\lambda t(y)\,dy}{2\lambda} \quad \forall\, x \in [0,1],\ m \geq M_3.$$

Let's indicate by $M = \max\{M_1, M_2, M_3\}$. For $m \geq M$ and $x \in [0,1]$,

$$\begin{aligned}
&\left|t_m^\lambda(x) - t^\lambda(x)\right| = \lambda \left|\frac{t_m(x)}{\int_0^\lambda t_m(y)\,dy} - \frac{t(x)}{\int_0^\lambda t(y)\,dy}\right| \\
\leq &\left|\frac{\lambda t_m(x)}{\int_0^\lambda t_m(y)\,dy} - \frac{\lambda t_m(x)}{\int_0^\lambda t(y)\,dy}\right| + \left|\frac{\lambda t_m(x)}{\int_0^\lambda t(y)\,dy} - \frac{\lambda t(x)}{\int_0^\lambda t(y)\,dy}\right| \\
\leq &\lambda |t_m(x)| \cdot \left|\frac{1}{\int_0^\lambda t_m(y)\,dy} - \frac{1}{\int_0^\lambda t(y)\,dy}\right| + \lambda \frac{|t_m(x) - t(x)|}{\int_0^\lambda t(y)\,dy} \\
\leq &\lambda(K+1) \cdot \frac{\varepsilon}{2\lambda(K+1)} + \frac{\lambda}{\int_0^\lambda t(y)\,dy} \cdot \frac{\varepsilon \int_0^\lambda t(y)\,dy}{2\lambda} = \varepsilon,
\end{aligned}$$

where in the last step we exploit the previous inequalities. As a result, $t_m^\lambda \to t^\lambda$ uniformly for any $\lambda \in [\delta, 1]$ when $m \to +\infty$. □

Theorem 6 guarantees that the uniform convergence of the full distribution t_m to t (Assumption 1) is enough to claim that any λcut distribution of t_m converges to the corresponding λcut distribution of t.

Q3. Convergence of KL Divergence Given the uniform convergence of the λcut distributions, we still need to verify whether their KL divergence converges, as expected, to the KL divergence of their limit. The following theorem shows that the KL divergence and the limit symbol can be interchanged.

Theorem 7 (Interchangeability of KL divergence and limit symbol)**.** *Let S, T and T_m be three real continuous random variables with distribution, respectively, $s, t, t_m \colon [0,1] \to (0, +\infty)$. Fix $\lambda^S, \lambda \in [\delta, 1]$, with $\delta > 0$. Let s^{λ^S}, t^λ and t_m^λ be λcut distributions. Assume that $s^{\lambda^S}(x)$, $t^\lambda(x)$, $t_m^\lambda(x) > 0$ for all $x \in [0,1]$, and that $t_m^\lambda \to t^\lambda$ uniformly for all $x \in [0,1]$. Then,*

$$\begin{aligned}
\lim_{m\to+\infty} KL\left(S^{\lambda^S} \,\middle\|\, T_m^\lambda\right) &= KL\left(S^{\lambda^S} \,\middle\|\, \lim_{m\to+\infty} T_m^\lambda\right) \\
&= KL\left(S^{\lambda^S} \,\middle\|\, T^\lambda\right),
\end{aligned}$$

where S^{λ^S}, T^λ and T_m^λ are the unique variables with, respectively, s^{λ^S}, t^λ and t_m^λ as distributions.

Proof. By definition,

$$KL\left(S^{\lambda^S} \,\middle\|\, T_m^\lambda\right) = \int_0^1 s^{\lambda^S}(x) \log\left(\frac{s^{\lambda^S}(x)}{t_m^\lambda(x)}\right) dx. \quad (5)$$

Because it is both positive and continuous, t_m^λ takes its (positive) maximum and minimum values on the interval $[0,1]$ and, therefore, the reciprocal function is bounded. Following the same procedure as in Theorem 6, it is easy to prove that

$$\frac{s^{\lambda^S}(x)}{t_m^\lambda(x)} \to \frac{s^{\lambda^S}(x)}{t^\lambda(x)} \quad \text{uniformly for } x \in [0,1].$$

