ECML PKDD PhD Forum

European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases

2025



September 15, 2025
Porto, Portugal

Organizers

Raquel Sebastião

IEETA - Institute of Electronics and Telematics Engineering of Aveiro, University of Aveiro, Aveiro, Portugal
CISeD – Research Centre in Digital Services, Polytechnic University of Viseu, Viseu, Portugal

Yun Sing Koh

School of Computer Science, University of Auckland, New Zealand

Accepted Submissions

Explore data in an explainable manner: an interpretable clustering algorithm
Fuyin Lai

Identifying Biological Features in Protein Language Model Embeddings for Functional Effect Prediction
Paulo Yanez Sarmiento

Tackling Data Scarcity in Machine Learning: Leveraging Symbolic Regression for Data Generation
Matthias Kuntz

Modelling of Forced Dynamical Systems with Laplace-Net
Bernd Zimmering

Introducing ISCNNs: Input Smooth Convex Neural Networks
Qinyan Zhou

High-Level Semantic Understanding in Multimodal Video Content
Oron Nir

Towards Efficient Model Explainability in Low-Latency Machine Learning Systems
Inês Silva

Robust Evaluation of Fairness and Explainability in AI Systems
Sérgio Jesus

EileForecast: An Open-Source Forecasting Module for Energy Management in SMEs
Claudia Ehrig

Toward Personalized Career Recommendations Using JobHop
Iman Johary

Regression-Based Estimation of Causal Effects in the Presence of Selection Bias and Confounding
Marlies Hafer

Time-varying Gaussian Process Bandit Optimization with Experts: no-regret in logarithmically-many side queries
Eliabelle Mauduit

Skewness-Robust Causal Discovery in Location-Scale Noise Models
Daniel Klippert

Interpretable Generative Stress Testing: A Framework to Audit Decision Boundaries through Synthetic Data Generation
Inês Gomes

Multimodal Machine Learning Analysis of School Attendance in New Zealand
Tingrui Qiao

Clustering Massive, Noisy, and Unstructured Textual Streams
Cesar Andrade

Online Learning from Drifting Capricious Data Streams with Flexible Hoeffding Tree
Ruirui Zhao

Identification of malicious devices in Federated Machine Learning Settings
Yohannes Biadgligne

Learning Topology Actions for Power Grid Control
Mohamed Hassouna

Adaptive Sampling Strategies for Learning from Imbalanced Regression Data Streams
Ehsan Aminian

Why t-SNE Works: Human-Aligned Evaluation and a Parametric Alternative
Shirin Mohebi

Leveraging Large Language Models and Semantic Enhancement for Vision-Language Grounding
Tahsir Ahmed Munna

Transfer Learning for Evolving Domains
Ricardo Pereira

Electrocardiogram for Biometric Recognition: Collectability, Stability and Application Challenges
Teresa Pereira

Symbolic Representation Analysis for Change Detection in Physiological Signals Using Compression-Based Similarity
Daniela Pais

A Machine Learning-based Data Quality Framework for Regularly Sampled Sensors
Catarina Canastra

Diffusion-based Adversarial Purification for Robust Deep Learning
Jia Fu

Information Leakage in Generative Models
Catarina Gomes

Improving Multi-Task Parameter-Efficient Fine-Tuning
Robert Belanec

Developing Robust Methods for Efficient Question Generation over Knowledge Base in Open Environments
Jiuyang Tang

Contents

1 Explore data in an explainable manner: an interpretable clustering algorithm **9**
Fuyin Lai, Jefrey Lijffijt, Tijl De Bie . . . . . .

2 Identifying Biological Features in Protein Language Model Embeddings for Functional Effect Prediction **11**
Paulo Yanez Sarmiento, Pia Francesca Rissom, Manuel Pfeuffer, Jordan Safer, Sumaiya Iqbal, Henrike Heyne, Nadja Klein, Bernhard Renard . . . . . .

3 Tackling Data Scarcity in Machine Learning: Leveraging Symbolic Regression for Data Generation **13**
Matthias Kuntz, Torben Weis . . . . . .

4 Modelling of Forced Dynamical Systems with Laplace-Net **14**
Bernd Zimmering, Cecília Coelho, Oliver Niggemann . . . . . .

5 Introducing ISCNNs: Input Smooth Convex Neural Networks **16**
Qinyan ZHOU, Andrea Simonetto, Manh-hung Le . . . . . .

6 High-Level Semantic Understanding in Multimodal Video Content **20**
Oron Nir, Ariel Shamir . . . . . .

7 Towards Efficient Model Explainability in Low-Latency Machine Learning Systems **22**
Inês Silva, Sérgio Jesus, Iker Perez, Carlos Soares, Rita Ribeiro

8 Robust Evaluation of Fairness and Explainability in AI Systems **24**
Sérgio Jesus, Pedro Saleiro, Rita P. Ribeiro, João Gama . . . .

9 EileForecast: An Open-Source Forecasting Module for Energy Management in SMEs **26**
Claudia Ehrig, Ursula Neumann, Catherine Cleophas, Germain Forestier . . . . . .

10 Toward Personalized Career Recommendations Using JobHop **31**
Iman Johary, Raphaël Romero, Alexandru Mara, Tijl De Bie . .

11 Regression-Based Estimation of Causal Effects in the Presence of Selection Bias and Confounding **33**
Marlies Hafer, Alexander Marx . . . . . .

12 Time-varying Gaussian Process Bandit Optimization with Experts: no-regret in logarithmically-many side queries **35**
Eliabelle Mauduit, Andrea Simonetto, Eloise Berthier . . . . . .

13 Skewness-Robust Causal Discovery in Location-Scale Noise Models **39**
Daniel Klippert, Alexander Marx . . . . . .

14 Interpretable Generative Stress Testing: A Framework to Audit Decision Boundaries through Synthetic Data Generation **41**
Inês Gomes, Jan N. van Rijn, Luís F. Teixeira . . . . . .

15 Multimodal Machine Learning Analysis of School Attendance in New Zealand **45**
Tingrui Qiao . . . . . .

16 Clustering Massive, Noisy, and Unstructured Textual Streams **47**
Cesar Andrade . . . . . .

17 Online Learning from Drifting Capricious Data Streams with Flexible Hoeffding Tree **49**
Ruirui Zhao . . . . . .

18 Identification of malicious devices in Federated Machine Learning Settings **54**
Yohannes Biadgligne, Yassine Baghoussi, Kai Li, Alípio Jorge .

19 Learning Topology Actions for Power Grid Control **56**
Mohamed Hassouna . . . . . .

20 Adaptive Sampling Strategies for Learning from Imbalanced Regression Data Streams **57**
Ehsan Aminian, Rita P. Ribeiro, João Gama . . . . . .

21 Why t-SNE Works: Human-Aligned Evaluation and a Parametric Alternative **59**
Shirin Mohebi,Jefrey Lijffijt, Tijl De Bie . . . . . .

22 Leveraging Large Language Models and Semantic Enhancement for Vision-Language Grounding **61**
Tahsir Ahmed Munna, Nuno Guimarães, Alípio Jorge . . . . . .

23 Transfer Learning for Evolving Domains **64**
Ricardo Pereira, Jacopo Bono, Pedro Ribeiro, Carlos Soares . .

24 Electrocardiogram for Biometric Recognition: Collectability, Stability and Application Challenges **66**
Teresa Pereira, Raquel Sebastião, Raquel Conceição, Vitor Sencadas

25 Symbolic Representation Analysis for Change Detection in Physiological Signals Using Compression-Based Similarity **68**
Daniela Pais, Susana Brás, Manuel Vico Ávalos, Raquel Sebastião

26 A Machine Learning-based Data Quality Framework for Regularly Sampled Sensors **70**
Catarina Canastra, António Casimiro, Pedro Ferreira . . . . . .

27 Diffusion-based Adversarial Purification for Robust Deep Learning **73**
Jia Fu, Anders Holst, Sepideh Pashami . . . . . .

28 Information Leakage in Generative Models **75**
Catarina Gomes, João P. Vilela, Ricardo Mendes . . . . . .

29 Improving Multi-Task Parameter-Efficient Fine-Tuning **77**
Robert Belanec, Ivan Srba, Maria Bielikova . . . . . .

30 Developing Robust Methods for Efficient Question Generation over Knowledge Base in Open Environments **79**

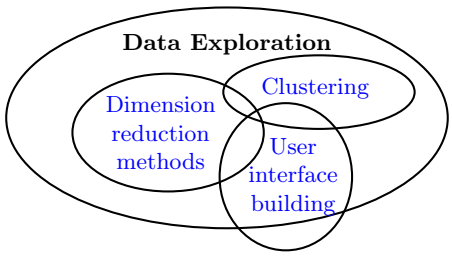
Explore data in an explainable manner: an interpretable clustering algorithm

Fuyin Lai

Supervisors: Jefrey Lijffijt, Tijl De Bie

Ghent University, Belgium

Abstract. Motivation to my research scope: Dimension reduction (DR) methods are frequently used for exploring high dimensional (HD) data. However, exploring data through embeddings is challenging because of the non-explainable embedding axis and distortion imposed by DR methods, especially non-linear DR method like t-SNE.



Black: General scope Blue: First project topics

Fig. 1: Research Scope

First Project (available on arXiv): To facilitate the exploration and interpretation of embeddings of HD data, I developed an Informative Clustering algorithm, named InfoClus. The idea is to partition the embedding into groups, each of which is given a sparse explanation using the original high-dimensional attributes. It works by optimizing an objective function that balances an information-theoretic quantification of the information users may gain from a partition and the corresponding best sparse explanation, with how complex the explanations are for users.

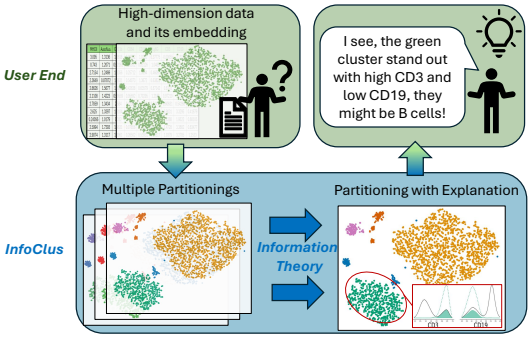
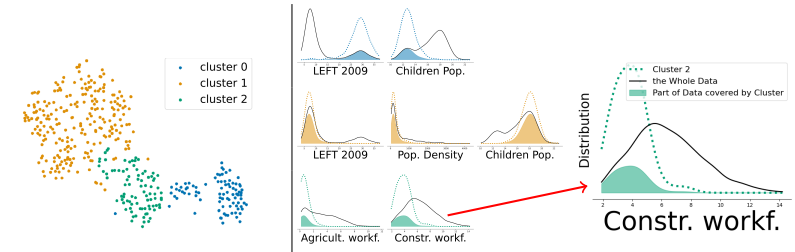


Fig. 2: InfoClus workflow

Early results: An interesting result is the experiment done on German-Socio-Eco dataset, which contains population and voting data from the 2009 elections. The partition found by InfoClus (see Fig. 3) clearly distinguish east and western Germany (cluster 0 and 1) without coordinates or other spatial information fed to InfoClus. Cluster 2 also highlight the industrial areas in Germany. We verify and show it by re-plotting all instances on the map of Germany (see Fig. 4). This interesting finding verifies the effectiveness of the measures InfoClus applies.



(a) InfoClus partitioning (b) Explanation for the left partitioning

Fig. 3: InfoClus analysis in GSE



(a) Cluster 0 (b) Cluster 1 (c) Cluster 2

Fig. 4: Visualization of InfoClus clusters on the map of Germany.

Future Scope: In the future, I plan to work further on helping users to understand the apparent patterns in embeddings and detect errors or distortions imposed by DR methods. Also, extending the data type to non-transaction based, like images, is also very interesting for data exploration.

Keywords: Data exploration · Dimension reduction · Explanation.

Identifying Biological Features in Protein Language Model Embeddings for Functional Effect Prediction

Paulo Yanez Sarmiento[1](✉), Pia Francesca Rissom[1,2], Manuel Pfeuffer[3], Jordan Safer[2], Sumaiya Iqbal[2], Henrike O. Heyne[1], Nadja Klein[4], and Bernhard Y. Renard[1](✉)

[1] Hasso Plattner Institute, Digital Engineering Faculty, University of Potsdam, Germany
{paulo.yanez,francesca.rissom,henrike.heyne,bernhard.renard}@hpi.de
[2] Bioinformatics and Machine Learning, Broad Institute of MIT and Harvard, USA
{jsafer,sumaiya}@broadinstitute.org
[3] School of Business and Economics, Humboldt-Universität zu Berlin, Germany
manuel.pfeuffer@hu-berlin.de
[4] Scientific Computing Center, Karlsruhe Institute of Technology, Germany
nadja.klein@kit.edu

Keywords: protein language models · embeddings · representation learning · interpretability · semi-structured neural networks

Extended Abstract

Similar to large language models in natural language processing, protein language models (PLMs) are trained in a task-agnostic, self-supervised manner on large datasets of protein sequences. Their embeddings have been shown to outperform traditional methods on a range of downstream tasks in computational biology [9]. In some cases, performance is further improved by models combining PLM embeddings with tabular biological features [2].

However, both approaches lack interpretability which is crucial for deriving biological insights and may be required for medical applications. The embeddings – numerical representations of the protein sequence – are not directly interpretable. When adding biological features to the model, it is unclear what actually drives the prediction as these features and the embeddings might be correlated.

To address this issue, we train a semi-structured neural network (SSN) with post-hoc orthogonalization [7]. Thereby, embeddings and tabular features are combined into a single model, and linear effects of the tabular features are removed from the embeddings by an orthogonal projection. SSNs have been used in a similar setting to remove tabular features from X-ray image embeddings [10]. In our work, we focus on predicting functional effects of mutated protein sequences – specifically, whether amino acid substitutions impact protein functions such as activity, binding, or stability. For our analyses, we use pre-trained

PLMs [6, 4, 1] and experimental benchmark datasets of protein sequences [5, 8]. Additional biological features are collected from public databases [3].

Our findings reveal that several tabular features significantly impact predictions based on PLM embeddings. Moreover, our ablation studies show that removing linear effects of these features from the embeddings decreases their predictive power. In further analyses, we investigate how fine-tuning the PLMs on a specific downstream task affects the resulting embeddings with respect to their correlation with biological features [8].

References

1. Elnaggar, A., Heinzinger, M., Dallago, C., Rehawi, G., Wang, Y., Jones, L., Gibbs, T., Feher, T., Angerer, C., Steinegger, M., et al.: Prottrans: Toward understanding the language of life through self-supervised learning. IEEE transactions on pattern analysis and machine intelligence **44**(10), 7112–7127 (2021)
2. Hsu, C., Nisonoff, H., Fannjiang, C., Listgarten, J.: Learning protein fitness models from evolutionary and assay-labeled data. Nature biotechnology **40**(7), 1114–1122 (2022)
3. Kwon, S., Safer, J., Nguyen, D.T., Hoksza, D., May, P., Arbesfeld, J.A., Rubin, A.F., Campbell, A.J., Burgin, A., Iqbal, S.: Genomics 2 Proteins portal: a resource and discovery tool for linking genetic screening outputs to protein sequences and structures. Nature Methods pp. 1–11 (2024)
4. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al.: Evolutionary-scale prediction of atomic-level protein structure with a language model. Science **379**(6637), 1123–1130 (2023)
5. Notin, P., Kollasch, A., Ritter, D., Van Niekerk, L., Paul, S., Spinner, H., Rollins, N., Shaw, A., Orenbuch, R., Weitzman, R., et al.: Proteingym: Large-scale benchmarks for protein fitness prediction and design. Advances in Neural Information Processing Systems **36** (2024)
6. Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., Guo, D., Ott, M., Zitnick, C.L., Ma, J., et al.: Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proceedings of the National Academy of Sciences **118**(15), e2016239118 (2021)
7. Rügamer, D.: A new PHO-rmula for improved performance of semi-structured networks. In: International Conference on Machine Learning. pp. 29291–29305. PMLR (2023)
8. Schmirler, R., Heinzinger, M., Rost, B.: Fine-tuning protein language models boosts predictions across diverse tasks. Nature Communications **15**(1), 7407 (2024)
9. Valentini, G., Malchiodi, D., Gliozzo, J., Mesiti, M., Soto-Gomez, M., Cabri, A., Reese, J., Casiraghi, E., Robinson, P.N.: The promises of large language models for protein design and modeling. Frontiers in Bioinformatics **3**, 1304099 (2023)
10. Weber, T., Ingrisch, M., Bischl, B., Rügamer, D.: Post-hoc Orthogonalization for Mitigation of Protected Feature Bias in CXR Embeddings. arXiv preprint arXiv:2311.01349 (2023)

Tackling Data Scarcity in Machine Learning: Leveraging Symbolic Regression for Data Generation

Matthias Kuntz[0009−0000−6593−3141], Torben Weis[0000−0001−6594−326X]

Chair of Distributed Systems, Faculty of Computer Science
University of Duisburg-Essen, Germany
{matthias.kuntz, torben.weis}@uni-due.de

Abstract. A key limitation of many machine learning (ML) methods is their strong dependence on large amounts of data. In contrast, the strength of symbolic regression (SR) is its ability to extract interpretable models even from small amounts of data. However, given the constrained complexity of the resulting symbolic expressions, SR may encounter difficulties in accurately capturing the intricate structures present within a data set, unlike traditional ML methods. This research seeks to combine the advantages of both approaches by using SR to generate synthetic data, which can then be used to train and enable other ML methods in environments where available data is limited. This approach enables us to bootstrap a non-SR ML model and fine-tune it during deployment as new non-synthetic data becomes available.
Our research will investigate multiple aspects:

1. Which type of traditional ML model works best for our approach?
2. The optimal ratio of real vs. synthetic data.
3. The importance of the complexity of the trained SR model that is used to synthesize the data.
4. The role of transfer learning, leveraging a model trained in a data-rich domain to improve performance in data-scarce settings when combined with our proposed approach.

The long-term objective is to design efficient learning systems that can cope with data scarcity, evolving environments, and resource limitations, that can eventually support online learning through SR-facilitated knowledge transfer.

Keywords: machine learning · data scarcity · symbolic regression · synthetic data.

Modelling of Forced Dynamical Systems with Laplace-Net

Bernd Zimmering[1 [0000−0001−8292−9945]], Cecília Coelho[1 [0009−0009−4502−937X]], and Oliver Niggemann[1 [0000−0001−8747−3596]]

Institute for Artificial Intelligence, Helmut Schmidt University Hamburg, Germany
{bernd.zimmering, cecilia.coelho, oliver.niggemann}@hsu-hh.de

The field of Scientific Machine Learning (SciML) is fast evolving, with Machine Learning (ML) algorithms used to either: (i) learn differential equations from data, such as Neural Ordinary Differential Equations (NODEs) [5]; or (ii) bypass compute-intensive solvers, as in Physics-Informed Neural Networks (PINNs) [4]. While PINNs are mainly applied in settings where fully known equations are available, many engineering systems lack complete models and suffer from limited data availability. NODEs, and related continuous-time neural models, are attractive in engineering due to their alignment with classical modelling [8], but standard NODEs do not capture essential non-local effects such as memory or delays.

To address this, [7] introduces an improved solver for Neural Fractional Differential Equations (NFDEs) [2], which can model arbitrary memory effects but are still limited regarding delays and require expensive solvers. Laplace-Net [6] overcomes these challenges with a modular, solver-free architecture that operates in the Laplace domain. It explicitly separates system dynamics from external inputs and initial conditions. This structure mirrors established principles from control engineering (e.g. transfer functions), enabling interpretation of the model components by domain experts and adaptation to unseen input signals. Furthermore, it can learn delays and non-local behaviour similar to NFDEs. Laplace-Net outperforms Laplace Neural Operators [1] and Long Short-Term Memory networks [3] in six out of eight datasets from different dynamical systems.

Despite these advances, there is a lack of systematic comparison regarding the impact of prior knowledge on model performance and computational cost in real engineering scenarios. Previous work [9] showed that NODEs can directly integrate prior knowledge, reducing data requirements—a key advantage when measurements are costly or scarce. Future research in this work plan will benchmark NODEs and Laplace-Net against state-of-the-art algorithms on engineering-relevant tasks, quantifying computational effort, data efficiency, and practical applicability. The goal is to provide actionable guidance for practitioners and assess the robustness of these models' performance in real-world engineering contexts.

