UNIVERSITÉ DE ROUEN NORMANDIE | litis | INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES ROUEN NORMANDIE | ALLIANCE SORBONNE UNIVERSITÉ | utc Recherche LMAC

CALL FOR A POSTDOCTORAL RESEARCHER IN:

Domain Adaptation for Anomaly Detection on Highly Heterogeneous Logs-based Data, with Application to Electric Vehicle Charging Stations

Topics Deep learning; Log-based data, Anomaly Detection, Domain Adaptation, Electric Vehicle Charging Stations

Abstract Log-based anomaly detection (AD) is critical for reliability, safety, and security in complex cyber-physical and cloud systems. While deep learning has advanced log-based, these models often fail under domain shift: new environments (different organizations, clusters, configurations, software versions, observability stacks, and workloads) generate logs with changing vocabularies, templates, event semantics, and class priors. This proposal aims at developing a general framework that (*i*) learns deep latent representations for heterogeneous logs, (*ii*) constructs a *reference distribution* of normal behavior under *domain shift*, and (*iii*) enables *distributed* learning across geographically dispersed devices while preserving data locality. The core idea is to represent each machine/site as a distribution in a shared latent space, yielding a "distribution of distributions" view that supports distributed based decision rules or one-class modeling. The framework will be validated first on available benchmarks log datasets (e.g., server/network monitoring) and then transferred to Electric Vehicle Charging Stations (EVCS). Expected outcomes include new adaptation objectives, distributed reference construction algorithms, robust evaluation protocols.

1 Context

Logs are ubiquitous in cloud-native and distributed systems, capturing events from microservices, kernels, containers, network components, databases, and security controls. Log-based anomaly detection has advanced via sequence modeling and representation learning [1, 2, 3, 4]. However most methods implicitly assume *train and test are drawn from the same distribution*. In practice, monitoring systems are deployed in new domains where:

- Syntactic heterogeneity: different software versions produce different logtemplates, while similar log messages may imply different system states in different contexts, i.e., identical tokens can imply different states across systems, and different tokens can describe the same state;
- Temporal drift: upgrades, incident remediation, and workload evolution shift normal behavior;
- Label scarcity: anomalies are rare; labeling is expensive and often inconsistent across sites.

This motivates *unsupervised* and *semi-supervised* domain adaptation where labeled data exist primarily in a source domain, while the target domain has limited or no labels [5, 6]. Let the labeled source domain be $\mathcal{D}_s = \{(x_i^s, y_i^s)\}_{i=1}^{n_s}$ and the unlabeled target domain be $\mathcal{D}_t = \{x_j^t\}_{j=1}^{n_t}$. Each x denotes a log sequence (e.g., a session, trace window, or sliding-time window) derived from raw log lines; $y \in \{0, 1\}$ denotes normal vs anomaly (or multi-class incident categories when available). The objective is to learn an anomaly scoring function $f_\theta(x)$ that minimizes target risk despite distribution shift using labels only in $\mathcal{D}_s$ and unlabeled $\mathcal{D}_t$.

Related works

Early log AD methods rely on parsing and sequence modeling (e.g., LSTM-based DeepLog) [1], and later unsupervised approaches detect sequential and quantitative anomalies [2]. Log representation learning has expanded with attention and self-supervision [3]. Domain adaptation methods include adversarial feature alignment (DANN) [5] and discrepancy minimization such as Maximum Mean Discrepancy (MMD) [7]. Optimal transport has emerged as a flexible distribution alignment principle [8]. For anomaly detection, one-class objectives (e.g., Deep SVDD) are natural building blocks [9]. This proposal integrates these approaches into a log-specific, deployment-oriented framework. A recent comprehensive study synthesizes ML techniques for log-based anomaly detection and highlights evaluation challenges [4].

2 Scientific objectives: general to application

The general goal of this proposal aims at developing a principled method to construct and maintain a **reference distribution** of normality across heterogeneous domains and machines, learned via deep representations, and used for anomaly detection via distributions divergences and/or one-class objectives. Then apply the general framework to **anomaly detection for electric vehicle charging stations**, where the same charger model can behave differently depending on geography, grid constraints, user behavior, network quality, and operator policies.

We consider a set of machines (or sites) indexed by $k \in \{1, \ldots, K\}$. Each machine produces log-derived sequences or windows $x_{k,t}$ over time t. We learn a deep encoder $h_{k,\theta}(\cdot)$ mapping raw log windows to latent vectors: $z_{k,t} = h_{k,\theta}(x_{k,t})$. For each machine is represented by an empirical distribution $P_k(t)$ in latent space and defined over a time interval (e.g., day/week) to track drift. Across machines, we model $\{P_k(t)\}_{k=1}^{K}$ as samples from a higher-level meta-distribution $\mathcal{M}(t)$ capturing the population: $P_k(t) \sim \mathcal{M}(t)$. The key tasks are: (i) Learn $h_{k,\theta}$ that is robust to domain shift (domain adaptation) and construct a **reference distribution** (or reference set) for normal behavior $\mathcal{M}_{\text{ref}}(t)$. Define anomaly decision functions from distances between distributions (OT / divergence) or one-class objectives and detect when $P_k(t)$ or $\mathcal{M}(t)$ shifts over time.