Likewise, by exploiting the continuity of the logarithm function and both the continuity and the boundedness of the λ-cut distributions, the following convergence naturally holds for $x \in [0,1]$

$$s^{\lambda^S}(x) \log\left(\frac{s^{\lambda^S}(x)}{t_m^\lambda(x)}\right) \to s^{\lambda^S}(x) \log\left(\frac{s^{\lambda^S}(x)}{t^\lambda(x)}\right) \quad \text{uniformly.}$$

Let's now fix $\varepsilon > 0$. By definition of uniform convergence, there exists $M \in \mathbb{N}$ such that, for all $m \geq M$ and for any $x \in [0,1]$,

$$\left|s^{\lambda^S}(x) \log\left(\frac{s^{\lambda^S}(x)}{t_m^\lambda(x)}\right) - s^{\lambda^S}(x) \log\left(\frac{s^{\lambda^S}(x)}{t^\lambda(x)}\right)\right| < \varepsilon.$$

Consequently, for all $m \geq M$,

$$\begin{aligned}
&\left|KL\left(S^{\lambda^S}\middle\|T_m^\lambda\right) - KL\left(S^{\lambda^S}\middle\|T_m^\lambda\right)\right| \\
= &\left|\int_0^1 s^{\lambda^S}(x) \log\left(\frac{s^{\lambda^S}(x)}{t_m^\lambda(x)}\right) dx - \int_0^1 s^{\lambda^S}(x) \log\left(\frac{s^{\lambda^S}(x)}{t^\lambda(x)}\right) dx\right| \\
\leq &\int_0^1 \left|s^{\lambda^S}(x) \log\left(\frac{s^{\lambda^S}(x)}{t_m^\lambda(x)}\right) - s^{\lambda^S}(x) \log\left(\frac{s^{\lambda^S}(x)}{t^\lambda(x)}\right)\right| dx \leq \varepsilon,
\end{aligned}$$

which proves the thesis. □

Provided that the limit can pass outside the KL divergence function, next we discuss the convergence of the transfer predictive threshold when solving the optimization problem at each step $m \in \mathbb{N}$.

Q4. Interchangeability of $\arg\min$ **and limit** The final step to prove that the transfer threshold λ_m^T converges to the actual value λ^T (first main result in Eq. 3) consists of exchanging the limit symbol with the argmin function. Although in general the equality may not hold, in our scenario we only need to show that one of the two inequalities is true. If we proved that

$$\begin{aligned}
&\limsup_{m\to+\infty} \arg\min_{\lambda\in[\delta,1]} \left\{KL\left(S^{\lambda^S}\middle\|T_m^\lambda\right)\right\} \subseteq \\
&\arg\min_{\lambda\in[\delta,1]} \left\{\lim_{m\to+\infty} KL\left(S^{\lambda^S}\middle\|T_m^\lambda\right)\right\}
\end{aligned}$$

we could conclude through Theorem 5 that the right hand side contains only one solution and, in turn, that the equality holds. For this task, we first need to introduce the following theorem, which states that the KL divergence is a continuous function when the threshold λ varies in $[\delta, 1]$.

Theorem 8 (Continuity of KL divergence)**.** *Given two continuous random variables S and T_m with density, respectively, $s(x)$ and $t_m(x)$ defined on $[0,1]$, and a constant $\delta > 0$, then the function*

$$f_m(\lambda) \coloneqq KL\left(S^{\lambda^S} \middle|\middle| T_m^{\lambda}\right) \text{ is continuous in } [\delta, 1],$$

where S^{λ^S} and T_m^{λ} are the random variable with, respectively, λcut distribution equal to s^{λ^S} and t_m^{λ}, as defined in definition 1.