References

1. Cao, Q., Goswami, S., Karniadakis, G.E.: Laplace neural operator for solving differential equations. Nature Machine Intelligence **6**(6), 631–640 (Jun 2024).

https://doi.org/10.1038/s42256-024-00844-4
2. Coelho, C., Costa, M.F.P., Ferrás, L.L.: Neural Fractional Differential Equations: Optimising the Order of the Fractional Derivative. Fractal and Fractional **8**(9), 529 (Sep 2024). https://doi.org/10.3390/fractalfract8090529
3. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–80 (Jan 1997). https://doi.org/10.1162/neco.1997.9.8.1735
4. Raissi, M., Perdikaris, P., Karniadakis, G.E.: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics **378**, 686–707 (Jan 2019). https://doi.org/10.1016/j.jcp.2018.10.045
5. Ricky T. Q., C., Yulia, R., Jesse, B., David, D.: Neural Ordinary Differential Equations. In: NeurIPS (Jan 2018)
6. Zimmering, B., Coelho, C., Gupta, V., Maleshkova, M., Niggemann, O.: Breaking free: Decoupling forced systems with laplace neural networks. In: Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD). Porto, Portugal (2025)
7. Zimmering, B., Coelho, C., Niggemann, O.: Optimising neural fractional differential equations for performance and efficiency. In: Proceedings of the 1st ECAI Workshop on "Machine Learning Meets Differential Equations: From Theory to Applications". Proceedings of Machine Learning Research, vol. 255, pp. 1–22. PMLR (Oct 2024)
8. Zimmering, B., Niggemann, O.: Integrating continuous-time neural networks in engineering: Bridging machine learning and dynamical system modeling (Jan 2024). https://doi.org/10.24405/15313
9. Zimmering, B., Roche, J.P., Niggemann, O.: Enhancing Nonlinear Electrical Circuit Modeling with Prior Knowledge-Infused Neural ODEs. In: 2024 IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA). pp. 01–08. IEEE, Padova, Italy (Sep 2024). https://doi.org/10.1109/ETFA61755.2024.10711112

Introducing ISCNNs: Input Smooth Convex Neural Networks

Qinyan Zhou[1] and Manh Hung Le[1] and Andrea Simonetto[1]

Unité de Mathématiques Appliquées, ENSTA, Institut Polytechnique de Paris, 91120 Palaiseau, France

Abstract. In a wide range of scientific and engineering applications, such as economics, operations research, optimal transport, and energy systems, it is essential to approximate unknown convex functions based on limited data. Traditional convex regression methods, including convex programs [3], approximate kernel-based techniques [2, 6], as well as piecewise-linear fitting approaches [4], often have poor scalability in high-dimensional spaces, may recover convex functions only with some errors, and may not have enough flexibility to recover the complicated convex behavior seen in real datasets.

Input Convex Neural Networks (ICNNs) [1] provide a promising alternative by encoding convexity directly into the architecture of neural networks. By enforcing weight constraints and adopting convexity-preserving activation functions, ICNNs guarantee convexity with respect to the input. Moreover, ICNN maintains the expressiveness and scalability of deep learning, making it an attractive choice for developing data-driven convex modeling in complex and high-dimensional spaces.

In this work, we introduce Input Smooth Convex Neural Networks (ISCNNs), a variant of ICNNs whose outputs are provably Lipschitz continuous. We focus on a previously underexplored property of ICNNs: the Lipschitz continuity of their output. Smooth convex functions are especially important in contexts where second-order information affects optimization performance. We investigate how the choice of activation function impacts the second-order smoothness of the ICNN output. In particular, we demonstrate that the widely used SoftPlus activation function produces significantly smoother convex functions compared to popular non-smooth alternatives such as ReLU.

By comparing numerical experiments with traditional non-smooth convex regression and smooth convex regression methods [7], we found that although smooth convex regression achieves the lowest approximation error in low-dimensional settings, its scalability is limited due to high computational cost. In contrast, Softplus-based ISCNN maintains a highly competitive accuracy in high-dimensional problems while showing significant computational advantages. We also experimented with a different activation function, which seems even more promising.

Future work will explore the use of ISCNNs in learning how to regularise in inverse problems arising from partial differential equations, as proposed in [5].

Keywords: Input Convex Neural Networks · Lipschitz Continuous ICNNs · Smooth Convex Regression

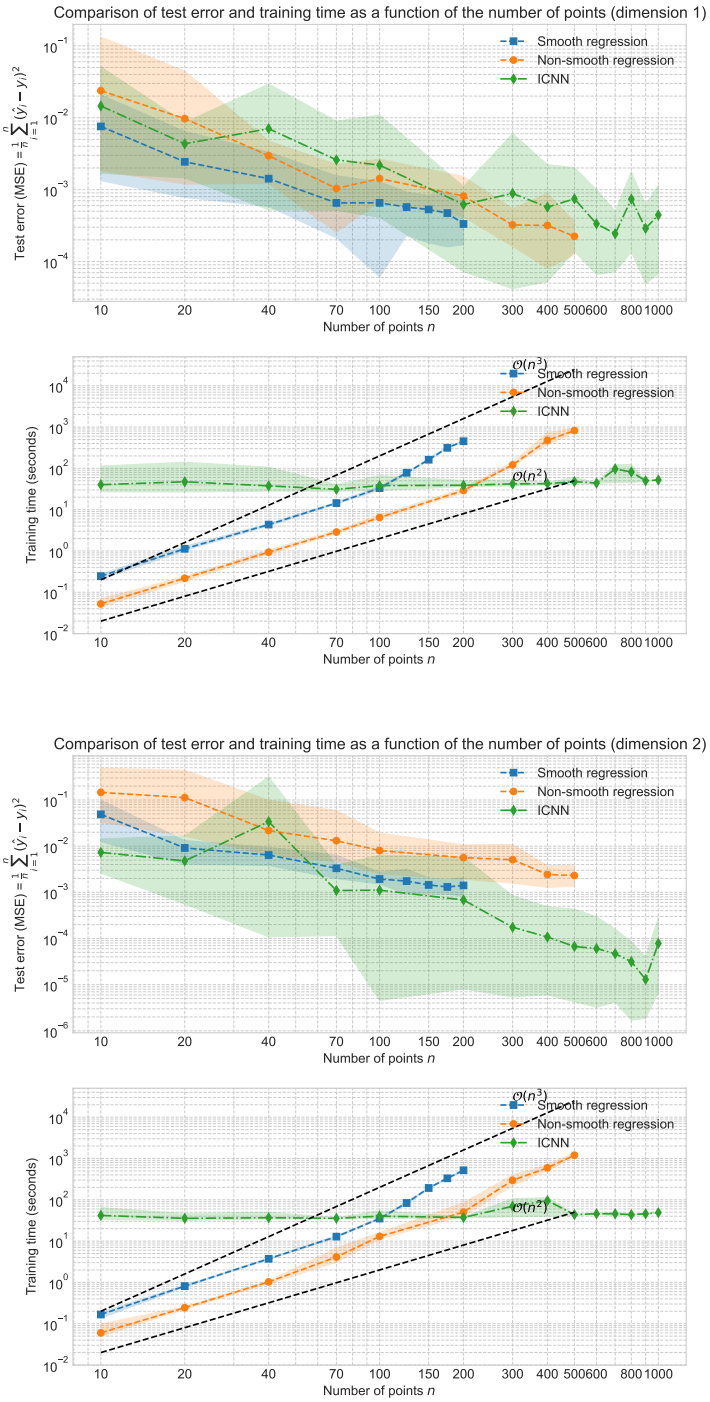


Fig. 1: Test error and training time versus the number of points for dimensions 1 and 2. Results are averaged over 10 independent realisations. In each plot, the shaded region represents the range between the minimum and maximum values, while the lines with markers correspond to the average values. The test error is measured using the Mean Squared Error (MSE), defined as MSE $= \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$, where $\hat{y}_i$ is the predicted output and y_i denotes the corresponding true value. In our setting, the true value is given by $y_i = \|x_i\|_2^2$.

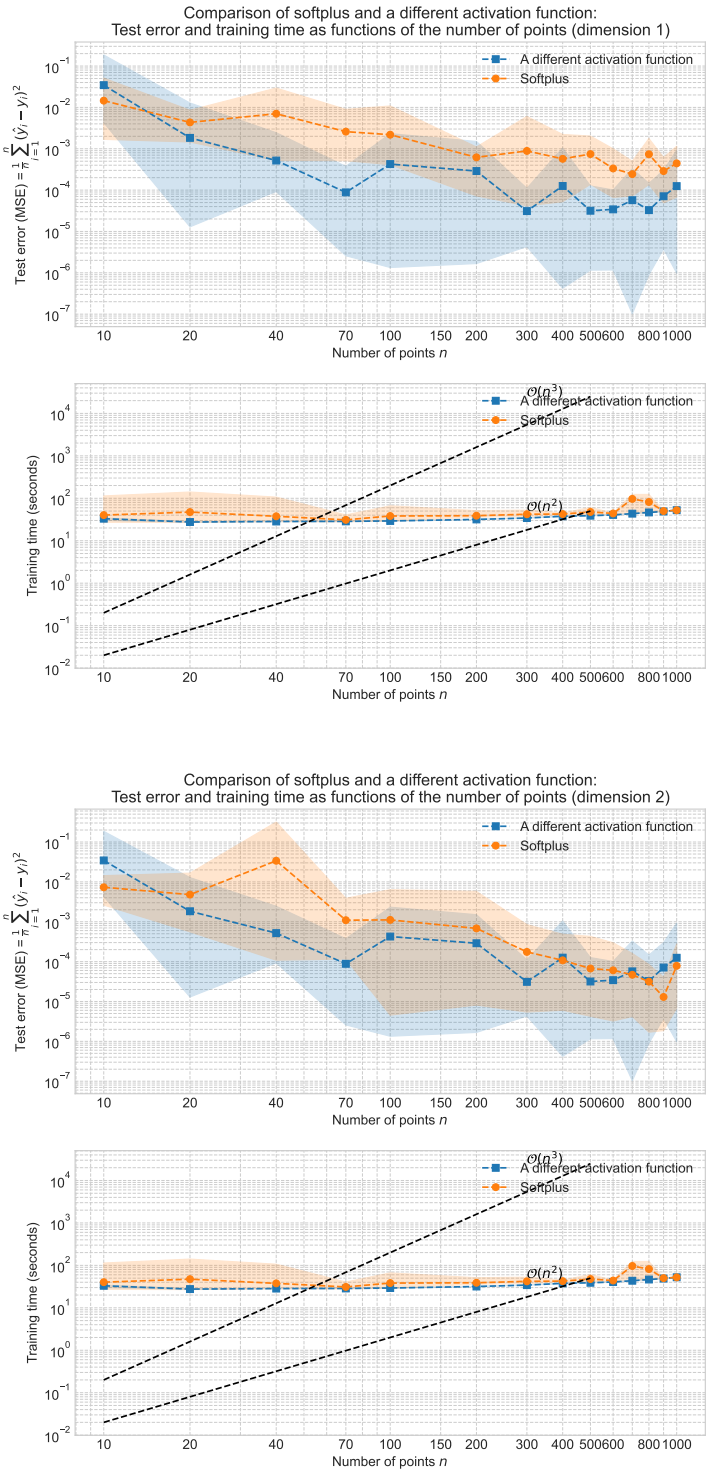


Fig. 2: Comparison between Softplus and a different activation function (0 for $x < 0$ and $\frac{1}{2}x^2$ for $x \geq 0$: test error and training time versus number of points for dimensions 1 and 2. Results are averaged over 10 independent realisations. In each plot, the shaded region represents the range between the minimum and maximum values, while the lines with markers correspond to the average values.

References

1. Amos, B., Xu, L., Kolter, J.Z.: Input convex neural networks. In: International conference on machine learning. pp. 146–155. PMLR (2017)
2. Aubin-Frankowski, P.C., Szabó, Z.: Hard shape-constrained kernel machines. Advances in neural information processing systems **33**, 384–395 (2020)
3. Chen, W., Mazumder, R.: Subgradient regularized multivariate convex regression at scale. SIAM Journal on Optimization **34**(3), 2350–2377 (2024)
4. Magnani, A., Boyd, S.P.: Convex piecewise-linear fitting. Optimization and Engineering **10**(1), 1–17 (2009). https://doi.org/10.1007/s11081-008-9045-3, https://doi.org/10.1007/s11081-008-9045-3
5. Mukherjee, S., Dittmer, S., Shumaylov, Z., Lunz, S., Öktem, O., Schönlieb, C.B.: Learned convex regularizers for inverse problems. arXiv preprint arXiv:2008.02839 (2020)
6. Muzellec, B., Bach, F., Rudi, A.: Learning psd-valued functions using kernel sums-of-squares. arXiv preprint arXiv:2111.11306 (2021)
7. Simonetto, A.: Smooth strongly convex regression. In: 2020 28th European Signal Processing Conference (EUSIPCO). pp. 2130–2134. IEEE (2021)

High-Level Semantic Understanding in Multimodal Video Content

PhD Student: Oron Nir[1][0000−0002−8443−5014]
Supervising Prof.: Ariel Shamir[1][0000−0001−7082−7845]

CANVAS Lab, Reichman University, Herzeliya, Israel, arik@runi.ac.il

1 Extended Abstract

My PhD research explores high-level semantic understanding of video content, particularly focusing on multimodal analysis and non-photorealistic video styles, such as animation. Given the exponential growth of video content, effective understanding and retrieval necessitate sophisticated, scalable methods that surpass traditional low-level recognition tasks.

My work titled "*CAST: Character labeling in Animation using Self-supervision by Tracking*," [1], addresses the challenge of representing stylistically diverse animated characters. Using novel SSL by tracking approach, CAST refines semantic embeddings specifically tailored for animation styles. This facilitates efficient character ReID and clustering within animated content, overcoming limitations inherent in standard vision models trained on natural imagery. A follow-up work was presented at IEEE VIS [2].

In my VCR [3] paper, I developed a multimodal framework for contextual retrieval that captures high-level semantic topics in longform videos, specifically focusing on educational videos by integrating ASR, OCR, and frame captioning into a unified semantic representation using LLMs.

My recent ECML-PKDD accepted paper on "*Unimodal Strategies for Density-Based Clustering*" [4] expands on ideas initially explored in CAST. It uncovers key properties of DBSCAN and leverages them to optimize hyperparameter search, improving clustering effectiveness.

Currently, I'm integrating these concepts into a person-centric Video Anomaly Detection system aimed at public safety. The challenge is calibrating scene-specific normal behavior by combining person ReID, contrastive SSL (CAST), and unsupervised domain adaptation to dynamically adjust VLM priors.

My interdisciplinary research spans CV, NLP, ML, and IR, intersecting several ECML-PKDD communities. I seek feedback on technical innovation strategies in crowded domains, approaches to making video content more searchable, and guidance on future career milestones.

References

1. Nir, O., Rapoport, G., Shamir, A.: CAST: Character labeling in Animation using Self-supervision by Tracking. In: Computer Graphics Forum, vol. 41, no. 2, pp. 135–145 (2022)

2. Nir, O., Neeman, A., Shamir, A.: Visualizing the CAST of Characters Labeled in Animation (Poster). In: IEEE VIS (2022)
3. Nir, O., Vidra, I., Neeman, A., Kinarti, B., Shamir, A.: VCR: Video Representation for Contextual Retrieval. In: Proceedings of the International Conference on Computing, Machine Learning and Data Science, pp. 1–9 (2024)
4. Nir, O., Tenenbaum, J., Shamir, A.: Unimodal Strategies in Density-based Clustering. In: ECML-PKDD (2025, accepted)
5. Cohen, N.Z., Nir, O., Shamir, A.: Conditional Balance: Improving Multi-Conditioning Trade-Offs in Image Generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2025)
6. Nir, O., Parmet, Y., Werner, D., Adin, G., Halachmi, I.: 3D Computer-vision System for Automatically Estimating Heifer Height and Body Mass. In: Biosystems Engineering, vol. 173, pp. 4–10 (2018)
7. Wang, C.Y., Nir, O., Vemprala, S., Kapoor, A., Ofek, E., McDuff, D., Gonzalez-Franco, M.: CityLifeSim: A High-Fidelity Pedestrian and Vehicle Simulation with Complex Behaviors. In: IEEE 2nd International Conference on Intelligent Reality (ICIR), pp. 11–16. IEEE (2022)
8. Ronen, R., Berezin, H., Preizler, R., Kasturi, G., Ezzour, A.J., Bhanavase, S., Nir, O.: Generic Automated Lead Ranking in Dynamics CRM. In: Proceedings of the 15th ACM Conference on Recommender Systems, pp. 757–759 (2021)
9. Vardi, B., Nir, O., Shamir, A.: CLIP-UP: CLIP-Based Unanswerable Problem Detection for Visual Question Answering. arXiv preprint arXiv:2501.01371 (2025, under submission)

Towards Efficient Model Explainability in Low-Latency Machine Learning Systems

Inês Oliveira e Silva[1,3] (✉), Sérgio Jesus[1,3], Iker Perez Lopez[3], Carlos Soares[2], and Rita P. Ribeiro[1]

[1] Faculdade de Ciências da Universidade do Porto
[2] Faculdade de Engenharia da Universidade do Porto
[3] Feedzai, Porto, Portugal
{ines.silva,iker.perez}@feedzai.com

Extended Abstract

Explainability is essential for trustworthy AI [1], yet Shapley-based attribution methods [6, 5] are often computationally intensive for deployment in high-throughput, low-latency settings [3]. This PhD research addresses the need for efficient, real-time explainability techniques that preserve interpretability and usability.

We focus on combining amortised and prototype-based explainability to enable fast, practical model understanding [3, 2]. Prototype-based methods [4], often used in domains such as financial transaction monitoring, offer human-interpretable, example-driven explanations that support regulatory compliance. However, generating and caching dense sets of prototypes and attributions is challenging in high-dimensional spaces, due to memory constraints and the complexity of covering the data manifold effectively.

We propose latent structuring and prototype selection around the geometry of attribution outputs instead of raw inputs. This enables meaningful interpolation in latent spaces, improving the utility of cached explanations. Our focus is on sequential tabular data, such as event streams in financial applications, where attribution-aware compression, caching and nearest-neighbour retrieval can support low-latency explanations.

We build on amortisation approaches (e.g. FastSHAP [3]), using autoencoders to learn mappings from inputs to Shapley values while enforcing latent structure aligned with explainability goals. Since Shapley attributions are deterministic for a fixed model, we hypothesise that expressive models can serve as effective interpolation engines, forming stable grids over which prototype instances can be selected or generated.

This research aims to design fast and faithful explainability pipelines for real-world, regulated settings, and balancing computational efficiency with human-centred design. Evaluation will involve collaboration with domain experts in financial fraud detection to ensure practical relevance and assess usability in real-world settings.

References

1. S. Ali, T. Abuhmed, S. El-Sappagh, K. Muhammad, J. M. Alonso-Moral, R. Confalonieri, R. Guidotti, J. Del Ser, N. Díaz-Rodríguez, and F. Herrera. Explainable artificial intelligence (xai): What we know and what is left to attain trustworthy artificial intelligence. *Information fusion*, 99:101805, 2023.
2. I. Covert, C. Kim, S.-I. Lee, J. Zou, and T. Hashimoto. Stochastic amortization: A unified approach to accelerate feature and data attribution. *arXiv preprint arXiv:2401.15866*, 2024.
3. N. Jethani, M. Sudarshan, I. C. Covert, S.-I. Lee, and R. Ranganath. Fastshap: Real-time shapley value estimation. In *International conference on learning representations*, 2021.
4. O. Li, H. Liu, C. Chen, and C. Rudin. Deep learning for case-based reasoning through prototypes: A neural network that explains its predictions. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.
5. S. M. Lundberg, G. G. Erion, and S.-I. Lee. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*, 2018.
6. S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

Robust Evaluation of Fairness and Explainability in AI Systems

Sérgio Jesus[1,4] (✉), Pedro Saleiro[2], Rita P. Ribeiro[1,3], and João Gama[3]

[1] Faculdade de Ciências da Universidade do Porto
[2] Opnova, Porto, Portugal
[3] LIAAD, INESCTEC Universidade do Porto
[4] Feedzai, Porto, Portugal
sergio.jesus@feedzai.com

Extended Abstract

The increasing deployment of AI systems in high-stakes domains requires rigorous assessment of properties such as fairness and explainability, which represent core pillars of the area of Responsible AI. However, current evaluation practices remain fragmented and unstandardized, and often unadapted to real-world constraints. This research project addresses these challenges by proposing robust, application-centred evaluation methodologies that align with practical deployment settings and respond to future regulatory expectations.