2.1 Objectives

- Domain adaptation for heterogeneous logs. How can we learn deep representations of logs that transfer across domains (machines/sites/vendors) without requiring target labels?
- Reference distributions. How can we estimate $\mathcal{M}_{\text{ref}}(t)$ via distributed learning when raw data remains local, and machines are non-IID?
- OT / one-class decision functions at the distribution level. What anomaly scores built on divergences between $\mathcal{M}(t)$ and $\mathcal{M}_{\text{ref}}$ are robust under heterogeneity and drift?
- Temporal drift, rupture detection, and irregular sampling: How do we handle machines whose time series are sampled at different rates, and how do we detect abrupt and gradual distributional changes [10].
- Application to EVCS. How can we incorporate station ecology (geography, site type, connectivity, usage patterns) and spatial information to improve adaptation and reduce false alarms?

3 Expected achievements

- A unified framework for domain-adapted log anomaly detection under extreme heterogeneity. alignment.
- Reference distribution construction methods (clustering/OT/hierarchical): a general distribution-of-distributions framework for heterogeneous data detection.
- OT- and one-class-based decision functions robust to non-IID clients and drift.
- Anomaly detection methods that handle heterogeneous sampling rates: decision functions based on OT and/or one-class-based decision functions that are robust to non-IID clients and drift.
- An EVCS-focused blueprint.

4 Benchmarking: datasets and baselines

We will validate on a mix of public and partner datasets. Public candidates include HDFS logs used widely in log AD studies [1, 2]. Baselines include DeepLog [1], LogAnomaly [2], attention-based log models [3], and generic DA methods such as DANN [5] and MMD-based alignment [7] adapted to log embeddings.

Application in EVCS. Detecting malfunctions in EVCS is now an industrial challenge for the sustainability of the existing fleet. One of the obstacles to detecting anomalies is the high degree of heterogeneity in the network, due to different contexts (geographical, temporal) and types of terminal. We will focus on operationally meaningful anomaly categories, including:

- Session failures: failed authorization, handshake failures, unexpected disconnects, incomplete transactions.
- Degraded performance: unusually slow charging, repeated retries, abnormal heartbeats/timeouts.
- Device faults: connector/contactor errors, meter faults, thermal alerts, firmware crash loops.
- Network/backend incidents: gateway instability, roaming API failures, database latency spikes.
- Security/fraud signals: suspicious repeated auth attempts, anomalous transaction patterns, tamper-like events.

Requirements

- PhD in Applied Mathematics / Computer Science / Data Science / Machine Learning / or a related field.
- Strong publication record in high-impact international journals and conferences in the fields of machine learning, or deep learning.
- Extensive experience with advanced deep learning using tools such as PyTorch within large-scale, multi-GPU computing environments.
- Ability to work independently and collaborate effectively within a team.

Application procedure

The research will take place within the LITIS laboratory located at INSA Rouen, France. The Postdoc will be jointly supervised by Gilles Gasso, Maxime Bérar (LITIS) and Mokhtar Z. Alaya (LMAC - UTC). Interested candidates should submit the following documents in PDF format to: maxime.berar@univ-rouen.fr, gilles.gasso@insa-rouen.fr, and alayaelm@utc.fr

- Cover letter outlining research interests, expertise, and reasons for interest in the position.
- Curriculum vitae highlighting relevant qualifications and experience.
- Publications or research papers.
- Contact information for at least two references.

Preselected candidates will be invited to an interview.

Additional practical information

- This project is funded by ANR SHARP: Machine Learning for Safe Vehicle Charging Points.
- Duration of the contract is 1 year. The position will start in September 2026.
- The monthly gross salary ranges from 2510€ to 2584€ depending on experience.

References

[1] M. Du, F. Li, G. Zheng, and V. Srikumar, "Deeplog: Anomaly detection and diagnosis from system logs through deep learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2017.

[2] W. Meng, Y. Liu, Y. Zhu, S. Zhang, D. Pei, Y. Liu, R. Chen, and Y. Zhang, "Loganomaly: Unsupervised detection of sequential and quantitative anomalies in unstructured logs," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI)*, 2019.

[3] S. Nedelkoski, J. Cardoso, and O. Kao, "Self-attentive classification-based anomaly detection in unstructured logs," in *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, 2020.

[4] S. Ali, C. Boufaied, D. Bianculli, P. Branco, and L. Briand, "A comprehensive study of machine learning techniques for log-based anomaly detection," *Empirical Softw. Engg.*, vol. 30, no. 5, Jun. 2025. [Online]. Available: https://doi.org/10.1007/s10664-025-10669-3

[5] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016.

[6] G. Wilson and D. J. Cook, "A survey of unsupervised deep domain adaptation," *ACM Transactions on Intelligent Systems and Technology*, vol. 11, no. 5, pp. 1–46, 2020.

[7] M. Long, Y. Cao, J. Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," in *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, 2015.

[8] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy, "Joint distribution optimal transportation for domain adaptation," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[9] L. Ruff, R. A. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft, "Deep one-class classification," in *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018.

[10] M. Ye, X. Fang, B. Du, P. C. Yuen, and D. Tao, "Heterogeneous federated learning: State-of-the-art and research challenges," *ACM Comput. Surv.*, vol. 56, no. 3, Oct. 2023. [Online]. Available: https://doi.org/10.1145/3625558