Proof. By definition, for any $\lambda \in [\delta, 1]$,

$$t_m^{\lambda}(x) = t_m(\lambda x) \cdot \frac{\lambda}{\int_0^{\lambda} t_m(y)\, dy}.$$

The reciprocal function of the integral depends on λ and is continuous because it is bounded and t_m is continuous with respect to both x and λ. Because the product of continuous functions is still a continuous function, $t_m^{\lambda}(x)$ is continuous on $[\delta, 1]$. As a result,

$$\begin{aligned} f_m(\lambda) &= \int_0^1 s^{\lambda^S}(x) \log\left(\frac{s^{\lambda^S}(x)}{t_m^{\lambda}(x)}\right) dx \\ &= \int_0^1 s^{\lambda^S}(x) \left[\log\left(s^{\lambda^S}(x)\right) - \log\left(t_m^{\lambda}(x)\right)\right] dx \\ &= \int_0^1 s^{\lambda^S}(x) \log\left(s^{\lambda^S}(x)\right) dx - \int_0^1 s^{\lambda^S}(x) \log\left(t_m^{\lambda}(x)\right) dx \end{aligned}$$

is continuous on $[\delta, 1]$ because of combination of operations that preserve the continuity. In the last line only the second integral depends on λ through the function $t_m^{\lambda}(x)$, hence preserving the continuity of the integrand function. Thus, $f_m(\lambda)$ is continuous. □

This guarantees that the hypotheses of the following theorem hold.

Theorem 9 (Interchangeability of $\arg\min$ and limit symbol)**.** *Let $f_m \colon [\delta, 1] \to \mathbb{R}$ be a continuous sequence of functions for a fixed $\delta > 0$. Assume that f_m converges pointwise to a continuous function $f \colon [\delta, 1] \to \mathbb{R}$. Then,*

$$\begin{aligned} \limsup_{m \to +\infty} \left(\arg\min_{\lambda \in [\delta, 1]} f_m(\lambda) \right) &\subseteq \arg\min_{\lambda \in [\delta, 1]} \left(\lim_{m \to +\infty} f_m(\lambda) \right) \\ &= \arg\min_{\lambda \in [\delta, 1]} f(\lambda). \end{aligned}$$

Proof. To prove the result, we take an element in the set on the left hand side and show that the element belongs to the set on right hand side. Let $\bar{\lambda} \in [\delta, 1]$ be such that

$$\bar{\lambda} \in \limsup_{m \to +\infty} \left(\arg\min_{\lambda \in [\delta, 1]} f_m(\lambda) \right).$$

Then, by definition of limit superior, for all $m \in \mathbb{N}$, $\exists\, K(m) \in \mathbb{N}$, $K(m) \geq m$, such that $\bar{\lambda} \in \arg\min_{\lambda \in [\delta, 1]} f_{K(m)}(\lambda)$. For $\lambda \in [\delta, 1]$,

$$f_{K(m)}(\bar{\lambda}) \leq f_{K(m)}(\lambda).$$

Table 2: The number of examples, variables, and the contamination factor for each considered dataset. IoT datasets (last 9) contamination factor vary among a list of values.

Dataset	# Data	# Vars	γ
`store1-hour1`	1703	11	0.1274
`store1-hour2`	1703	11	0.1192
`store1-hour3`	1703	11	0.1262
`store1-hour4`	1703	11	0.1315
`store2-hour1`	1704	11	0.0370
`store2-hour2`	1704	11	0.0082
`store2-hour3`	1704	11	0.0540
`store2-hour4`	1704	11	0.0769
`store3-hour1`	1276	11	0.0337
`store3-hour2`	1276	11	0.0713
`store3-hour3`	1276	11	0.0165
`store3-hour4`	1276	11	0.0681
`turbine-15`	838	10	0.101
`turbine-21`	385	10	0.078
`Webcam`	2000	115	List
`Security Camera 838`	2000	115	List
`Security Camera 737`	2000	115	List
`Security Camera 1003`	2000	115	List
`Security Camera 1002`	2000	115	List
`Ennio Doorbell`	2000	115	List
`Ecobee Thermostat`	2000	115	List
`Danmini Doorbell`	2000	115	List
`Baby Monitor`	2000	115	List