In the domain of fairness, we introduce a novel stress-testing framework that systematically injects bias patterns into datasets to uncover fairness vulnerabilities in ML models [3]. This methodology underpins the Bank Account Fraud (BAF) suite [6], a set of realistic benchmarks based on real-world fraud detection data. These benchmarks enable more complete studies of fairness-performance trade-offs across algorithms. To support reproducible experimentation, we developed Aequitas Flow [7], an open-source package for fairness-aware model evaluation that integrates metrics, visualizations, and hyperparameter optimization methods [2].

For explainability, we propose XAI Test [5], a methodology inspired by randomized controlled trials (RCTs) [8], to quantify the real-world utility of explanations. Two large-scale experiments involving professional fraud analysts evaluate the impact of different explanation techniques on decision accuracy, speed, and perceived usefulness [4, 1]. Findings reveal that while explanations can improve decision speed, their effects on accuracy vary significantly by method, highlighting the importance of empirical, human-centred validation.

Together, these contributions support a unified framework for evaluating AI systems under realistic, multi-objective constraints, bridging theoretical desiderata and practical needs. Future work will explore evaluations tailored to emergent models and applications, such as LLMs, and strategies for the online selection and adaptation of explainability solutions.

References

1. K. Amarasinghe, K. T. Rodolfa, S. Jesus, V. Chen, V. Balayan, P. Saleiro, P. Bizarro, A. Talwalkar, and R. Ghani. On the importance of application-grounded experimental design for evaluating explainable ml methods. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(19):20921–20929, Mar. 2024.
2. A. Cruz, C. G. Belém, S. Jesus, J. Bravo, P. Saleiro, and P. Bizarro. FairGBM: Gradient boosting with fairness constraints. In *The Eleventh International Conference on Learning Representations*, 2023.
3. S. Jesus, D. Alves, J. Pombal, A. Cruz, J. Veiga, J. Ferreira, C. Belém, M. Sampaio, P. Saleiro, and P. Bizarro. Obtaining a generated dataset with a predetermined bias for evaluating algorithmic fairness of a machine learning model, Aug. 22 2023. US Patent 11,734,612.
4. S. Jesus, C. Belém, V. Balayan, J. a. Bento, P. Saleiro, P. Bizarro, and J. a. Gama. How can i choose an explainer? an application-grounded evaluation of post-hoc explanations. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, page 805–815, New York, NY, USA, 2021. Association for Computing Machinery.
5. S. Jesus, C. Belém, V. Balayan, D. Polido, J. Sousa, J. Carvalhais, A. M. Ruela, M. Almeida, P. Saleiro, and P. Bizarro. Human-in-the-loop evaluation for explainable artificial intelligence, Jan. 27 2022. US Patent App. 17/382,075.
6. S. Jesus, J. Pombal, D. Alves, A. Cruz, P. Saleiro, R. P. Ribeiro, J. Gama, and P. Bizarro. Turning the tables: Biased, imbalanced, dynamic tabular datasets for ml evaluation. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 33563–33575. Curran Associates, Inc., 2022.
7. S. Jesus, P. Saleiro, I. O. e Silva, B. M. Jorge, R. P. Ribeiro, J. Gama, P. Bizarro, and R. Ghani. Aequitas flow: Streamlining fair ml experimentation. *Journal of Machine Learning Research*, 25(354):1–7, 2024.
8. R. Kohavi, R. Longbotham, D. Sommerfield, and R. M. Henne. Controlled experiments on the web: survey and practical guide. *Data mining and knowledge discovery*, 18(1):140–181, 2009.

EileForecast: An Open-Source Forecasting Module for Energy Management in SMEs

Claudia Ehrig[1,2], Ursula Neumann[1], Catherine Cleophas[2], and Germain Forestier[3]

[1] Fraunhofer IIS, Fraunhofer Institute for Integrated Circuits IIS, Fraunhofer Center for Applied Research on Supply Chain Services SCS, 90411 Nuremberg, Germany
[2] Kiel Univeristy, Research Group Service Analytics, 24098 Kiel, Germany
[3] University of Upper Alsace, IRIMAS, 68 093 Mulhouse, France

1 Background

Small and Medium-sized Enterprises (SMEs) represent over 99% of European businesses [9]. Therefore, increasing their energy efficiency is highly relevant in the pursuit of climate goals. While Energy Management Systems (EMS) offer a tool to monitor, analyse and optimise energy consumption and production, only a fraction of SMEs use them [10].

2 Problem

Although open-source EMS provide a cost-effective solution, the internal forecasting methods (see internal tools in Table 1) have not yet been evaluated for their usage within EMS, i.e., on individual industrial load time series. This is also true for methods in publicly available software packages for Short-Term Load Forecasting (STLF), which can be integrated into EMS (see external tools in Table 1). As forecasts form the basis for decision-making regarding efficiency measures (see Table 2), their reliability and accuracy are essential for EMS.

3 Contribution

To motivate SMEs to increase their energy efficiency, we design the open-source STLF Python package `EileForecast`, which offers methods evaluated on SME data and can be integrated into EMS. We show the effectiveness of the package in a case study with a German SME.

4 Preliminary Benchmark Results

Starting point for the package design is a benchmark of STLF methods provided by existing EMS, external packages, and academic literature on 53 public industrial SME datasets from the Open Power System Data dataset `CoSSMic` [12] and

Table 1. Forecasting possibilities of open-source EMS (internal, I) and open-source energy forecasting packages (external, E).

Tool	I/E	Functionalities	Models
OpenEMS [5]	I	Automatic retraining	Seasonal Naive, Seasonal Average, LSTM
OPEN [13]	I	-	Moving Average
EMHASS [4]	I	HO, evaluation	Seasonal Naive, Seasonal Average, Linear Regression, KNN Regression, Elastic Net, more regression models
OpenSTEF [5]	E	Transfer learning, feature integration & selection, example notebooks	Arima, Linear Regression, XGBoost, LightGBM, Quantile Regression
ProLoaf [8]	E	HO, example notebooks	Seasonal Naive, Exponential Smoothing, Arima, Garch, Encoder-Decoder NN
Energy Service Generics Framework[14]	E	Embed any model as a service	Any
EileForecast [ours]	E	HO, example notebooks, benchmarking, foundation models	Most accurate models of a thorough benchmark on STLF

Table 2. Energy efficiency measures and corresponding forecasts in industrial case studies.

Study	Goal	Decision	Target Series	Forecast Methods	Results
[6]	Consistent consumption	Production scheduling	Electricity demand of single processes	Expert system	16% peak reduction, 20% reduction of standard deviation, and mean/max factor increase over 20% for one day
[7]	Identify inefficiencies of fans	Reduce fans	Fan absorption	FFNNs	Potential savings of 140 MWh in three years
[11]	Avoid peaks	Peak alert	Overall load	Linear model	Recall of up to 75.4% for 1h-ahead forecast and 86.9% for 24h interval forecast
[2]	Reduce peaks and contracted power	Scheduling workdays and production	Overall consumption and single machines	Fuzzy NN	Forecasts were satisfactory for the authors

the `LoadProfile` [1] datasets. As SMEs are just starting to use EMS, we evaluate the models for increasing data availability, from one week up to one year. Figure 1 shows the superiority of the time series foundation model TimesFM [4] and the tree-based model XGBoost [3] among the five models already evaluated.

5 EileForecast

We provide the models from the benchmark that perform consistently well across the datasets in our module `EileForecast`. The latter will be published as open-source software and lends itself to diverse use cases by forecasting various target series, set in the configuration file. It consists of two Docker containers, which can be integrated within the data and EMS infrastructure; see Figure 2.

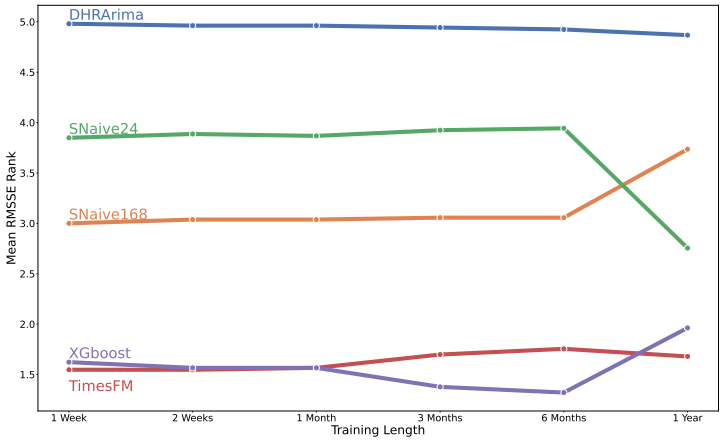


Fig. 1. Mean RMSSE ranks of the models over the datasets for increasing data availability and the hourly 24-step forecast. Lower is better. Analogue results for the metric of MASE.

References

1. Braeuer, F.: Load profile data of 50 industrial plants in germany for one year. Zenodo (2020)
2. Cárdenas, J.J., Romeral, L., Garcia, A., Andrade, F.: Load forecasting framework of electricity consumptions for an intelligent energy management system in the user-side. Expert Systems with Applications **39**(5), 5557–5565 (2012)
3. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. pp. 785–794 (2016)

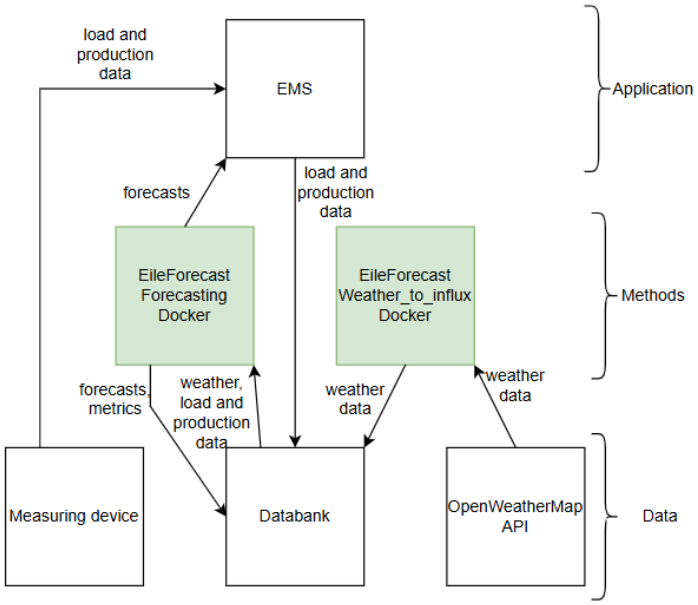


Fig. 2. EileForecast provides two self-contained Docker containers (green): Weather_to_influx, which pulls weather data from OpenWeatherMap (https://openweathermap.org/api) and writes it to an Influx database, and Forecasting, which pulls weather, load and production data from the database and creates forecasts for the EMS.

4. Das, A., Kong, W., Sen, R., Zhou, Y.: A decoder-only foundation model for time-series forecasting. In: Forty-first International Conference on Machine Learning (2024)
5. Feilmeier, S., wgerbl, ebakir, fabianfnc, Sahutoglu, H., pooran c, Obermeier, M., Asen, S., Sagar, H., Rieger, L., da Kai, Ihara, H., Kyle, luzpaz, S, F., Sicking, T., Lehne, C., dependabot support, greemo.: Openems/openems: 2025.5.0 (2025.5.0). Zenodo. https://doi.org/10.5281/zenodo.15313038 (2025), accessed: 2025-05-04
6. Grigoraș, G., Neagu, B.C.: An advanced decision support platform in energy management to increase energy efficiency for small and medium enterprises. Applied Sciences **10**(10), 3505 (2020)
7. Grimaccia, F., Niccolai, A., Mussetta, M., D'Alessandro, G.: Iso 50001 data driven methods for energy efficiency analysis of thermal power plants. Applied Sciences **13**(3), 1368 (2023)
8. Gürses-Tran, G., Oppermann, F., Monti, A.: Proloaf: Probabilistic load forecasting for power systems. SoftwareX **23**, 101487 (2023)
9. Katsinis, A.A., Lagüera, G.J., Di, B.L., Odenthal, L., Hell, M., Lozar, B., et al.: Annual report on european smes 2023/2024, sme performance review. Publications Office of the European Union (2024)
10. Knayer, T., Kryvinska, N.: The influence of energy management systems on the progress of efficient energy use in cross-cutting technologies in companies. Energy Efficiency **16**(3), 12 (2023)
11. Komatsu, H., Kimura, O.: Peak demand alert system based on electricity demand forecasting for smart meter data. Energy and Buildings **225**, 110307 (2020)
12. Minde, A.: Open power system data. 2020. data package household data. Version 2020-04-15. : $//data.open-power-system-data.org/household_data/2020-04-15/$. (2017)
13. Morstyn, T., Collett, K.A., Vijay, A., Deakin, M., Wheeler, S., Bhagavathy, S.M., Fele, F., McCulloch, M.D.: Open: An open-source platform for developing smart local energy system applications. Applied Energy **275**, 115397 (2020)
14. Wölfle, D., Förderer, K., Riedel, T., Landwich, L., Mikut, R., Hagenmeyer, V., Schmeck, H.: Open energy services–forecasting and optimization as a service for energy management applications at scale. arXiv preprint arXiv:2402.15230 (2024)

Toward Personalized Career Recommendations Using JobHop

Iman Johary, Raphael Romero, Alexandru Mara, and Tijl De Bie

Department of Electronics and Information Systems, Ghent University, Belgium

1 Extended Abstract

In future work, I plan to develop machine learning based methods for personalized career path recommendations. These methods will leverage both historical career patterns and individual characteristics to generate adaptive, goal-oriented career guidance. By bridging dataset creation and predictive modeling, this research aims to contribute to scalable, equitable, and personalized labor market tools.

The goal of my PhD research is to develop data-driven methods for career path recommendation, helping individuals navigate complex labor markets and make informed career decisions. As a foundational step toward this objective, I developed JobHop [1], a large-scale, publicly available dataset derived from anonymized resumes provided by VDAB, the public employment service of Flanders, Belgium. JobHop is a novel dataset built from 442,000 anonymized resumes. These resumes, primarily in Dutch but also in English and French, were preprocessed using a two-step pipeline. First, I employed large language models (LLMs) for structural information extraction, transforming unstructured text into structured JSON records containing job titles, companies, employment periods, and educational qualifications. Second, I implemented a data normalization step, mapping the extracted job titles to standardized ESCO occupation codes using both a baseline embedding-based similarity pipeline and a commercial classifier. This process resulted in a temporally annotated ESCO dataset comprising over 2.3 million work experiences. JobHop fills a gap in career modeling research by providing a longitudinal, structured, and publicly accessible resource for analyzing career dynamics. It enables large-scale studies of occupational mobility and education-to-employment pathways. Beyond its research utility, it can also serve as a benchmark dataset for training and evaluating algorithms in job matching and career forecasting. By addressing the scarcity of publicly available datasets with detailed occupational trajectories, JobHop supports downstream applications such as career path prediction and skill gap analysis. In future work, I plan to develop machine learning based methods for personalized career path recommendations. These methods will leverage both historical career patterns and individual characteristics to generate adaptive, goal-oriented career guidance. By bridging dataset creation and predictive modeling, this research aims to contribute to scalable, equitable, and personalized labor market tools.

Keywords: Career Path Recommendation · Large Language Models · ESCO classification

References

1. Johary, I., Romero, R., Mara, A.C., Bie, T.D.: Jobhop: A large-scale dataset of career trajectories (2025), https://arxiv.org/abs/2505.07653

Regression-Based Estimation of Causal Effects in the Presence of Selection Bias and Confounding

Marlies Hafer[1,2] and Alexander Marx[1,2]

[1] Department of Statistics, TU Dortmund University
[2] Research Center Trustworthy Data Science and Security, University Alliance Ruhr
{marlies.hafer,alexander.marx}@tu-dortmund.de

We consider the problem of estimating the expected causal effect $E[Y \mid do(X)]$ for a target variable Y when treatment X is set by intervention, focusing on continuous random variables. In settings without selection bias or confounding, $E[Y \mid do(X)] = E[Y \mid X]$, which can be estimated using standard regression methods. However, regression fails when systematic missingness induced by selection bias, or confounding distorts the data. Boeken et al. [2023] show that when training data is subject to selection, proxy variables unaffected by this process can, under certain constraints, be used to correct for selection bias to recover $E[Y \mid X]$, and hence $E[Y \mid do(X)]$, reliably. When data is additionally affected by confounding, however, this equality is no longer valid. In this work, we consider a more general setting and propose a framework that incorporates both selection bias and confounding, as illustrated in Figure 1 based on an application example. Specifically, we derive theoretical conditions ensuring identifiability and recoverability of causal effects under access to external data and proxy variables. We further introduce a two-step regression estimator (TSR), capable of exploiting proxy variables to adjust for selection bias while accounting for confounding. We show that TSR coincides with prior work if confounding is absent, but achieves a lower variance. Extensive simulation studies validate TSR's correctness for scenarios that include both selection bias and confounding with proxy variables, as exemplified in Figure 2 .

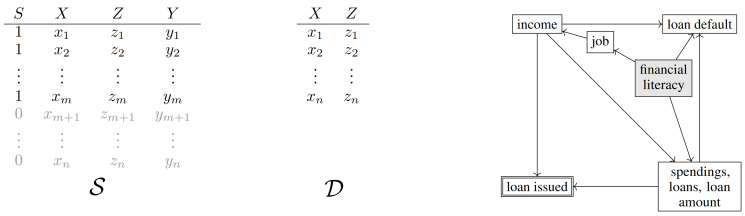


Fig. 1. Left: Dataset composition: $\mathcal{S}$ is affected by a selection mechanism, $\mathcal{D}$ contains unbiased data. Right: Example applicable to our framework. When estimating the causal effect of income X on loan default Y, the node *loan issued* represents the selection variable S, *financial literacy* is an unobserved confounder, and all other nodes serve as proxies (Z).

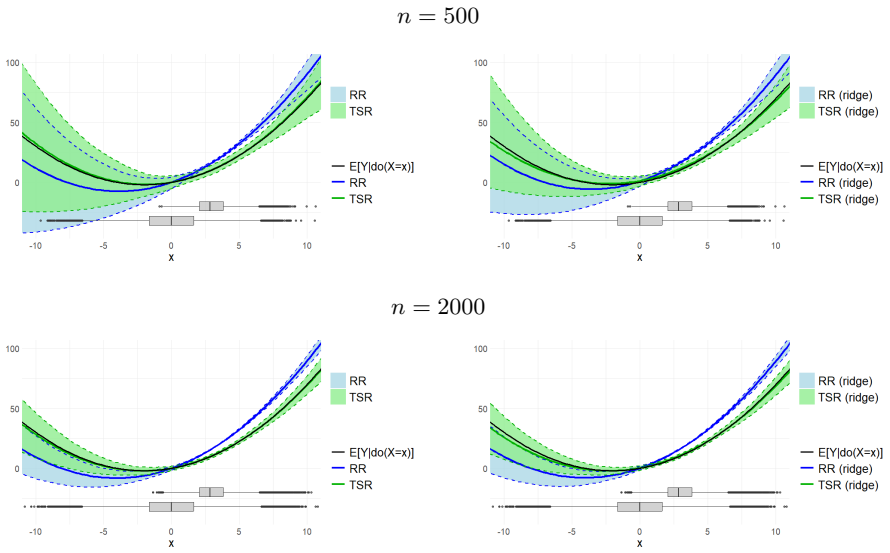


Fig. 2. Comparison of the central 95%-areas of TSR and RR for the DAG in Figure 1 with sample size $n \in \{500, 2000\}$ and $\mathcal{S} \cap \mathcal{D} = \emptyset$. The upper boxplot represents the distribution of X in $\mathcal{S}$ and the lower in $\mathcal{D}$. The curves for RR and TSR display the mean estimation over the 100 simulation runs.

References

1. Philip Boeken, Noud de Kroon, Mathijs de Jong, Joris M. Mooij, and Onno Zoeter: Correcting for selection bias and missing response in regression using privileged information. UAI, 2023.

Time-varying Gaussian Process Bandit Optimization with Experts: no-regret in logarithmically-many side queries

Eliabelle Mauduit[1] ✉, Eloïse Berthier[2], and Andrea Simonetto[1]

[1] Unité des Mathématiques Appliquées, ENSTA, Institut Polytechnique de Paris, 91120 Palaiseau, France {eliabelle.mauduit,andrea.simonetto}@ensta.fr
[2] U2IS, ENSTA, Institut Polytechnique de Paris, 91120 Palaiseau, France eloise.berthier@ensta.fr

Abstract. We consider the problem of time-varying Bayesian optimization with bandit feedback, where the underlying reward function evolves over time and is assumed to lie in a Reproducing Kernel Hilbert Space. While classical Gaussian Process Upper Confidence Bound (GP-UCB) algorithms have been shown to yield sublinear regret in stationary environments, such guarantees generally do not hold in time-varying settings, where the reward function can drift unpredictably. This poses significant challenges for both the theoretical analysis and the design of effective algorithms.