Given that $\lim_{m \to +\infty} K(m) = +\infty$, the inequality

$$f(\bar{\lambda}) = \lim_{m \to +\infty} f_{K(m)}(\bar{\lambda}) \leq \lim_{m \to +\infty} f_{K(m)}(\lambda) = f(\lambda)$$

holds for all $\lambda \in [\delta, 1]$. As a result, $\bar{\lambda} \in \arg\min_{\lambda \in [\delta, 1]} f(\lambda)$. □

We apply this theorem by using $f_m(\lambda) = KL(S^{\lambda^S} || T_m^{\lambda})$, which is continuous according to Theorem 8. In addition, the pointwise convergence to $f(\lambda) = KL(S^{\lambda^S} || T^{\lambda})$ holds by the Theorem 7.

Experiments

Data. Table 2 shows the properties of the 23 real-world anomaly detection datasets. The water data are proprietary and shared with the researchers under an NDA. They cannot be made public without consent of the providing company. The wind turbine data can be downloaded from http://www.industrial-bigdata.com/Data (Zhang et al. 2018). The IoT data can be found at https://archive.ics.uci.edu/ml/datasets/detection_of_IoT_botnet_attacks_N_BaIoT# (Meidan et al. 2018; Mirsky et al. 2018). While water and wind turbines data preserve their original structure, we subsample 2000 examples from the IoT data setting the contamination factor as $[0.03, 0.05, 0.08, 0.10, 0.15, 0.20, 0.25]$ when used as source domain, and as 0.01 when used as target domain.

Hyperparameters. Because our goals are i) to recover the contamination factor, and ii) prove the impact of more accurate estimates on the anomaly detector's performance, we