To address this, we introduce SparQ-GP-UCB, a novel framework where we handle time-variations of the reward through uncertainty injection (UI), which consists of increasing the noise variance of past observations. To compensate for the information loss induced by UI, we use a sparsification and querying mechanism, whose key idea is to restore no-regret guarantees by strategically updating a limited number of tale observations.

This formulation naturally leads to a sparse inference problem under heteroscedastic noise, which we solve by extending sparse GP techniques to accommodate time-varying uncertainty profiles.

Theoretically, we show that recovering no-regret is achievable in this setting, provided we are allowed to make a logarithmic number of side queries per time step. This result provides a concrete answer to the open question of how to achieve no-regret in the time-varying setting.

We validate our approach on a real-world dataset of ozone level measurements across New York City, spanning a full year of daily data. In this setting, the reward function (representing the spatial ozone concentration) naturally varies over time, reflecting seasonal and environmental changes. Experimental results show that SparQ-GP-UCB consistently outperforms existing baselines, achieving significantly lower cumulative regret over time.

These findings highlight the practical effectiveness and theoretical soundness of SparQ-GP-UCB in dynamically evolving environments. By requiring only minimal expert input, our method bridges the gap between the idealized assumptions of static optimization and the realities of time-varying systems, making it applicable to a broad range of real-world scenarios where continual adaptation is essential.

1 Figures

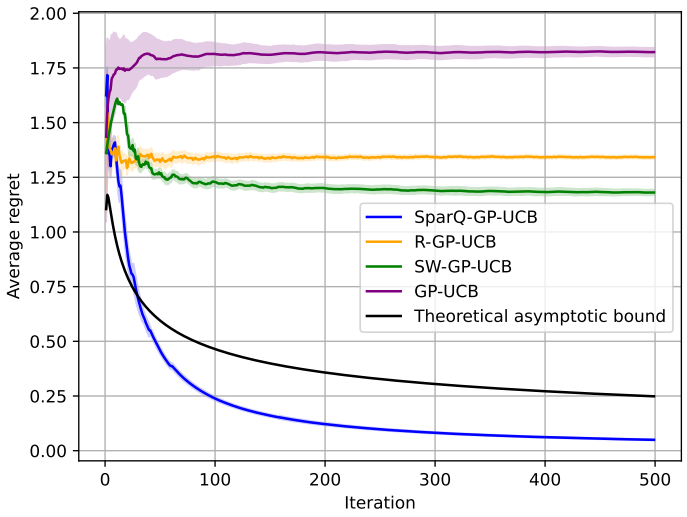


Fig. 1. Average regret of GP-UCB variants on time varying synthetic data.

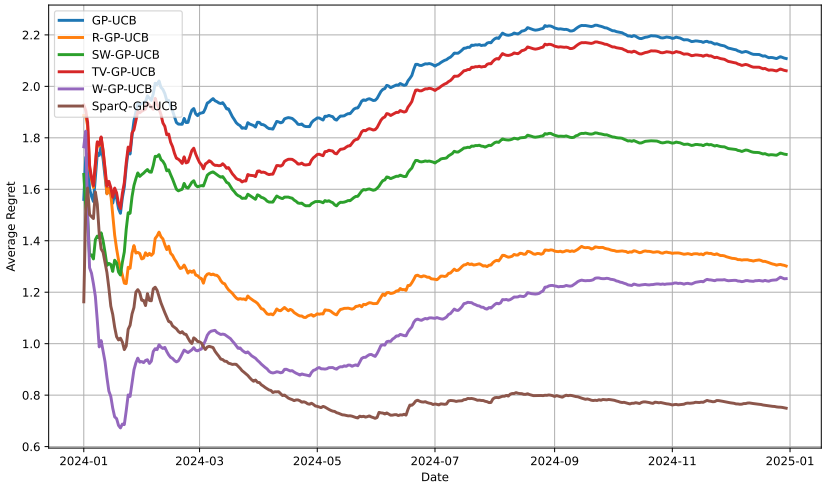


Fig. 2. Numerical performances of GP-UCB variants on real data.

2 Figures

References

1. Besbes, O., Gur, Y., Zeevi, A.: Stochastic multi-armed-bandit problem with non-stationary rewards. In: Advances in Neural Information Processing Systems. vol. 27 (2014)
2. Bogunovic, I., Scarlett, J., Cevher, V.: Time-varying Gaussian Process bandit optimization. In: International Conference on Artificial Intelligence and Statistics. vol. 51, pp. 314–323 (2016)
3. Brunzema, P., Von Rohr, A., Trimpe, S.: On controller tuning with time-varying Bayesian optimization. In: Conference on Decision and Control. pp. 4046–4052 (2022)
4. Burt, D.R., Rasmussen, C.E., van der Wilk, M.: Convergence of sparse variational inference in Gaussian Processes regression. Journal of Machine Learning Research **21**(131), 1–63 (2020)
5. Deng, Y., Zhou, X., Kim, B., Tewari, A., Gupta, A., Shroff, N.B.: Weighted Gaussian Process bandits for non-stationary environments. In: International Conference on Artificial Intelligence and Statistics. vol. 151, pp. 6909–6932 (2022)
6. Derezinski, M., Calandriello, D., Valko, M.: Exact sampling of determinantal point processes with sublinear time preprocessing. In: Advances in Neural Information Processing Systems. vol. 32 (2019)
7. Imamura, H., Charoenphakdee, N., Futami, F., Sato, I., Honda, J., Sugiyama, M.: Time-varying Gaussian Process bandit optimization with non-constant evaluation time. arXiv:2003.04691 (2020)
8. Jadbabaie, A., Rakhlin, A., Shahrampour, S., Sridharan, K.: Online Optimization : Competing with Dynamic Comparators. In: International Conference on Artificial Intelligence and Statistics. vol. 38, pp. 398–406 (2015)
9. Kirschner, J., Krause, A.: Information directed sampling and bandits with heteroscedastic noise. In: Conference On Learning Theory. pp. 358–384 (2018)
10. Koltchinskii, V., Giné, E.: Random matrix approximation of spectra of integral operators. Bernoulli **6**(1), 113–167 (2000)
11. Kulesza, A.: Determinantal point processes for machine learning. Foundations and Trends in Machine Learning **5**(2–3), 123–286 (2012)
12. Makarova, A., Usmanova, I., Bogunovic, I., Krause, A.: Risk-averse heteroscedastic Bayesian optimization. In: Advances in Neural Information Processing Systems. vol. 34, pp. 17235–17245 (2021)
13. Russac, Y., Faury, L., Cappé, O., Garivier, A.: Self-concordant analysis of generalized linear bandits with forgetting. In: International Conference on Artificial Intelligence and Statistics. vol. 130, pp. 658–666 (2021)
14. Scholkopf, B., Tsuda, K., Vert, J.P.: Kernel Methods in Computational Biology. The MIT Press (2004)
15. Shawe-Taylor, J., Williams, C., Cristianini, N., Kandola, J.: On the eigenspectrum of the gram matrix and the generalization error of kernel-PCA. IEEE Transactions on Information Theory **51**(7), 2510–2522 (2005)
16. Snelson, E., Ghahramani, Z.: Sparse Gaussian Processes using pseudo-inputs. In: Advances in Neural Information Processing Systems. vol. 18 (2005)
17. Srinivas, N., Krause, A., Kakade, S.M., Seeger, M.W.: Information-theoretic regret bounds for Gaussian Process optimization in the bandit setting. IEEE Transactions on Information Theory **58**(5), 3250—-3265 (2012)

18. Titsias, M.: Variational learning of inducing variables in sparse Gaussian Processes. In: International Conference on Artificial Intelligence and Statistics. vol. 5, pp. 567–574 (2009)
19. Titsias, M.K.: Variational model selection for sparse Gaussian Process regression. Tech. rep., University of Manchester, UK (2008)
20. Van Vaerenbergh, S., Lázaro-Gredilla, M., Santamaria, I.: Kernel recursive least-squares tracker for time-varying regression. IEEE Transactions on Neural Networks and Learning Systems **23**, 1313–1326 (2012)
21. Werge, N., Akgül, A., Kandemir, M.: BOF-UCB: A Bayesian-optimistic frequentist algorithm for non-stationary contextual bandits. arXiv:2307.03587 (2023)
22. Wu, Q., Iyer, N., Wang, H.: Learning contextual bandits in a non-stationary environment. In: International ACM SIGIR Conference on Research and Development in Information Retrieval (2018)
23. Zhao, P., Zhang, L., Jiang, Y., Zhou, Z.H.: A simple approach for non-stationary linear bandits. In: International Conference on Artificial Intelligence and Statistics. vol. 108, pp. 746–755 (2020)
24. Zhou, X., Shroff, N.: No-regret algorithms for time-varying Bayesian optimization. In: Annual Conference on Information Sciences and Systems. pp. 1–6 (2021)

Skewness-Robust Causal Discovery in Location-Scale Noise Models

Daniel Klippert[1,2] and Alexander Marx[1,2]

[1] Department of Statistics, TU Dortmund University
[2] Research Center Trustworthy Data Science and Security, University Alliance Ruhr
{daniel.klippert,alexander.marx}@tu-dortmund.de

Causal discovery aims to infer cause-and-effect relationships from observational data. In this work, we focus on the bivariate setting, i.e., our goal is to distinguish whether X causes Y, or Y causes X, assuming that there is no unobserved confounder. As this problem cannot be solved in general, assumptions on the underlying structural causal model are required. In our case, we investigate location-scale noise models (LSNMs), a flexible class of models in which the effect Y is modeled based on the cause X as

$$Y = f(X) + g(X)N,$$

where N is an independent noise source. In our experiments, we find that state-of-the-art causal discovery algorithms, which estimate LSNMs assuming standard Gaussian noise (Immer et al., 2023), struggle with identifying the correct causal direction in settings with skewed noise. Skewed noise distributions are common in many real-world domains, including finance, medicine, and environmental sciences. Incorporating skewness allows for greater modeling flexibility and robustness. To address the limitation of existing methods in effectively handling skewed noise, we propose a likelihood-based algorithm for causal discovery under LSNMs with skewed noise. Our method assumes a skew-normal distribution (Azzalini, 1985) for the noise and builds on an expectation maximization algorithm to estimate model parameters via penalized splines. To identify the causal direction based on the estimated models, we employ both independence testing and likelihood scoring. Our benchmarks include synthetically generated datasets and established causal discovery benchmarks. Since Azzalini's skew normal distribution entails the normal distribution as a special case, the method remains applicable in settings with symmetric noise. Preliminary results show that our method achieves near perfect accuracy across various noise distributions ranging from symmetric to strongly skewed.

Table 1. Comparison of the accuracy on the newly proposed benchmarks with skewed LSNMs. Each benchmark considers a different error distribution and consists of 100 datasets, each containing 1000 bivariate data points. The errors were generated from a skew-normal distribution (SN) with skewness parameters of -8 or 20, and from an asymmetric generalized normal (AGN) distribution with skewness parameters of -0.5 or 1. We compare our proposed method to LOCI (Immer et al., 2023) using independence testing (IT) and likelihood scoring (LS). The results confirm that the proposed method improves upon the baseline and achieves near perfect accuracy.

Method\Benchmark	SN_{-8}	SN_{20}	$AGN_{-0.5}$	AGN_{1}
Proposed IT	0.98	0.99	0.92	0.93
Proposed LS	**1.00**	**1.00**	**1.00**	**0.99**
LOCI IT	0.77	0.62	0.79	0.63
LOCI LS	**1.00**	0.90	0.61	0.75

References

Azzalini, A.: A class of distributions which includes the normal ones. Scandinavian journal of statistics, 171–178 (1985)

Immer, A., Schultheiss, C., Vogt, J. E., Schölkopf, B., Bühlmann, P., Marx, A.: On the identifiability and estimation of causal location-scale noise models. In: International Conference on Machine Learning (pp. 14316–14332). PMLR. (2023).

Interpretable Generative Stress Testing

A Framework to Audit Decision Boundaries through Synthetic Data Generation

Inês Gomes[1,2,4 [0009−0006−3104−4622]], Jan N. van Rijn[4 [0000−0003−2898−2168]], and Luís F. Teixeira[1,3 [0000−002−4050−7880]]

[1] Faculty of Engineering, University of Porto
{ines.gomes,luisft}@fe.up.pt
[2] Artificial Intelligence and Computer Science Laboratory, Portugal
[3] INESC TEC, Portugal
[4] Leiden Institute of Advanced Computer Science (LIACS), Netherlands
j.n.van.rijn@liacs.leidenuniv.nl

Abstract. *What happens at the decision boundaries of our classifiers? Are our models truly generalizing or simply memorizing? Which features are responsible for low confidence in predictions?*
With the widespread adoption of deep learning across critical domains, these questions are more important than ever. While deep learning enables significant advances [5], it can challenge interpretability, trust, transparency and responsible use [1]. These concerns increased the demand for auditing tools that expose risks before deployment [7].
In this context, we propose the *Interpretable Generative Stress Testing* framework, which identifies, analyses, and explains low-confidence predictions in image classifiers. The framework is built on three pillars, shown in Fig. 1. The synthetic data generation pillar creates realistic images near the model decision boundary to reveal its vulnerabilities. The pattern finding pillar analyzes these images to uncover common patterns and identify scenarios the model struggles with. The explanation pillar applies interpretability methods to highlight the features contributing to low confidence predictions, combined with textual explanations that translate technical insights into clear, accessible language. These results are integrated into model cards [6] to enhance transparency and support the development of responsible systems.
The proof of concept of this framework [4] used GASTeN [2] for synthetic data generation, which limited its applicability to binary classifiers and small-scale datasets. Fig. 2 illustrates the framework output for MNIST, focusing on the decision boundary between classes *7* and *1*. Ongoing work uses diffusion-based generative models to audit any decision boundary across various image classifiers and datasets [3]. Fig. 3 shows synthetic images generated in the decision boundary between *papillon* and *chihuahua* classes. Future work will improve prototype selection, enhance explanation quality, and integrate textual descriptions.

Keywords: Model Auditing · Generative Artificial Intelligence · Prototype-based Explainable Artificial Intelligence · Stress Testing · Responsible Artificial Intelligence · Computer Vision

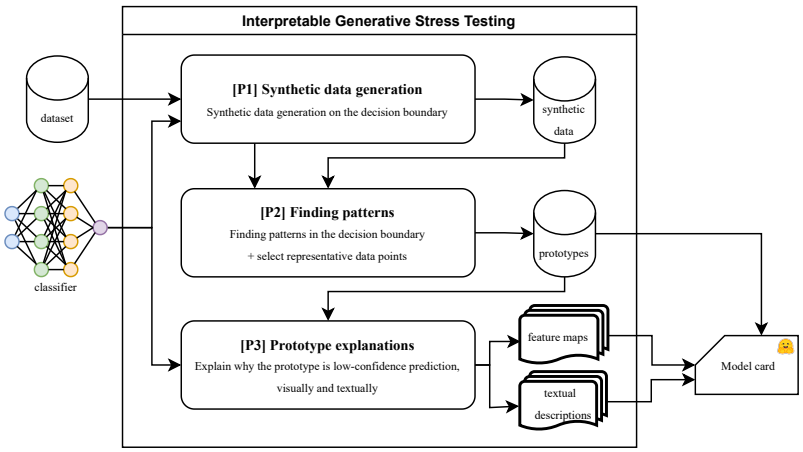


Fig. 1. Given a dataset and an image classifier, the framework identifies and explains low-confidence predictions near decision boundaries. It has three pillars: (1) generating synthetic samples near the boundary, (2) detecting patterns and selecting prototypes, and (3) providing visual and textual explanations for those cases. Insights can fill model cards.

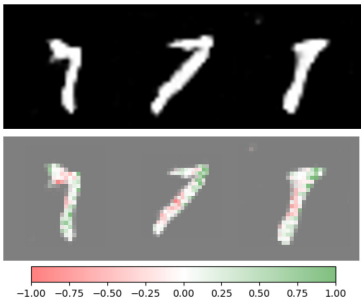


Fig. 2. Prototypes and saliency maps for auditing a binary CNN classifier on MNIST digits 7 and 1. The top row shows three representative synthetic prototypes generated by the framework. The bottom row displays corresponding saliency maps computed with GradientSHAP. Green regions indicate features contributing to classification as a 7, while red regions highlight features associated with classification as a 1.



Fig. 3. Images generated using Stable Diffusion with classifier-guided sampling. A ConvNeXt model trained on ImageNet-1k provides feedback during generation. The process targets the decision boundary between the papillon and chihuahua classes.

Acknowledgements This work was partially funded by projects AISym4Med and CRAI and LIACC. European Union under the Horizon Europe Framework Programme Grant Agreement Nº: 101095387; Agenda "Center for Responsible AI", nr. C645008882-00000055, investment project nr. 62, financed by the Recovery and Resilience Plan (PRR) and by European Union — NextGeneration EU. Funded by the European Union — NextGenerationEU; UID/00027 of the LIACC — Artificial Intelligence and Computer Science Laboratory — funded by Fundação para a Ciência e a Tecnologia, I.P./ MCTES through the national funds. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

References

1. Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F.: Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion **58**, 82–115 (2020)
2. Cunha, L., Soares, C., Restivo, A., Teixeira, L.F.: GASTeN: Generative Adversarial Stress Test Networks. In: Advances in Intelligent Data Analysis XXI. pp. 91–102 (2023)
3. Gomes, I., Restivo, A., Santos, M., Soares, C., van Rijn, J.N., Teixeira, L.F.: Stress Testing Classifiers around the Decision Boundary with Latent Diffusion. In: Submitted to British Machine Vision Conference (2025)
4. Gomes, I., Teixeira, L.F., van Rijn, J.N., Soares, C., Restivo, A., Cunha, L., Santos, M.: Finding Patterns in Ambiguity: Interpretable Stress Testing in the Decision Boundary. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 8316–8321 (2024)
5. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016)
6. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I.D., Gebru, T.: Model Cards for Model Reporting. In: Proceedings of the Conference on Fairness, Accountability, and Transparency. pp. 220–229 (2019)
7. Raji, I.D., Smart, A., White, R.N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., Barnes, P.: Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In: Conference on Fairness, Accountability, and Transparency. pp. 33–44 (2020)

Multimodal Machine Learning Analysis of School Attendance in New Zealand

Tingrui Qiao

University of Auckland

Background. Regular school attendance supports academic learning [2], social development [3], and access to support systems [6,1]. Traditional research relies on surveys completed by parents or teachers, but attendance is a student-centred issue. Hearing directly from students offers more nuanced insights for policy. To support this, the *Our Voices* programme developed the *Our Journey* app (Fig. 1), which collects youth perspectives through open-ended text and image responses. Participants from *Growing Up in New Zealand*, a national longitudinal study, used the app to share experiences across themed areas of life.

Unstructured Data. Responses from the app require interpretation, traditionally done via time-consuming qualitative methods. We introduced Thematic-LM [4], a multi-agent system using large language models to automate thematic analysis. We also developed Thematic Bottleneck Models [5], which integrate codes and themes into prediction tasks, improving interpretability and subjective understanding. We found that students with regular attendance often describe educational goals, supportive families, and positive peers. Absenteeism is linked to negative social experiences and daily struggles, such as poor sleep.

Survey Data. In contrast to personal narratives, survey data from *Growing Up in New Zealand* offers a broader sociological context. Standard methods struggle with variation across waves, lose meaning through one-hot encoding, and omit open-ended responses. Our Survey-as-Text Modelling approach treats multi-wave surveys as coherent text, preserving contextual meaning and enabling comparison. We also propose gradient-guided counterfactual analysis to identify key predictors, highlighting mental health, bullying, family support, and teacher fairness. This work was accepted at the ECML-PKDD 2025 applied data science track. Future work involves a joint analysis of unstructured data and survey data.