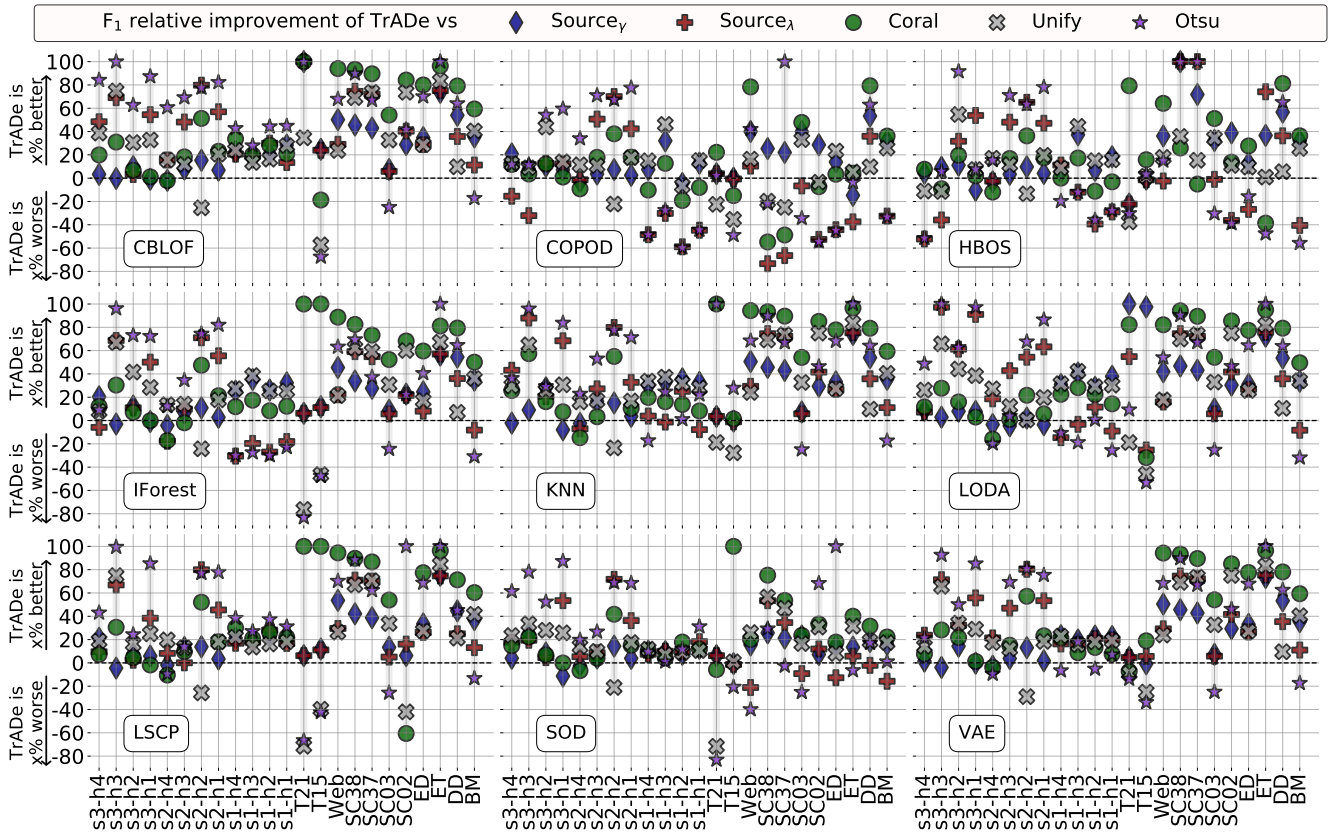

Figure 4: F_1 *relative improvement* of TRADE over the baselines, divided by anomaly detector h and averaged per target domain, over all the 206 experiments. Positive values mean that TRADE performs better than the baseline. Overall, TRADE shows positive F_1 scores improvements against all the baselines in at least 16 target domains for 7 out of 9 anomaly detectors. However, when averaging over the target domains, TRADE performs worse than SOURCE$_\lambda$ only when COPOD is used as anomaly detector.

Table 3: Comparison of the performance of TRADE, which uses an ensemble of detectors to estimate the contamination factor, to a variant that only uses a single-detector. Performance is measured in terms of the accuracy of the estimate of γ as measured by mean absolute error (MAE). We report the number of times TRADE wins (lower MAE), draws (absolute differences ≤ 0.001), and loses (higher MAE) versus each variant. Each variant is identified by the name of the considered anomaly detector.

TRADE Variant	**MAE on γ** Wins	Draws	Loses
KNNO	**105**	2	27
IFOREST	**74**	5	55
CBLOF	**91**	5	38
COPOD	**95**	3	36
LODA	**85**	4	45
VAE	**101**	1	32
HBOS	**113**	1	20
SOD	**101**	9	24
LSCP	**82**	0	52
Total	**847**	30	329

select sensible hyperparameters and set the remaining as default[8]: KNN has $k = 200$, CBLOF has $n\_clusters = 30$, HBOS has $n\_bins = 25$, SOD has $n\_neighbors = 25$ and $ref\_set = 10$, IFOREST has $n\_estimators = 5$, LODA has $n\_bins = 50$ and $n\_random\_cuts = 200$, LSCP has $local\_region\_size = 150$ and uses 3 LOF with $k = 50$, VAE has $[10, 2, 2, 10]$ as layer structure and $epochs = 10$, while COPOD has no hyperparameters.

Results (Q2). Figure 4 shows the F_1 *relative improvement* for each anomaly detector h over all the 206 experiments, averaged by target domain. For 7 out of 9 anomaly detectors, TRADE performs on average better than each baseline in at least 16 out of 23 target domains. When using HBOS as anomaly detector, TRADE has an average improvement between 11.5% and 21% against all the baselines, despite showing negative improvements in 13 out of 23 experiments against SOURCE_λ.

Results (Q3). Table 3 illustrates the difference between using TRADE and its single-detector variants for estimating the contamination factor γ. Overall, TRADE results in better estimates of the contamination factor (lower MAE) than the single-detector variants. In the worst case, it wins 74 times and loses 55 times compared with the IFOREST variant.

[8]implementation is available on PYOD: https://pyod.readthedocs.io/en/latest/