References

1. Allison, M.A., Attisha, E., Lerner, M., De Pinto, C.D., Beers, N.S., Gibson, E.J., Gorski, P., Kjolhede, C., O'Leary, S.C., Schumacher, H., et al.: The link between school attendance and good health. Pediatrics **143**(2) (2019)
2. Guleker, R., Keci, J.: The effect of attendance on academic performance. Mediterranean Journal of Social Sciences **5**(23) (2014)
3. Kearney, C.A., Benoit, L., Gonzálvez, C., Keppens, G.: School attendance and school absenteeism: A primer for the past, present, and theory of change for the future. In: Frontiers in Education. vol. 7, p. 1044608. Frontiers (2022)



Fig. 1. *Our Journey* provides a gamified, youth-centred platform where users can share their views while exploring themed islands that cover various aspects of their lives.

4. Qiao, T., Walker, C., Cunningham, C., Koh, Y.S.: Thematic-lm: A llm-based multi-agent system for large-scale thematic analysis. In: Proceedings of the ACM on Web Conference 2025. pp. 649–658 (2025)
5. Qiao, T., Walker, C., Cunningham, C.W., Jang-Jones, A., Morton, S.M.B., Meissel, K., Koh, Y.S.: Understanding school attendance through multimodal modelling of student narratives. In: AI for Children: Healthcare, Psychology, Education
6. Rocque, M., Jennings, W.G., Piquero, A.R., Ozkan, T., Farrington, D.P.: The importance of school attendance: Findings from the cambridge study in delinquent development on the life-course effects of truancy. Crime & Delinquency **63**(5), 592–612 (2017)

Clustering Massive, Noisy, and Unstructured Textual Streams

César Andrade*

Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal
up202101459@edu.fc.up.pt

Abstract. Many digital platforms continuously generate large volumes of short, user-generated texts — often noisy, informal, and unstructured. Clustering these texts is essential for downstream tasks such as recommendation systems, trend detection, and fraud monitoring. However, due to their brevity, lexical variability, and high dimensionality, short texts pose significant challenges for traditional clustering methods. Short Text Stream Clustering (STSC) methods typically follow two paradigms: similarity-based clustering [1–5] and probabilistic topic modeling [6–9]. The former uses embeddings and distance metrics to offer scalability and adaptability; the latter, such as LDA [6] and MStream [10], yields interpretable topics but struggles in the presence of noise, abbreviations, and sparse data, particularly in multilingual and real-time contexts.

This PhD project explores combining both paradigms to improve clustering of short textual data from real-world sources. It focuses on evaluating transformer-based embeddings (e.g., BERT[11], Sentence-BERT [12]), adapting topic models to handle informal input, and assessing hybrid methods that integrate semantic similarity with probabilistic structure. The study is guided by the following questions: Can combining similarity and topic modeling improve clustering in noisy short streams? Do multilingual or domain-specific embeddings enhance robustness? Is it feasible to deploy LLM-based representations in streaming pipelines without compromising scalability?

Recent work by Andrade et al. [13–15] has shown promising results by combining contextual outlier filtering, transformer embeddings, and evaluation metrics tailored for real-world data. These studies suggest that lexical normalization can be an effective strategy to reduce ambiguities caused by abbreviations and spelling variations, improving the reliability and consistency of hybrid clustering methods.

References

1. Zhang, Tian and Ramakrishnan, Raghu and Livny, Miron: BIRCH: an efficient data clustering method for very large databases. In: *ACM sigmod record*, volume 25, number 2, pp. 103–114 (1996)

* Part of the MAP Doctoral Program in Computer Science. Supervised by Prof. João Gama from the Faculty of Economics and Prof. Rita P. Ribeiro from the Faculty of Sciences, University of Porto

2. Carnein, Matthias and Trautmann, Heike: Optimizing data stream representation: An extensive survey on stream clustering algorithms. In: *Business & Information Systems Engineering*, volume 61, number 3, pp. 277–297 (2019)
3. Cao, Feng and Estert, Martin and Qian, Weining and Zhou, Aoying: Density-based clustering over an evolving data stream with noise. In: *Proceedings of the 2006 SIAM international conference on data mining*, pp. 328–339 (2006)
4. Lorbeer, Boris and Kosareva, Ana and Deva, Bersant and Softić, Dženan and Ruppel, Peter and Küpper, Axel. A-BIRCH: automatic threshold estimation for the BIRCH clustering algorithm. In: *INNS conference on Big Data*, pp. 169–178, 2016. Springer.
5. Shou, L., Wang, Z., Chen, K., & Chen, G.: Sumblr: continuous summarization of evolving tweet streams. In: *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pp. 533–542. (2013)
6. Blei, D. M., Ng, A. Y., & Jordan, M. I.: Latent dirichlet allocation. In: *Journal of machine Learning research*, volume 3, number Jan, pp. 993–1022. (2003)
7. Yin, J. and Zhang, W. and Chao, D. and Yu, X. and Liu, Z. and Wang, J.: Model-based clustering of short text streams. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2634-2642. ACM (2016)
8. Yin, J., & Wang, J.: A model-based approach for text clustering with outlier detection. In: *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, pp. 625–636. IEEE (2016)
9. J. Kumar, J. Shao, R. Kumar, S. Ud Din, C. B. Mawuli, and Q. Yang: A context-enhanced Dirichlet model for online clustering in short text streams. In: *Expert Systems with Applications*, vol. 228, p. 120262, 2023.
10. Yin, J., Wang, J., Xu, W., & Gao, M.: Model-based clustering of short text streams. In: *27th ACM Int. Conf. on Information and Knowledge Management*, pages 697–706. ACM (2018)
11. Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *arXiv preprint arXiv:1810.04805*, 2018.
12. Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing* (EMNLP), arXiv:1908.10084, 2019. `https://doi.org/10.48550/arXiv.1908.10084`.
13. Andrade, Cesar and Ribeiro, Rita P. and Gama, João: Topic model with contextual outlier handling: a study on electronic invoice product descriptions. In: *EPIA Conference on Artificial Intelligence*, pp. 365–377. Springer Nature Switzerland, Cham (2023)
14. Andrade, Cesar and Ribeiro, Rita P. and Gama, João: Community-Based Topic Modeling with Contextual Outlier Handling. In: *Conference of the Spanish Association for Artificial Intelligence*, pp. 173–183. Springer Nature Switzerland, Cham (2024)
15. Andrade, Cesar and Ribeiro, Rita P. and Gama, João: Evaluating Short Text Stream Clustering on Large E-commerce Datasets. In: *Brazilian Conference on Intelligent Systems*, pp. 245–259. Springer Nature Switzerland, Cham (2024)

Online Learning from Drifting Capricious Data Streams with Flexible Hoeffding Tree

Ruirui Zhao

National University of Defense Technology, Changsha 410073, China
zhaoruirui20@nudt.edu.cn

Capricious data streams—where features can appear and disappear arbitrarily—are common in practical scenarios such as sensor networks. In existing research, they are mainly handled based on linear classifiers, feature correlation or ensemble of trees [2,5,6,7,8,9]. There exist deficiencies such as limited learning capacity and high time cost. More importantly, the concept drift problem in them receives little attention, which is a fundamental challenge. Therefore, drifting capricious data streams are focused on in this work, and a new algorithm DCFHT (online learning from Drifting Capricious data streams with Flexible Hoeffding Tree) is proposed based on a single Hoeffding tree (Figure 1). DCFHT supports non-linear modeling, efficient adaptation to feature-space changes, and informed concept drift handling.

To conduct online learning from drifting capricious data streams through a single Hoeffding tree, there are two key issues to be addressed. (1) How to update the tree based on each instance with arbitrary disappearance and reappearance of features? (2) When a drift is detected, how to adapt the tree to new concept under capricious feature space? In DCFHT, a universal feature space is dynamically maintained, which is extended by all observable and disappearing features. The first issue is solved through reusing existing knowledge and restructuring the tree in real time. For the second issue, DCFHT starts training a backup learner from a detected drift until it outperforms the primary learner.

To the best of our knowledge, this is the first try to use a single decision tree to learn from capricious data streams and deal with concept drifts within them, addressing a critical challenge in data stream mining. DCFHT offers interpretability, relatively low runtime, and higher learning capacity. Its performance is validated on 22 public and 15 synthetic datasets (Table 1) under diverse scenarios. Results (Table 2, Figure 2, Table 3) show that it is not only more accurate, but also maintains relatively low runtime on capricious data streams.

Based on this work, I plan to improve the Hoeffding tree learning strategy, and I am working on the concept drift detection in varying feature spaces. I am therefore seeking feedback on potential technical innovations in these areas, as well as suggestions for related research directions.

References

1. Beyazit, E., Alagurajah, J., Wu, X.: Online learning from data streams with varying feature spaces. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 3232–3239 (2019)

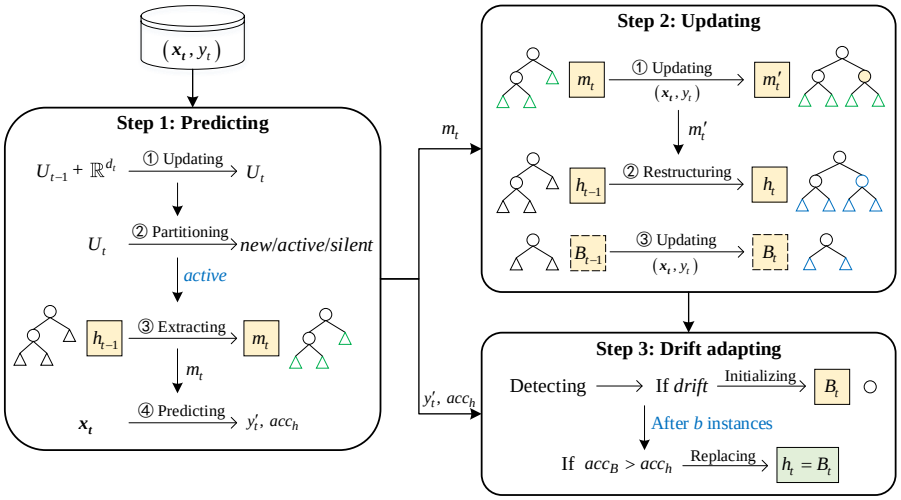


Fig. 1: The overall framework of DCFHT.

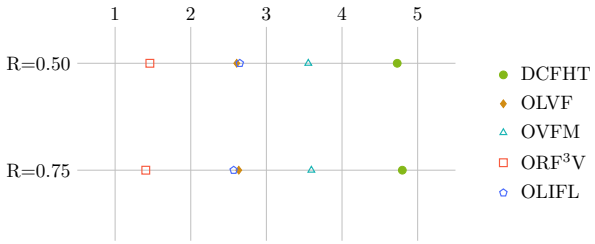


Fig. 2: Critical difference diagram showing the pairwise statistical comparison of the compared algorithms. The further to the right, the better. R means the removing ratio of features.

Table 1: The statistics of datasets used in the experiments

	No.	Dataset	#Instances	#Features	#Classes	Count	Type
Public	1	wpbc	198	33	2	151/47	continuous
	2	ionosphere	351	34	2	225/126	hybrid
	3	wdbc	569	30	2	357/212	continuous
	4	australian	690	14	2	383/307	hybrid
	5	wbc	699	9	2	459/241	discrete
	6	diabetes	768	8	2	500/268	continuous
	7	german	1000	24	2	700/300	hybrid
	8	svmguide3	1284	22	2	947/337	hybrid
	9	splice	3175	60	2	1648/1527	discrete
	10	DNA	3186 (1532)	180	3	767/765/1654	boolean
	11	kr-vs-kp	3196	36	2	1669/1527	discrete
	12	spambase	4601	57	2	2788/1813	continuous
	13	spam	6213	499	2	2069/4144	boolean
	14	isolet	7797 (600)	617	26	300 per class	continuous
	15	mushrooms	8124	112	2	3916/4208	boolean
	16	phishing	11055	46	2	6157/4898	boolean
	17	weather	18159	8	2	5698/12461	continuous
	18	magic04	19020	10	2	12332/6688	continuous
	19	a8a	32561	123	2	7841/24720	boolean
	20	electricity	45312	6	2	19237/26075	continuous
	21	covtype	581012 (20114)	54	7	2747/17367/211840/ 183301/35754/9492/20510	hybrid
	22	poker	829201 (765952)	10	10	415526/350426/39432/ 17541/3225/1657/ 1186/195/11/2	discrete
Synthetic	1	Hyper_{gs}	40000	30	2	19914/20086	continuous
	2	Hyper_{g}	40000	30	2	19913/20087	continuous
	3	Hyper_{i}	40000	30	2	19878/20122	continuous
	4	Hyper_{sg}	40000	30	2	19916/20084	continuous
	5	Hyper_{s}	40000	30	2	19916/20084	continuous
	6	RBF_{gs}	20000	30	2	9014/10986	continuous
	7	RBF_{g}	20000	30	2	9014/10986	continuous
	8	RBF_{r}	20000	30	2	9383/10617	continuous
	9	RBF_{sg}	20000	30	2	9014/10986	continuous
	10	RBF_{s}	20000	30	2	9014/10986	continuous
	11	Agrawal_{gs}	20000	30	2	9998/10002	hybrid
	12	Agrawal_{g}	20000	30	2	9998/10002	hybrid
	13	Agrawal_{r}	20000	30	2	9999/10001	hybrid
	14	Agrawal_{sg}	20000	30	2	9998/10002	hybrid
	15	Agrawal_{s}	20000	30	2	9999/10001	hybrid

The value in "()" is the number of instances when two classes are used. The column "*Count*" lists the distribution of classes, and 3 instances with missing values in *isolet* are excluded. The column "*Type*" refers to the type of features, and "*hybrid*" means there are both continuous and discrete features. For synthetic datasets, subscripts $s/g/i/r$ mean sudden/gradual/incremental/recurrent drifts, respectively; subscript gs means gradual drifts followed by sudden drifts, similarly, subscript sg means sudden drifts followed by gradual drifts.

Table 2: Comparison results on capricious data streams

R			0.50				0.75				
Dataset	DCFHT	OLVF ([1])	OVFM ([4])	ORF^{3}V ([8])	OLIFL ([9])	DCFHT	OLVF	OVFM	ORF^{3}V	OLIFL
wpbc	75.15±0.30	67.63±1.73•	66.92±2.33•	**75.28±1.38**	71.67±1.74•	75.30±0.53	74.85±1.73	67.63±2.78•	**75.33±0.91**	73.38±1.06•
ionosphere	**80.63±1.32**	61.91±1.65•	71.25±1.72•	47.43±3.67•	55.47±11.88•	**75.64±2.35**	63.62±0.61•	67.29±2.00•	48.97±2.59•	52.99±9.26•
wdbc	**90.21±1.21**	49.93±5.65•	89.53±0.63	44.17±3.74•	61.25±0.90•	**87.26±2.29**	57.61±4.46•	85.45±1.00	47.11±3.30•	65.54±1.41•
australian	75.57±1.32	62.48±1.02•	**75.91±1.27**	41.58±1.79•	50.83±5.35•	**70.46±2.10**	55.94±3.08•	69.70±1.22	43.86±1.65•	50.88±5.22•
wbc	**92.20±1.00**	53.23±2.13•	91.67±1.79	30.24±1.36•	63.92±15.65•	**88.78±1.53**	45.98±2.19•	86.92±1.83•	31.29±1.49•	46.35±7.96•
diabetes	**69.10±0.67**	56.47±1.39•	66.71±1.35•	58.51±2.22•	58.71±0.99•	**65.74±1.61**	63.48±0.78•	64.49±1.58	58.24±2.35•	58.61±1.92•
german	69.44±0.22	57.50±0.88•	67.92±1.24•	69.75±0.26∗	**69.85±0.16**∗	69.37±0.39	46.50±0.83•	68.27±0.91•	69.17±0.73	**69.87±0.29**∗
svmguide3	**74.00±1.06**	70.63±0.73•	73.62±0.90	73.42±0.33	73.28±0.32•	**73.56±0.10**	73.40±0.16•	73.04±0.47•	73.36±0.33	72.28±0.48•
splice	**76.88±0.85**	63.67±0.54•	66.65±0.73•	50.39±1.41•	53.01±0.93•	**66.74±1.20**	52.86±0.85•	59.12±0.67•	50.08±1.17•	53.01±0.72•
DNA	**88.48±1.40**	84.69±0.43•	86.01±0.73•	39.67±3.64•	73.94±1.58•	**79.09±1.48**	72.83±0.81•	77.37±1.47•	39.71±1.96•	66.81±0.57•
kr-vs-kp	**72.91±0.49**	62.43±0.54•	70.71±0.54•	43.90±0.68•	51.45±0.56•	**63.57±0.62**	53.13±0.59•	60.80±0.55•	44.59±0.68•	51.96±0.52•
spambase	**74.69±0.59**	59.39±0.56•	51.65±4.71•	60.32±0.18•	52.24±0.49•	**71.08±0.63**	53.78±0.32•	68.19±0.53•	59.11±0.37•	57.34±0.90•
spam	83.15±0.71	90.39±0.18∗	**90.42±0.30**∗	52.77±2.32•	68.75±0.07•	80.42±0.35	76.94±0.67•	**85.45±0.56**∗	49.55±2.44•	68.55±0.11•
isolet	92.55±0.93	**95.82±0.59**∗	91.58±1.61	39.50±4.22•	76.28±6.34•	**87.75±3.35**	80.80±5.59•	83.60±2.98•	40.10±4.18•	65.40±4.19•
mushrooms	90.55±0.86	93.72±0.08∗	**94.65±0.22**∗	44.03±0.90•	82.54±1.01•	85.39±0.58	80.92±0.54•	**87.16±0.24**∗	43.84±1.01•	72.45±0.51•
phishing	**85.03±0.43**	81.92±0.13•	84.09±0.45•	54.30±1.06•	75.05±0.61•	75.83±0.73	67.70±0.82•	**76.07±0.37**	51.29±1.49•	66.96±0.39•
weather	**68.79±0.23**	63.99±0.23•	68.38±0.35•	68.59±0.02•	62.22±0.20•	**68.31±0.12**	68.05±0.03•	66.70±0.31•	68.11±0.15•	64.88±0.26•
magic04	**70.23±0.53**	60.43±0.23•	68.84±0.20•	34.30±0.29•	53.24±2.23•	**68.07±0.50**	60.41±0.28•	66.93±0.17•	36.97±0.78•	52.36±1.86•
a8a	78.22±0.22	77.37±0.14•	**81.28±0.11**∗	75.91±0.01•	76.56±0.05•	76.88±0.18	75.97±0.03•	**78.55±0.08**∗	75.91±0.01•	75.35±0.04•
electricity	**65.70±0.40**	60.10±0.20•	65.55±0.13	45.03±0.55•	57.31±0.16•	**63.87±0.27**	56.31±0.18•	63.19±0.18•	47.91±1.08•	53.60±0.18•
covtype	**88.43±0.13**	86.89±0.14•	88.21±0.21•	86.34±0.00•	86.32±0.01•	86.44±0.12	86.44±0.01	**86.66±0.17**	86.34±0.00•	85.13±0.05•
poker	**61.08±0.02**	50.76±0.04•	54.37±0.04•	45.94±0.04•	52.19±0.05•	**58.98±0.03**	54.25±0.00•	53.40±0.04•	46.15±0.06•	52.01±0.05•
Hyper$_{gs}$	**70.08±0.11**	50.42±0.23•	66.08±0.32•	37.13±1.19•	51.22±0.09•	**63.02±0.12**	49.79±0.03•	59.86±0.39•	40.86±0.49•	50.68±0.21•
Hyper$_{g}$	**70.12±0.14**	50.51±0.20•	66.07±0.19•	37.36±1.11•	51.21±0.14•	**63.41±0.16**	49.79±0.02•	60.19±0.25•	40.46±0.46•	50.65±0.18•
Hyper$_{i}$	**68.96±0.15**	58.27±0.27•	61.54±0.30•	37.90±0.49•	62.48±0.15•	**63.06±0.21**	54.06±0.63•	56.58±0.42•	40.98±0.38•	56.15±0.09•
Hyper$_{sg}$	**69.93±0.15**	50.43±0.25•	65.91±0.31•	37.28±1.13•	51.21±0.11•	**63.09±0.13**	49.79±0.02•	60.23±0.34•	40.70±0.56•	50.60±0.14•
Hyper$_{s}$	**70.02±0.13**	50.52±0.24•	66.10±0.32•	37.16±1.20•	51.32±0.12•	**63.08±0.12**	49.79±0.03•	60.12±0.31•	40.84±0.63•	50.63±0.14•
RBF$_{gs}$	**64.91±0.24**	55.25±0.34•	56.64±0.33•	51.14±0.99•	57.79±0.22•	**59.67±0.31**	54.92±0.03•	53.59±0.30•	51.31±1.00•	55.45±0.10•
RBF$_{g}$	**64.90±0.18**	55.28±0.29•	56.70±0.30•	51.28±1.01•	57.84±0.24•	**59.86±0.23**	54.92±0.03•	53.57±0.29•	51.19±1.03•	55.44±0.10•
RBF$_{r}$	**64.21±0.31**	54.80±0.30•	56.45±0.30•	50.45±0.87•	57.18±0.26•	**59.92±0.25**	53.13±0.07•	53.30±0.39•	50.45±0.69•	54.83±0.14•
RBF$_{sg}$	**64.76±0.20**	55.25±0.22•	56.66±0.25•	51.23±0.95•	57.79±0.21•	**59.82±0.22**	54.92±0.02•	53.56±0.21•	51.15±1.01•	55.40±0.10•
RBF$_{s}$	**64.79±0.20**	55.09±0.17•	56.45±0.18•	51.66±0.91•	57.78±0.19•	**60.08±0.20**	54.92±0.03•	53.85±0.22•	50.76±0.85•	55.54±0.04•
Agrawal$_{gs}$	**55.50±0.36**	50.38±0.29•	50.56±0.85•	47.10±0.33•	50.26±0.26•	**54.16±0.39**	50.03±0.23•	50.97±0.30•	47.33±0.21•	50.17±0.23•
Agrawal$_{g}$	**55.52±0.38**	50.36±0.28•	50.20±0.52•	47.17±0.32•	50.25±0.23•	**54.24±0.36**	50.08±0.26•	50.96±0.30•	47.42±0.19•	50.22±0.25•
Agrawal$_{r}$	**57.75±0.34**	50.14±0.36•	49.94±0.48•	47.45±0.41•	49.95±0.30•	**55.06±0.36**	50.11±0.22•	50.05±0.29•	47.94±0.28•	49.87±0.31•
Agrawal$_{sg}$	**55.36±0.42**	50.42±0.32•	50.42±0.39•	47.24±0.32•	50.20±0.28•	**54.18±0.42**	50.05±0.23•	50.92±0.37•	47.44±0.32•	50.23±0.29•
Agrawal$_{s}$	**55.86±0.38**	50.06±0.33•	50.39±0.46•	47.39±0.28•	49.99±0.42•	**53.74±0.34**	50.09±0.19•	50.68±0.30•	47.76±0.31•	50.22±0.31•
Average	**72.31±10.89**	62.12±13.66	68.27±13.75	50.28±12.95	60.34±10.40	**68.24±10.44**	59.41±11.17	65.25±12.06	50.91±12.15	58.43±9.39
w/t/l	——	34/0/3	28/6/3	34/2/1	36/0/1	——	35/2/0	29/5/3	34/3/0	36/0/1

'•' denotes that DCFHT is significantly better than the compared method by paired t-test [3] with confidence level 0.05, and '∗' denotes that DCFHT is significantly worse. The win/tie/loss counts for DCFHT are summarized in the last row, abbreviated as w/t/l.

Table 3: Comparison of runtime on capricious data streams (in seconds, $R = 0.50$)

Dataset	DCFHT	OLVF	OVFM	ORF^{3}V	OLIFL	Dataset	DCFHT	OLVF	OVFM	ORF^{3}V	OLIFL
wpbc	0.105	**0.015**	1.409	1.468	0.466	electricity	12.868	**3.630**	103.984	295.578	24.320
ionosphere	0.443	**0.072**	23.053	0.632	0.868	covtype	12.472	**1.716**	412.947	141.075	70.171
wdbc	0.641	**0.072**	12.634	1.484	1.217	poker	243.095	**68.762**	4278.562	415.336	336.658
australian	0.792	**0.072**	13.984	0.820	0.754	Hyper$_{gs}$	19.858	**17.042**	1319.004	2428.388	21.262
wbc	0.487	**0.072**	6.770	1.158	0.500	Hyper$_{g}$	22.396	**11.413**	1157.273	880.738	21.196
diabetes	0.460	**0.075**	6.441	1.766	0.563	Hyper$_{i}$	22.916	**15.737**	706.280	855.118	21.247
german	0.788	**0.152**	28.152	1.728	1.580	Hyper$_{sg}$	19.163	**14.602**	373.134	907.687	21.223
svmguide3	0.505	**0.091**	7.657	11.284	1.851	Hyper$_{s}$	23.527	**11.675**	283.029	2170.050	21.232
splice	2.997	**0.272**	126.118	12.720	12.113	RBF$_{gs}$	11.515	**5.550**	283.658	543.334	10.641
DNA	3.243	**0.203**	228.150	23.498	16.176	RBF$_{g}$	8.223	**6.760**	217.067	511.600	10.609
kr-vs-kp	1.783	**0.269**	89.105	5.188	7.889	RBF$_{r}$	9.240	**3.480**	148.783	445.393	10.654
spambase	5.312	**0.424**	63.912	25.338	16.796	RBF$_{sg}$	8.800	**3.436**	149.186	363.006	10.628
spam	85.930	**11.398**	5489.517	331.493	175.240	RBF$_{s}$	10.420	**3.606**	176.064	457.880	10.650
isolet	4.580	**0.191**	79.623	43.756	19.514	Agrawal$_{gs}$	6.840	6.306	72.573	47.263	**4.096**
mushrooms	12.682	**0.888**	1668.878	39.589	16.593	Agrawal$_{g}$	6.174	**2.995**	72.704	47.190	4.091
phishing	15.032	**5.024**	319.655	37.548	30.448	Agrawal$_{r}$	7.120	**2.171**	72.844	48.049	4.111
weather	6.685	**6.601**	45.554	151.318	13.661	Agrawal$_{sg}$	6.991	**2.165**	72.825	47.140	4.137
magic04	6.535	**5.297**	54.952	1654.833	13.089	Agrawal$_{s}$	6.130	**3.661**	72.924	47.282	4.080
a8a	102.309	**3.667**	1952.353	238.346	178.093	***Average***	19.164	**5.934**	545.696	357.732	30.227

2. Chen, Z., He, Y., Wu, D., Zhang, W., Deng, Z.: $\ell_{1,\infty}$ -mixed norm promoted row sparsity for fast online CUR decomposition learning in varying feature spaces. In: Proceedings of the 2025 SIAM International Conference on Data Mining (SDM). pp. 124–133. SIAM (2025)
3. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. The Journal of Machine learning research **7**, 1–30 (2006)
4. He, Y., Dong, J., Hou, B.J., Wang, Y., Wang, F.: Online learning in variable feature spaces with mixed data. In: 2021 IEEE International Conference on Data Mining (ICDM). pp. 181–190. IEEE (2021)
5. He, Y., Schreckenberger, C., Stuckenschmidt, H., Wu, X.: Towards utilitarian online learning-a review of online algorithms in open feature space. In: Proc. of International Joint Conferences on Artificial Intelligence Organization. pp. 6647–6655 (2023)
6. He, Y., Wu, B., Wu, D., Beyazit, E., Chen, S., Wu, X.: Online learning from capricious data streams: A generative approach. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19. pp. 2491–2497 (7 2019)
7. He, Y., Yuan, X., Chen, S., Wu, X.: Online learning in variable feature spaces under incomplete supervision. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 4106–4114 (2021)
8. Schreckenberger, C., He, Y., Lüdtke, S., Bartelt, C., Stuckenschmidt, H.: Online random feature forests for learning in varying feature spaces. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 4587–4595 (2023)
9. You, D., Yan, H., Xiao, J., Chen, Z., Wu, D., Shen, L., Wu, X.: Online learning for data streams with incomplete features and labels. IEEE Transactions on Knowledge and Data Engineering **36**(9), 4820–4834 (2024)

Identification of malicious devices in Federated Machine Learning Settings

Yohannes Biadgligne[1][0000−0002−5808−1849], Yassine Baghoussi[2][0000−0002−1943−1471], Kai Li[3][0000−0002−0517−2392], and Alípio Jorge[4][0000−0002−5475−1382]

[1] INESC TEC, University of Porto, Porto, Portugal
yohannes.biadgligne@inesctec.pt
[2] INESC TEC, University of Porto, Porto, Portugal
yassine.baghoussi@inesctec.pt
[3] CISTER Research Centre
kai@isep.ipp.pt
[4] INESC TEC, University of Porto, Porto, Portugal
alipio.jorge@inesctec.pt

Abstract. Federated Learning (FL) enables collaborative model training across multiple edge devices while preserving data privacy [1]. However, its decentralized nature and reliance on client-generated data and updates make it vulnerable to security threats. One critical threat is poisoning attacks, where malicious clients manipulate local data or model updates before sending them to the central server. Once aggregated, these poisoned contributions can severely degrade the accuracy, reliability, and safety of the global model. With FL's growing adoption in privacy-sensitive domains like healthcare, finance, and smart infrastructure, along with compliance demands under regulations such as the General Data Protection Regulation (GDPR), ensuring the security and robustness of FL systems is essential.

Detecting and mitigating malicious devices is key to securing FL deployments. Approaches such as CONTRA [2], D2MIF [3], and FLDetector [4] offer valuable insights but often rely on narrow assumptions such as IID data, fixed attack models (e.g., label flipping or backdoor), or static attacker ratios. These constraints limit their effectiveness in real-world FL settings marked by non-IID data, dynamic participation, and evolving threats, highlighting the need for generalized, adaptive, and attack-agnostic identification strategies.

This PhD research aims to advance detection and defense methods by integrating attack-resilient optimization, adaptive anomaly detection, and cross-client behavioral analysis. Objectives include (1) identifying adversarial updates without access to raw client data or labels, (2) ensuring robustness under non-IID and resource-constrained settings, and (3) maintaining model performance for benign clients while minimizing false positives. As a foundation, we evaluated three widely used client-side optimizers—Adam, SGD, and RMSProp—under label-flipping attacks (LFAs). Results from 900 runs show SGD outperforms the others in 87% of cases, highlighting its resilience to LFAs [5]. Sample results from our experiment are summarized in Figure 1.

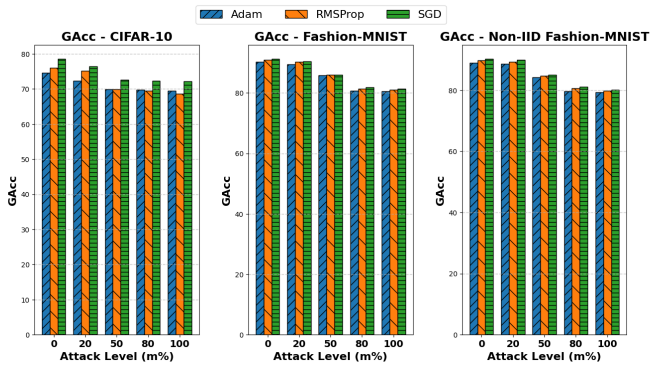


Fig. 1. Global accuracy of SGD, Adam, and RMSProp across different LFA and FL settings.

Keywords: Machine learning · Federated learning · Adversarial Attacks · Anomaly Detection · Poisoning Attacks .

References

1. McMahan, H.B., Moore, E., Ramage, D., Hampson, S., Agüera y Arcas, B.: Communication-Efficient Learning of Deep Networks from Decentralized Data. arXiv preprint arXiv:1602.05629 (2023)
2. Awan, S., Luo, B., Li, F.: CONTRA: Defending against poisoning attacks in federated learning. In: Computer Security – ESORICS 2021, 26(Part I), 455–475 (2021). https://doi.org/10.1007/978-3-030-88418-5_22
3. Liu, W., Lin, H., Wang, X., Hu, J., Kaddoum, G., Piran, M.J., Alamri, A.: D2MIF: A malicious model detection mechanism for federated learning empowered artificial intelligence of things. IEEE Internet of Things Journal 8(18), 13923–13936 (2021). https://doi.org/10.1109/JIOT.2021.3081606
4. Zhang, Z., Cao, X., Jia, J., Gong, N.Z.: FLDetector: Defending Federated Learning Against Model Poisoning Attacks via Detecting Malicious Clients. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 2545–2555. Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3534678.3539231
5. Biadgligne, Y., Baghoussi, Y., Kai, H., Jorge, A.: Resilience Under Attack: Benchmarking Optimizers Against Poisoning in Federated Learning for Image Classification Using CNN. In: Proceedings of the IWANN 2025, Coruña, Spain, 16-18 June (2025).

Learning Topology Actions for Power Grid Control

Mohamed Hassouna[1,2][0000−0001−9927−1625]

[1] Fraunhofer Institute for Energy Economics and Energy System Technology (IEE)
[2] Intelligent Embedded Systems, University of Kassel
Germany
mohamed.hassouna@iee.fraunhofer.de
mohamed.hassouna@uni-kassel.de

1 Abstract

The rising proportion of renewable energy in the electricity mix introduces significant operational challenges for power grid operators. Effective power grid management demands adaptive decision-making strategies capable of handling dynamic conditions. With the increase in complexity, more and more Deep Learning (DL) approaches have been proposed to find suitable grid topologies for congestion management. In this work, we contribute to this research by introducing a novel Imitation Learning (IL) approach that leverages soft labels derived from simulated topological action outcomes, thereby capturing multiple viable actions per state. Unlike traditional IL methods that rely on hard labels to enforce a single optimal action, our method constructs soft labels over actions, by leveraging effective actions that prove suitable in resolving grid congestion. To further enhance decision-making, we integrate Graph Neural Networks (GNNs) to encode the structural properties of power grids, ensuring that the topology-aware representations contribute to better agent performance. Our approach significantly outperforms state-of-the-art baselines, all of which use only topological actions, as well as feedforward and GNN-based architectures with hard labels. Most notably, it achieves a 17% better performance compared to the greedy expert agent from which the imitation targets were derived.

Keywords: Power Grids · Topology Control · Graph Neural Networks · Learning to Run a Power Network

Adaptive Sampling Strategies for Learning from Imbalanced Regression Data Streams

Ehsan Aminian[1,3], Rita P. Ribeiro[1,3], and Joao Gama[2,3]

[1] Faculty of Sciences, University of Porto, Porto, Portugal
[2] Faculty of Economics, University of Porto, Porto, Portugal
[3] LIAAD, INESC TEC, Porto, Portugal
ehaminian@gmail.com, rpribeiro@fc.up.pt, jgama@fep.up.pt

Abstract. A critical, yet underexplored challenge in machine learning is handling imbalanced regression data streams, where accurately predicting rare target values is more significant than predicting frequent ones [1]. Unlike imbalanced classification, regression involves a continuous target variable, making the identification and learning of underrepresented values complex, especially under the typical data stream constraints: single-pass processing, limited memory, and adaptation to concept drift[2]. This Ph.D research addresses this gap by developing novel adaptive sampling strategies tailored for imbalanced regression in data streams.
The primary contribution of the thesis is twofold. First, we propose Chebyshev-based undersampling (ChebyUS) and oversampling (ChebyOS) strategies. These methods leverage Chebyshev's inequality to differentiate between frequent and rare cases based on the distance of target values from their incrementally updated mean. This approach enables more balanced learning by either reducing the influence of frequent cases or amplifying rare ones, proving particularly effective when rare events are concentrated at the extremes of the target distribution[1]. Second, recognizing the limitation that rare cases may not always reside at distribution extremes, we introduce histogram-based undersampling (HistUS) and oversampling (HistOS) strategies[2]. These methods dynamically adapt to the evolving nature of data streams by constructing and maintaining an online histogram of target values[3]. This online histogram guides the sampling process, ensuring that rare instances occurring anywhere within the target distribution are adequately represented, offering greater flexibility and generalizability. The proposed sampling strategies were integrated into four online regression algorithms—Perceptron, FIMT-DD, AMRules, and TargetMean [1,2]—and evaluated on synthetic and real-world datasets. Results show consistent improvements over baselines in predicting rare values, as measured by RMSEϕ and SERA (Table 1).
A comparative analysis revealed that histogram-based methods offered broader applicability, whereas Chebyshev-based methods demonstrated superior efficiency in capturing rare extreme values (Table.2). This research advances the field by offering practical, scalable solutions for real-time applications where accurate prediction of rare events is crucial. Future work includes enhancing robustness to abrupt concept drifts.

Keywords: Imbalanced Data Streams Regression · Sampling Strategies

Table 1. Comparison of ChebyUS/ChebyOS and HistUS/HistOS with Baseline using $RMSE_\phi$ and $SERA$: average rank and number of significant wins (#w) at 95% confidence level.

Algorithm	**Cheby-based**	$RMSE_\phi$	#W	**Hist-based**	$RMSE_\phi$	#W	$SERA$	#W
AMRules	Baseline	6.32	(5)	Baseline	3.83	(0)	6.25	(0)
	ChebyUS	**5.61**	(8)	HistUS	2.42	(8)	**3.58**	(10)
	ChebyOS	**5.43**	(10)	HistOS	2.58	(8)	**3.92**	(6)
Perceptron	Baseline	4.68	(3)	Baseline	5.58	(1)	5.50	(1)
	ChebyUS	**2.25**	(9)	HistUS	3.83	(9)	**2.75**	(10)
	ChebyOS	**3.00**	(10)	HistOS	3.75	(9)	**2.75**	(9)
FIMT-DD	Baseline	3.68	(4)	Baseline	3.92	(1)	5.33	(0)
	ChebyUS	**2.68**	(10)	HistUS	2.96	(7)	**3.00**	(12)
	ChebyOS	**2.07**	(11)	HistOS	**2.33**	(10)	**3.25**	(8)
TargetMean	Baseline	5.82	(4)	Baseline	7.46	(0)	6.62	(1)
	ChebyUS	**4.96**	(8)	HistUS	6.00	(10)	**2.96**	(9)
	ChebyOS	**4.43**	(10)	HistOS	5.88	(10)	**3.12**	(9)

Table 2. Comparison of undersampling (ChebyUS vs HistUS) and oversampling (ChebyOS vs HistOS) approaches using $RMSE_\phi$ and $SERA$: average rank and number of significant wins (#w) at 95% confidence level.

Algorithm	**Approach**	**Undersampling**				**Oversampling**			
		$RMSE_\phi$	#w	$SERA$	#w	$RMSE_\phi$	#w	$SERA$	#w
AMRules	Cheby	**2.83**	(5)	**4.92**	(5)	3.50	(0)	6.00	(1)
	Hist	3.08	(2)	5.42	(5)	**3.00**	(7)	**4.50**	(6)
Perceptron	Cheby	**4.21**	(6)	4.42	(3)	5.42	(0)	5.08	(1)
	Hist	4.46	(5)	4.42	(5)	**4.33**	(9)	**3.08**	(8)
FIMT-DD	Cheby	4.71	(2)	4.83	(3)	**2.92**	(3)	5.08	(3)
	Hist	**3.79**	(6)	**4.50**	(8)	3.50	(1)	**3.58**	(6)
TargetMean	Cheby	**5.92**	(7)	**3.29**	(7)	7.25	(1)	5.33	(1)
	Hist	7.00	(1)	4.21	(2)	**6.08**	(9)	**3.33**	(8)

References

1. E. Aminian, R. P. Ribeiro, and J. Gama. Chebyshev approaches for imbalanced data streams regression models. *Data Mining and Knowledge Discovery*, 35(6):2389–2466, 2021. Springer.
2. Ehsan Aminian, Rita P. Ribeiro, and Joao Gama. Histogram Approaches for Imbalanced Data Streams Regression. *arXiv preprint arXiv:2501.17568*, 2025.
3. J. Gama and C. Pinto. Discretization from data streams: applications to histograms and data mining. In *Proceedings of the 2006 ACM symposium on Applied computing*, pages 662–667, 2006.

Why t-SNE Works: Human-Aligned Evaluation and a Parametric Alternative

Shirin Mohebi

Supervisors: Jefrey Lijffijt, Tijl De Bie

Ghent University, Belgium

Abstract. Motivation: t-SNE is widely used for dimensionality reduction (DR) in bioinformatics, vision, and NLP. Despite its limitations, it still outperforms newer parametric methods in practice. This paradox raises:
Why does t-SNE still dominate, despite poor scalability, non-parametric design, and weak global structure?
Are parametric alternatives, deep nets, kernel embeddings, parametric t-SNE, inferior, or is there a gap in adoption, tooling, or metrics?
What properties of t-SNE make it feel right to users, even when other methods may outperform it in benchmarks?

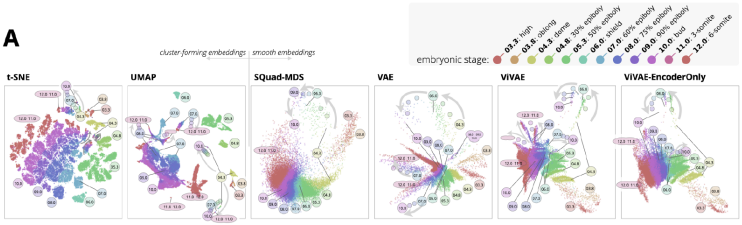


Fig. 1: From Novak *et al.* [1], Fig. 3A (CC-BY-NC-ND).

Methodology: To answer these questions, We will survey DR, comparing recent parametric approaches (e.g., Parametric t-SNE, autoencoders, UMAP-based models) and non-parametric ones (e.g., t-SNE, UMAP, LargeVis). The survey spans multiple real and synthetic datasets and will incorporate both established metrics (trustworthiness, continuity, k-NN preservation, global structure) and qualitative factors like interpretability and tuning effort.

Contributions: We hypothesize that current metrics may not fully capture why t-SNE is preferred in practice. We will test whether these metrics align with user judgement and visual utility. Based on the gaps we identify, we propose to design new, human- or task-aligned metrics. These insights will guide a new parametric DR method that combines the strengths of t-SNE with the benefits of parametric approaches.

Keywords: Dimensionality reduction · t-SNE · Evaluation metrics · Parametric methods · Visualization

References

1. Novak, D. et al.: *Interpretable models for scRNA-seq data embedding with multi-scale structure preservation.* bioRxiv (2024). https://doi.org/10.1101/2023.11.23.568428

Leveraging Large Language Models and Semantic Enhancement for Vision-Language Grounding

Tahsir Ahmed Munna[1,2], Nuno Guimarães[1,2], Alípio Jorge[1,2]

[1] FCUP - University of Porto
[2] LIAAD - INESC TEC
tahsir.a.munna@inesctec.pt,nuno.r.guimaraes@inesctec.pt,amjorge@fc.up.pt

Abstract. Recent advances in natural language processing and computer vision, particularly through Large Language Models (LLMs), have opened new frontiers in multimodal learning [8, 13, 4]. For example, in the medical domain, multimodal learning can contribute towards medical report generation [8], clinical decision support [12], and visual question answering [6]. However, aligning complex medical narratives with their corresponding visual representations remains a critical bottleneck in clinical decision-making, where precision and interpretability are top priorities [15]. This challenge is further amplified in low-resource language varieties, such as European Portuguese, where the scarcity of annotated datasets hinders the development of effective Vision-Language Grounding (VLG) systems [9].

This work addresses these limitations by introducing a two-fold framework focused on Extractive VLG (EVLG) and Abstractive VLG (AVLG). EVLG enables precise localisation of medical findings in images based on textual descriptions and vice versa, while AVLG supports bidirectional translation between visual data and medical narratives. To overcome data scarcity, the approach leverages synthetic clinical narratives generated via LLMs and employs cross-lingual transfer learning [10]. Semantic Enhancement Techniques (SETs), including medical ontologies [14] and knowledge graphs [3], are integrated to improve both linguistic fidelity and grounding accuracy.

The methodology combines multimodal pre-training, joint embedding spaces, and attention mechanisms to construct a robust system capable of generalizing across imaging modalities and linguistic variations. The initial evaluation will be conducted using existing state-of-the-art methods [11, 7] on benchmark datasets [2, 1]. Human expert evaluation will also be incorporated to assess the interpretability and reliability of our approach compared to baseline methods. Beyond healthcare, the framework also holds potential in domains requiring precise text-image alignment, such as autonomous navigation [5].

Acknowledgments

This work is co-financed by Component 5 - Capitalization and Business Innovation, integrated in the Resilience Dimension of the Recovery and Resilience Plan within the scope of the Recovery and Resilience Mechanism (MRR) of the European Union (EU), framed in the Next Generation EU, for the period 2021 - 2026, within project HfPT, with reference 41.

References

1. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., et al.: Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33, pp. 590–597 (2019)
2. Johnson, A.E., Pollard, T.J., Berkowitz, S.J., Greenbaum, N.R., Lungren, M.P., Deng, C.y., Mark, R.G., Horng, S.: Mimic-cxr, a de-identified publicly available database of chest radiographs with free-text reports. Scientific data **6**(1), 317 (2019)
3. Lecu, A., Groza, A., Hawizy, L.: Knowledge graph-driven retrieval-augmented generation: Integrating deepseek-r1 with weaviate for advanced chatbot applications. arXiv preprint arXiv:2502.11108 (2025)
4. Li, C., Wong, C., Zhang, S., Usuyama, N., Liu, H., Yang, J., Naumann, T., Poon, H., Gao, J.: Llava-med: Training a large language-and-vision assistant for biomedicine in one day. Advances in Neural Information Processing Systems **36**, 28541–28564 (2023)
5. Li, Y., Yang, L., Yang, M., Yan, F., Liu, T., Guo, C., Chen, R.: Navblip: a visual-language model for enhancing unmanned aerial vehicles navigation and object detection. Frontiers in Neurorobotics **18**, 1513354 (2025)
6. Lim, Q.Z., Lee, C.P., Lim, K.M., Anbananthen, K.S.M.: Vlmt: Vision-language multimodal transformer for multimodal multi-hop question answering. arXiv preprint arXiv:2504.08269 (2025)
7. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015)
8. Luo, L., Tang, B., Chen, X., Han, R., Chen, T.: Vividmed: Vision language model with versatile visual grounding for medicine. arXiv preprint arXiv:2410.12694 (2024)
9. Lupascu, M., Rogoz, A.C., Stupariu, M.S., Ionescu, R.T.: Large multimodal models for low-resource languages: A survey. arXiv preprint arXiv:2502.05568 (2025)
10. Munna, T.A., Fernandes, A.L., Silvano, P., Guimarães, N., Jorge, A.: Using llms to generate patient journeys in portuguese: an experiment. In: Proceedings of the Eighth International Workshop on Narrative Extraction from Texts held in conjunction with the 47th European Conference on Information Retrieval. pp. 161–173. CEUR-WS (2025), `https://ceur-ws.org/Vol-3964/paper14.pdf`
11. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: Bleu: a method for automatic evaluation of machine translation. In: Proceedings of the 40th annual meeting of the Association for Computational Linguistics. pp. 311–318 (2002)

12. Patil, A., Patil, V., Sankpal, S., Patankar, T.S., Bhute, H.: Multimodal decision support system for improved diagnosis and healthcare decision making. Journal of Biology and Health Science (2025)
13. Wang, W., Lv, Q., Yu, W., Hong, W., Qi, J., Wang, Y., Ji, J., Yang, Z., Zhao, L., XiXuan, S., et al.: Cogvlm: Visual expert for pretrained language models. Advances in Neural Information Processing Systems **37**, 121475–121499 (2024)
14. Wang, Y., Yin, C., Zhang, P.: Multimodal risk prediction with physiological signals, medical images and clinical notes. Heliyon **10**(5) (2024)
15. Zhu, K., Xia, P., Li, Y., Zhu, H., Wang, S., Yao, H.: Mmedpo: Aligning medical vision-language models with clinical-aware multimodal preference optimization. In: Proceedings of the 42nd International Conference on Machine Learning (ICML) (2025), `https://icml.cc/virtual/2025/poster/44599`, poster presented at ICML 2025

Transfer Learning for Evolving Domains

Ricardo Ribeiro Pereira[1,2] (✉), Jacopo Bono[1],
Pedro Ribeiro[2], and Carlos Soares[2]

[1] Feedzai, Portugal
[2] University of Porto, Portugal
ricardo.ribeiro@feedzai.com

Extended Abstract

Transfer learning (TL) explores how to leverage knowledge from various tasks or domains (sources) to enhance the predictive performance of machine learning (ML) models in related tasks or domains (targets) [1]. Most TL methods assume fixed data availability scenarios, such as fully labelled source domains and unlabelled or sparsely labelled target domains. However, real-world ML applications often involve evolving data conditions, where data from new domains arrives gradually and labels become available over time.

The goal of my PhD is to study this problem, which we call Transfer Learning for Evolving Domains (TrED). TrED frames TL as a continuum of stages that align with existing TL sub-problems. It begins with multiple labelled source domains and no data from a new target domain, similar to Domain Generalisation [2, 6]. As target data becomes available, Domain Adaptation [3, 4] techniques may help improve performance. Eventually, the target domain can mature into a source for newer domains or be included in a Multi-Domain Learning [5] solution.

So far, we have conducted a review of the literature on TL and its aforementioned sub-problems. We identified common techniques used under different data availability assumptions and hypothesised which approaches may generalise across the stages of TrED. To support our experimental work, we developed an evaluation framework that simulates evolving data conditions, including domain shift, label delay, and time-based sampling.

Moving forward, we plan to explore feature representation techniques that support effective knowledge transfer across domains. Then, we will propose techniques to address specific challenges of each TrED stage, such as performance estimation, data scarcity, model updates, and negative transfer. Ultimately, our goal is to develop original TL solutions capable of operating robustly across the full spectrum of data availability conditions.

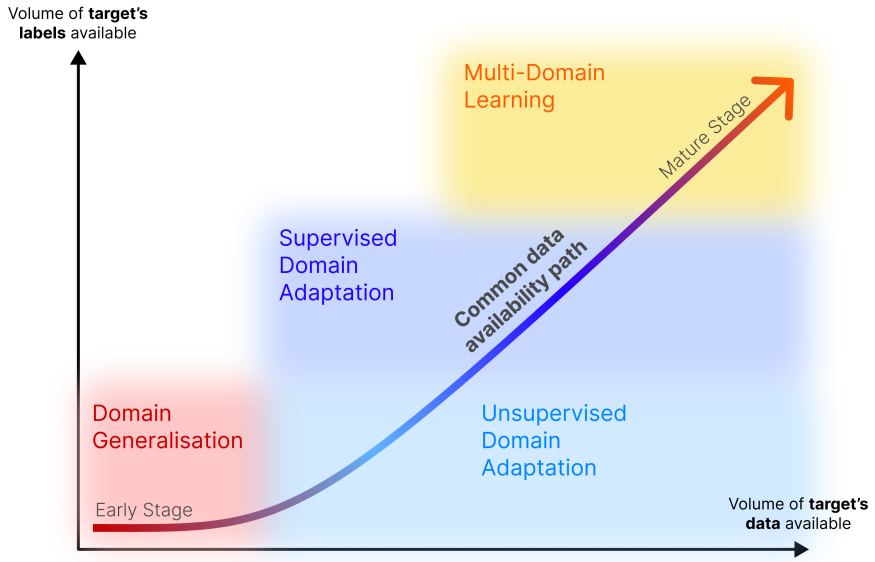


Fig. 1. Transfer learning can be divided in different sub-problems that make different assumptions about target domain data and labels available. Once a new domain is included, the available data typically evolves over time according to the depicted arrow.

References

1. S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.
2. J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, and S. Y. Philip. Generalizing to unseen domains: A survey on domain generalization. *IEEE transactions on knowledge and data engineering*, 35(8):8052–8072, 2022.
3. M. Wang and W. Deng. Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135–153, 2018.
4. G. Wilson and D. J. Cook. A survey of unsupervised deep domain adaptation. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(5):1–46, 2020.
5. Y. Yang and T. M. Hospedales. A unified perspective on multi-domain and multi-task learning. *arXiv preprint arXiv:1412.7489*, 2014.
6. K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C. C. Loy. Domain generalization: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4396–4415, 2022.

Electrocardiogram for Biometric Recognition: Collectability, Stability and Application Challenges

Teresa M.C. Pereira[1,2,3] (✉), Raquel Sebastião[1,4], Raquel C. Conceição[2], and Vitor Sencadas[3]

[1] IEETA, DETI, LASI, Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal
[2] IBEB, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
[3] CICECO, DEMaC, Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal
[4] CISeD, Instituto Politécnico de Viseu, Campus Politécnico, 3504-510 Viseu
teresamcp@ua.pt

Biometric recognition systems play a crucial role in modern security, offering advantages over traditional methods such as passwords. Electrocardiogram (ECG)-based recognition has gained attention among emerging biometric modalities due to its universality, liveness detection capability [1], and strong resistance to spoofing. Unlike fingerprints or facial features, ECG signals provide dynamic physiological information unique to each individual, which makes them difficult to replicate and, thus, a highly secure option for recognition systems [2]. However, for the implementation of ECG biometrics, some challenges arise, including the discomfort and setup complexity of traditional Ag/AgCl electrodes, which require skin preparation and can lead to skin irritation, as well as high signal variability caused by physical activity, emotional states, and long-term physiological changes [3].

This PhD project aims to overcome these limitations by developing a novel solid-state ECG electrode based on Polydimethylsiloxane (PDMS) and silver-ink technology. The proposed solution enhances biocompatibility, eliminating the need for skin preparation, and allows an easier and faster setup for fingertip ECG acquisition, which is suitable for daily or mobile use. Additionally, the project investigates how various physiological conditions affect ECG signals and biometric recognition performance, with a strong focus on intra- and inter-subject variability. Three ECG acquisition protocols are designed: (1) multi-session recordings to evaluate stability over time, (2) physical activity to assess motion artifacts and postural effects, and (3) emotional elicitation via video stimuli to analyze changes induced by fear and happiness. Signals will undergo a preprocessing step - removal of noises and normalization - and features will be extracted through fiducial, non-fiducial, and hybrid approaches. For classification, both machine learning, e.g., Support Vector Machines (SVM) and k-Nearest Neighbors, and deep learning, e.g., Convolutional Neural Networks (CNNs) models will be implemented. Thus, the main goal of the project is to develop a reproducible,

robust, and practical ECG-based biometric recognition system that is accurate and reliable under diverse conditions, for real-world applications.



Fig. 1. Graphical Abstract of the proposed PhD project.

Acknowledgement: This work was funded by national funds through FCT — Fundação para a Ciência e a Tecnologia, I.P., under the PhD grant UI/BD/153605/2022 (https://doi.org/10.54499/UI/BD/153605/2022) (T.M.C.P.), under unit 00127-IEETA, and within the RD units UID/00645/2025 (IBEB), 2022.08973.PTDC, and CICECO-Aveiro Institute of Materials UIDB/50011/2020 (DOI 10.54499/UIDB/50011/2020), UIDP/50011/2020 (DOI 10.54499/UIDP/50011/2020) LA/P/0006/2020 (DOI 10.54499/LA/P/0006/2020), financed by national funds through the FCT/MCTES (PIDDAC).

References

1. Ranjeet Srivastva, Ashutosh Singh, and Yogendra Narain Singh. Plexnet: A fast and robust ECG biometric system for human recognition. *Information Sciences*, 558:208–228, 2021.
2. Gokhan Guven, Hakan Gurkan, and Umit Guz. Biometric identification using fingertip electrocardiogram signals. *Signal, Image and Video Processing*, 12:933–940, 2018.
3. Miguel Carvalho and Susana Brás. Addressing intra-subject variability in electrocardiogram-based biometric systems through a hybrid architecture. *Biomed. Signal Process. Control.*, 87:105465, 2024.

Symbolic Representation Analysis for Change Detection in Physiological Signals Using Compression-Based Similarity

Daniela Pais[1*], Susana Brás[1], Manuel Vico Ávalos[2,3], and Raquel Sebastião[1,4]

[1] IEETA, DETI, LASI, University of Aveiro, Aveiro, Portugal
[2] Unidade Local de Saúde Viseu Dão-Lafões, Viseu, Portugal
[3] Faculty of Health Sciences, University of Beira Interior, Portugal
[4] CISeD – Research Centre in Digital Services, Polytechnic University of Viseu, Campus Politécnico, Viseu, Portugal
[*]danielapais@ua.pt

This work aims to develop a method to detect changes in physiological signals over time to support real-time identification of noise, sensor errors, or context-specific events in monitoring and decision support systems. The initial focus is on Electrocardiogram (ECG) signals collected during postoperative recovery, where detecting changes may help clinicians assess and manage patient pain. While this clinical scenario serves as the primary motivation, a key goal is to design an approach that can be adapted to other physiological signals and application domains. The proposed method combines symbolic time series representation with compression-based similarity analysis. A sliding window is applied to the signal, and each window is transformed into a symbol sequence using a modified Symbolic Aggregate approXimation (SAX) algorithm. The original SAX reduces dimensionality by segmenting a time series and mapping each segment to a symbol based on breakpoints obtained from a Gaussian distribution [2], which may inadequately represent physiological data. To overcome this limitation, we estimate subject-specific ECG probability distributions using Kernel Density Estimation (KDE) and derive breakpoints tailored to the individual. Change detection is performed by computing similarity between symbolic representations using the Normalized Compression Distance (NCD), a measure based on Kolmogorov complexity [1]. Two strategies are employed: (i) sequential comparison, where each symbolic window is compared to the previous one to detect local changes; and (ii) reference-based comparison, where each window is compared to a baseline to identify deviations from typical patterns. Initial experiments on postoperative ECG data show the method can detect variations in similarity during patient recovery, suggesting sensitivity to temporal changes in physiological state. While no consistent correlation has yet been observed between these variations and the timing of analgesic administration, this may be influenced by confounding clinical factors. Current efforts are directed at refining the method by improving probability density estimation and optimizing key parameters, including window size, SAX representation settings, and the compression algorithm. Future work will focus on validating the approach on other physiological signals and evaluating its utility in broader change detection tasks.

Acknowledgments. This work is supported by FCT - Fundação para a Ciência e Tecnologia, I.P., under the PhD grant no. UI/BD/153374/2022 with DOI identifier https://doi.org/10.54499/UI/BD/153374/2022 (D.P.). This work is also funded by FCT, through national funds, within the scope of the UID/00127 project (IEETA/UA, http://www.ieeta.pt/).

References

1. Li, M., Chen, X., Li, X., Ma, B., Vitanyi, P.M.B.: The similarity metric. IEEE Transactions on Information Theory **50**(12), 3250–3264 (2004). https://doi.org/10.1109/TIT.2004.838101
2. Lin, J., Keogh, E., Wei, L., Lonardi, S.: Experiencing sax: a novel symbolic representation of time series. Data Mining and Knowledge Discovery **15**, 107–144 (2007). https://doi.org/https://doi.org/10.1007/s10618-007-0064-z

A Machine Learning-based Data Quality Framework for Regularly Sampled Sensors

Catarina Canastra[1], António Casimiro[1], and Pedro M. Ferreira[1]

LASIGE, Faculty of Sciences, University of Lisbon, Portugal
ccanastra@lasige.di.fc.ul.pt

1 Introduction

Sensor faults often affect measurements, compromising their quality and decreasing the dependability of cyber-physical systems (CPS) that are consuming the data. Data quality is a topic of concern in CPSs, which has been surveyed for several application domains [1–3].

2 Challenges

Assessing and improving data quality enhances the dependability of CPS. While multiple data quality frameworks with distinct foundations and architectures have been developed [4, 5], these frameworks: a) are typically designed ad hoc in a system dependent way, as they are building on particular concepts [4–6]; b) lack dependability sense, for instance focusing solely on assessing data quality [5, 7]; c) do not have broad applicability, as error detection [8, 9] and error correction [6, 8, 9] are typically focused on specific types of errors; and d) lack seamless configuration, as they are designed for particular system requirements [7, 10, 11].

3 Methodology

This PhD aims at proposing and developing a data quality framework leveraging machine learning techniques to enhance the dependability of CPSs that regularly collect data from sensors. Figure 1 presents the components of the framework, from the collection of data quality requirements and contextual information, to data and information provision to the consuming system. To accomplish this objective and address the identified challenges, the following specific objectives were defined:

1. Define a conceptual framework for sensor data quality by identifying semantically orthogonal and relevant quality concepts;
2. Define a data quality architecture that supports offline and online operation of data quality pipelines;
3. Define and implement machine learning-based data quality assessment (as shown in Figure 2) and improvement approaches capable of addressing different types of errors;

4. Implement data quality pipelines for a broad class of CPSs;
5. Demonstrate the broad applicability and effectiveness of the data quality framework, architecture, and pipelines.

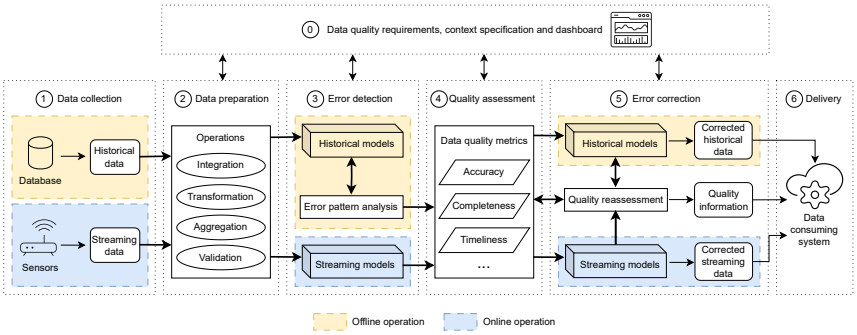


Fig. 1. High-level architecture of the data quality framework. A supervision dashboard is used for user interaction, requirements specification, context definition, and data visualization. The framework may operate in offline or online mode, consuming, respectively, historical data or real-time sensor data. Input data is prepared through integration, normalization, aggregation, and validation to ensure it is ready for quality checks, starting with error detection. The latter can be done using machine learning models, which are trained and tested using previously collected and prepared data. The quality of data is assessed through data quality metrics such as accuracy, completeness, and timeliness, in light of the predefined requirements and context data. Erroneous or low quality data are corrected also using machine learning models, which are specialized for each type of error. The resulting corrected data, quality metrics, and original data, are delivered to the consuming system.

References

1. Sha, K., Zeadally, S.: Data Quality Challenges in Cyber-Physical Systems. Journal of Data and Information Quality, 2015. https://doi.org/10.1145/2740965
2. Castaño, F., Strzelczak, S., Villalonga, A., Haber, R.E., Kossakowska, J.: Sensor Reliability in Cyber-Physical Systems Using Internet-of-Things Data: A Review and Case Study. Remote Sensing 11, 2019. https://doi.org/10.3390/rs11192252
3. Byabazaire, J., O'Hare, G., Delaney, D.: Data Quality and Trust: Review of Challenges and Opportunities for Data Sharing in IoT. Electronics, 2020. https://doi.org/10.3390/electronics9122083
4. Geisler, S., Quix, C., Weber, S., Jarke, M.: Ontology-Based Data Quality Management for Data Streams. Journal of Data and Information Quality, 2016. https://doi.org/10.1145/2968332
5. Kuemper, D., Iggena, T., Toenjes, R., Pulvermueller, E.: Valid. IoT: a framework for sensor data quality analysis and interpolation. In: Proceedings of the 9th ACM Multimedia Systems Conference, 2018. https://doi.org/10.1145/3204949.3204972

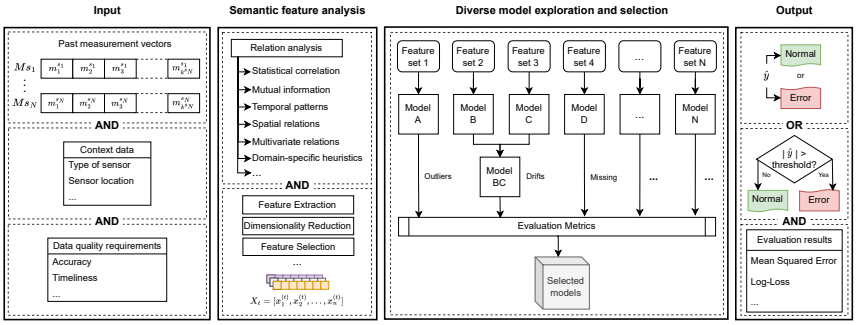


Fig. 2. Machine learning approach for developing error detection models. Input data typically includes past measurement vectors, context data, and data quality requirements. Semantic feature analysis allows uncovering temporal and spatial relations within single and multiple data streams, and preparing various feature sets with potentially meaningful information for the models. Multiple models, diverse feature sets, and combinations thereof, can be used to handle error detection. Various evaluation strategies and metrics analyze the output of the selected models to detect normal or erroneous data.

6. D'Aniello, G., Gaeta, M., Hong, T.-P.: Effective Quality-Aware Sensor Data Management. IEEE Transactions on Emerging Topics in Computational Intelligence, 2017. https://doi.org/10.1109/TETCI.2017.2782800
7. Cappiello, C., Samá, W., Vitali, M.: Quality awareness for a Successful Big Data Exploitation. In: Proceedings of the 22nd International Database Engineering and Applications Symposium, 2018. https://doi.org/10.1145/3216122.3216124
8. Sousa, I., Casimiro, A., Cecílio, J.: Artificial Neural Networks for Real-Time Data Quality Assurance. ACM SIGAda Ada Letters 42, 2022. https://doi.org/10.1145/3577949.3577966
9. Jesus, G., Casimiro, A., Oliveira, A.: Using Machine Learning for Dependable Outlier Detection in Environmental Monitoring Systems. ACM Transactions on Cyber-Physical Systems 5, 2021. https://doi.org/10.1145/3445812
10. Karthik, N., Ananthanarayana, V.S.: Context Aware Trust Management Scheme for Pervasive Healthcare. Wireless Personal Communications 105, 2019. https://doi.org/10.1007/s11277-018-6091-9
11. Yuan, T., Xiong, W., Liu, S.: Large-Scale Data Quality Challenges, Framework and Evaluation in Metro Systems. IEEE Transactions on Big Data, 2025. https://doi.org/10.1109/TBDATA.2024.3474215

Diffusion-based Adversarial Purification for Robust Deep Learning

Jia Fu[1,2], Anders Holst[1,2], and Sepideh Pashami[1,3]

[1] RISE Research Institutes of Sweden, Kista 164 40, Sweden
{jia.fu, anders.holst, sepideh.pashami}@ri.se
[2] KTH Royal Institute of Technology, Stockholm 114 28, Sweden
[3] Halmstad University, Halmstad 301 18, Sweden

Keywords: Adversarial Defense · Vision-Language Models · Diffusion Process.

1 Background

Deep learning models, notably the popular Vision-Language Models (VLMs) [1, 8], are vulnerable to adversarial perturbations [14], which are imperceptible to humans but can drastically mislead model inference [6]. While adversarial fine-tuning [10] improves robustness under specific threat models, it often fails to generalize against diverse attacks. Generative purification [13], though attack-agnostic, typically suffers from considerable processing overhead. This highlights the need for defense mechanisms that are both efficient and versatile. Our research is based on diffusion process [12], where Gaussian noise is injected through a forward diffusion to disrupt the adversarial pattern and then adversarial noise is removed together with Gaussian noise by a reverse diffusion [9].

2 Methodology

We introduce DiffCAP [4], a diffusion-based cumulative adversarial purification framework for VLMs. DiffCAP adds Gaussian noise to adversarial perturbed images progressively until their latent embeddings stabilize, indicated by a semantic similarity threshold between consecutive embeddings. We establish a certified recovery region within the forward diffusion process, theoretically proving that adversarial perturbations can be neutralized after sufficient noise injection. Furthermore, we quantify the embedding convergence rate, justifying our adaptive similarity threshold acquisition. Finally, a pretrained diffusion model [3] reverses the noisy images to clean representation suitable for VLM inference.

Complementarily, we propose DiffPAD [5], a diffusion-based adversarial patch decontamination aimed at mitigating localized patch attacks [2]. DiffPAD first leverages conditional diffusion to restore downsampled images. The resulting reconstruction errors correlate linearly with the size of adversarial patches, enabling precise patch localization via a dynamic binarization and sliding window approach. Subsequently, DiffPAD replaces the closed-form solution to guide diffusion for inpainting masked areas.

3 Discussion

Extensive experiments across various attack settings, target models, datasets, and task scenarios demonstrate that both DiffCAP and DiffPAD achieve state-of-the-art robustness compared to selected adversarial defenses. The future work will extend their applicability to a wider range of perturbation types such as textual adversarial modifications [7] and irregularly shaped visual patches [11].

References

1. Awadalla, A., Gao, I., Gardner, J., Hessel, J., Hanafy, Y., Zhu, W., Marathe, K., Bitton, Y., Gadre, S., Sagawa, S., et al.: Openflamingo: An open-source framework for training large autoregressive vision-language models. arXiv preprint arXiv:2308.01390 (2023)
2. Brown, T.B., Mané, D., Roy, A., Abadi, M., Gilmer, J.: Adversarial patch. arXiv preprint arXiv:1712.09665 (2017)
3. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021)
4. Fu, J., Wu, Y., Chen, Y., Peng, K., Zhang, X., Cevher, V., Pashami, S., Holst, A.: Diffcap: Diffusion-based cumulative adversarial purification for vision language models. arXiv preprint arXiv:2506.03933 (2025)
5. Fu, J., Zhang, X., Pashami, S., Rahimian, F., Holst, A.: Diffpad: Denoising diffusion-based adversarial patch decontamination. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 6602–6611. IEEE (2025)
6. Goodfellow, I.J., Shlens, J., Szegedy, C.: Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572 (2014)
7. Guo, C., Sablayrolles, A., Jégou, H., Kiela, D.: Gradient-based adversarial attacks against text transformers. arXiv preprint arXiv:2104.13733 (2021)
8. Liu, H., Li, C., Li, Y., Lee, Y.J.: Improved baselines with visual instruction tuning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 26296–26306 (2024)
9. Nie, W., Guo, B., Huang, Y., Xiao, C., Vahdat, A., Anandkumar, A.: Diffusion models for adversarial purification. arXiv preprint arXiv:2205.07460 (2022)
10. Schlarmann, C., Singh, N.D., Croce, F., Hein, M.: Robust clip: Unsupervised adversarial fine-tuning of vision embeddings for robust large vision-language models. arXiv preprint arXiv:2402.12336 (2024)
11. Sharif, M., Bhagavatula, S., Bauer, L., Reiter, M.K.: A general framework for adversarial examples with objectives. ACM Transactions on Privacy and Security (TOPS) **22**(3), 1–30 (2019)
12. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456 (2020)
13. Yoon, J., Hwang, S.J., Lee, J.: Adversarial purification with score-based generative models. In: International Conference on Machine Learning. pp. 12062–12072. PMLR (2021)
14. Zhao, Y., Pang, T., Du, C., Yang, X., Li, C., Cheung, N.M.M., Lin, M.: On evaluating adversarial robustness of large vision-language models. Advances in Neural Information Processing Systems **36**, 54111–54138 (2023)

Information Leakage in Generative Models

Catarina Gomes[1], Ricardo Mendes[2], and João P. Vilela[1]

[1] CRACS/INESC TEC and Dept. of Computer Science Faculty of Sciences, University of Porto, Porto, Portugal
[2] CISUC and Dept. of Informatics Engineering University of Coimbra Coimbra, Portugal
up201804545@edu.fc.up.pt, rscmendes@dei.uc.pt, jvilela@fc.up.pt

1 Research Challenges and Problem

Despite the vast use of generative models (GMs) to replace real private data, there is evidence of the side-effect in privacy-protection [8] that motivates the need to discern fair from harmful modeling (with Information Leakage, IL). IL occurs when GMs replicate memorized training samples [1–3] or in more subtle ways by generating samples that share a particular source of uniqueness from the training set [4, 7] (Subtle Memorization, SM). However, the state of the art (SOTA) fails to capture SM by focusing on exact replication as the only sign of IL, which underestimates leakage and overestimates compliance with data protection laws. SOTA also fails to characterize inherently high-risk data by focusing on population-level indicators (duplication [1, 3, 4, 6]) and model-related factors (model size [3, 6]), potentially missing causes at the sample-level. A limited comprehension of unknown forms of IL and its causes, prevent a better understanding of privacy implications of GMs and motivates us to study IL that advents from memorization in GMs.

2 Research Questions and Goals

To address these gaps, the following questions will be addressed:

*Q*1) *What subtle forms of memorization negatively impact privacy in GMs?*

We aim to design a theoretical framework to define and measure SM by extracting training properties instead of images from popular GMs [5], enhancing this work with the extended study of the effect of several factors. This study could be extended to the natural language domain, where the definition of exact extraction is not established [6].

*Q*2) *How to characterize inherently high-risk data?*

We aim to model privacy risk at the sample-level identifying the underlying characteristics that contribute to exact and subtle IL with statistical indicators at the sample-level, starting to quantify the training data's role on its own leakage.

*Q*3) *What approaches can be developed to control IL in GMs?*

We aim to answer Q3 using the underlying characteristics of high-risk data to control exact and subtle IL.

The research problem and these contributions will be presented as a poster, where the relevance within the current research landscape, the intended purpose, and the potential impact will be clearer.

Acknowledgments. This work is funded by national funds through FCT - Fundação para a Ciência e a Tecnologia, I.P., under the support UID/50014/2023 (https://doi.org/10.54499/UID/50014/2023).

References

1. Carlini, N., Hayes, J., Nasr, M., Jagielski, M., Sehwag, V., Tramèr, F., Balle, B., Ippolito, D., Wallace, E.: Extracting training data from diffusion models. In: Proceedings of the 32nd USENIX Conference on Security Symposium. SEC '23, USENIX Association, USA (2023)
2. Carlini, N., Liu, C., Erlingsson, Ú., Kos, J., Song, D.: The secret sharer: Evaluating and testing unintended memorization in neural networks. In: 28th USENIX Security Symposium (USENIX Security 19). pp. 267–284. USENIX Association, Santa Clara, CA (2019)
3. Carlini, N., Tramèr, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., Roberts, A., Brown, T., Song, D., Erlingsson, Ú., Oprea, A., Raffel, C.: Extracting training data from large language models. In: 30th USENIX Security Symposium (USENIX Security 21). pp. 2633–2650. USENIX Association (2021)
4. Naseh, A., Roh, J., Houmansadr, A.: Memory triggers: Unveiling memorization in text-to-image generative models through word-level duplication. arXiv preprint arXiv:2312.03692 (2023)
5. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10684–10695 (June 2022)
6. Schwarzschild, A., Feng, Z., Maini, P., Lipton, Z., Kolter, J.Z.: Rethinking llm memorization through the lens of adversarial compression. In: Globerson, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J., Zhang, C. (eds.) Advances in Neural Information Processing Systems. vol. 37, pp. 56244–56267. Curran Associates, Inc. (2024)
7. Somepalli, G., Singla, V., Goldblum, M., Geiping, J., Goldstein, T.: Diffusion art or digital forgery? investigating data replication in diffusion models. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6048–6058 (2023)
8. Stadler, T., Oprisanu, B., Troncoso, C.: Synthetic data – anonymisation groundhog day. In: 31st USENIX Security Symposium (USENIX Security 22). pp. 1451–1468. USENIX Association, Boston, MA (Aug 2022)

Improving Multi-Task Parameter-Efficient Fine-Tuning

Robert Belanec[1,2] (✉), Ivan Srba[2], and Maria Bielikova[2]

[1] Faculty of Information Technology, Brno University of Technology, Brno, Czechia
[2] Kempelen Institute of Intelligent Technologies, Bratislava, Slovakia
{name.surname}@kinit.sk

The growing scale of the successful transformer-based large language models (LLMs) contributed to the increase in computational requirements for training and evaluation. To tackle this problem, parameter-efficient fine-tuning [2] (PEFT) methods were introduced. Multi-task PEFT methods give us the benefit of having only a single expert model per multiple tasks. However, current multi-task PEFT methods often rely on a specific selection of source tasks and do not investigate the sensitivity of PEFT methods to the various multi-task learning factors influencing the performance and efficiency of multi-task PEFT.

Our goal within the dissertation is to improve the multi-task PEFT methods and explore their multi-task learning factors that affect their efficiency and performance. To achieve it, we propose two research objectives (RO): **RO1**: *Detailed investigation of multi-task learning factors affecting the performance and efficiency of multi-task PEFT methods*, and **RO2**: *Contribute to the field of multi-task PEFT methods that are robust to multi-task learning factors, effectively transfer to multi-lingual settings, and generalize well to in-domain and new tasks.*

To answer our research objective, our **preliminary work** includes **replication of the ATTEMPT [1] method**, showing performance instabilities and sensitivity to source task selection. We introduce **PEFT-Factory**, an open-source benchmark for unified evaluation of PEFT methods. Finally, we propose **Task Prompt Vectors** (TPV), a modular and more efficient method of combining soft prompts, introduced in our paper accepted to the ECML PKDD 2025 (attached to this submission).

Our **future plans** for the dissertation include designing a novel benchmark specifically tailored to evaluate multi-task learning factors (RO1), improving the efficiency and robustness of multi-task PEFT by introducing novel PEFT methods (RO2), and evaluating multi-task PEFT in a multilingual environment by utilizing unlearning techniques (RO2).

Acknowledgments. This work was partially funded by the European Union, under the project lorAI - Low Resource Artificial Intelligence, GA No. 101136646; and by the European Union NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I01-03-V04-00006.

References

1. Asai, A., Salehi, M., Peters, M., Hajishirzi, H.: ATTEMPT: Parameter-efficient multi-task tuning via attentional mixtures of soft prompts. In: Goldberg, Y., Kozareva, Z., Zhang, Y. (eds.) Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. pp. 6655–6672. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Dec 2022). https://doi.org/10.18653/v1/2022.emnlp-main.446, https://aclanthology.org/2022.emnlp-main.446
2. Han, Z., Gao, C., Liu, J., Zhang, J., Zhang, S.Q.: Parameter-efficient fine-tuning for large models: A comprehensive survey. Transactions on Machine Learning Research (2024), https://openreview.net/forum?id=lIsCS8b6zj

Developing Robust Methods for Efficient Question Generation over Knowledge Base in Open Environments [PhD Forum]

Runhao Zhao[1], Jiuyang Tang[1(✉)], Xianghan Wang[2], Xiaolu Liu[2], Lei He[2], Lizhen Wu[3], and Sijie He[3]

[1] Laboratory for Big Data and Decision, National University of Defense Technology, Changsha, 410073, China
{runhaozhao,jiuyang_tang}@nudt.edu.cn
[2] College of Systems Engineering, National University of Defense Technology, China
{lxl_sunny,helei}@nudt.edu.cn
[3] Graduate School, National University of Defense Technology, China
lzwu@nudt.edu.cn,woshidou_dou@126.com

Abstract. Knowledge base question generation (KBQG) generates natural language questions from a knowledge base (KB) and target answers. While effective on **high-quality**, **small-scale**, **sufficient annotated data** and **compute**, existing methods falter in open environments with **low data quality**, **large-scale KBs**, and **limited training resources**. This PhD project proposes a set of robust KBQG methods suited for such settings, enhancing adaptability and knowledge management. Partial results have been published in **Information Processing & Management** and **CIKM 2024**.

Keywords: KBQG · open environments · robust methods

1 Introduction

Knowledge base question generation (KBQG) supports question answering and dialogue systems by generating questions from knowledge bases (KBs) [3, 4]. Methods fall into template-based [5, 9], neural network-based [1, 2, 7, 8], and LLM-based [6,12] categories. However, these often fail in open environments with (Fig. 1): 1)*Low Data Quality:* Diverse user inputs not matched by rigid training data [13]; 2) *Large-scale Data:* Massive KBs like YAGO [10, 11] exceed model capacity; 3) *High Training Costs:* Scarce labels and compute hinder training and generalization [12].

2 Research Design

This PhD project address key challenges as follows (Tab. 1):

- *KBQG in Low-Quality Data Scenarios:* We propose BC-RLHF, a bias-corrected reinforcement learning with human feedback framework, improving question

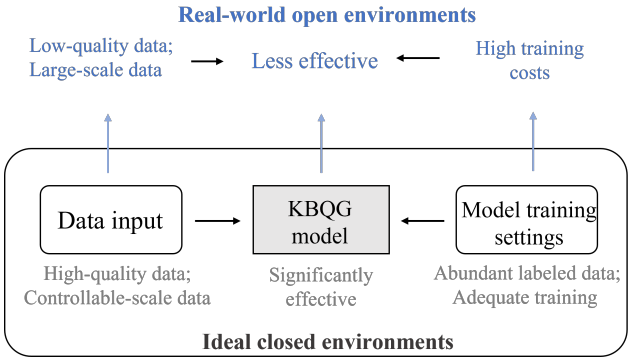


Fig. 1. Key challenges faced by KBQG in open environments

diversity and naturalness (**Published in Information Processing & Management [13]**).

- *KBQG in Large-Scale Data Scenarios:* A bidirectional graph partitioning and subgraph reasoning method is being developed to enhance scalability.
- *Low-Cost Training for KBQG:* We introduce RoleAgentQG, a zero-shot multi-agent system combining LLMs and small models for efficient training (**Accepted at CIKM 2024 [12]**).
- *KBQG Integration Verification:* Future work will unify these methods into a scenario-adaptive system and evaluate performance in open environments.

Table 1. Progress of PhD Research Tasks

Research Task	KBQG in Low-Quality Data	KBQG in Large-Scale Data	Low-Cost Training for KBQG	KBQG Integration Verification
Progress	Completed [13]	Ongoing	Completed [12]	Ongoing

References

1. Bi, S., Liu, J., Miao, Z., Min, Q.: Difficulty-controllable question generation over knowledge graphs: A counterfactual reasoning approach. Information Processing & Management (2024), https://api.semanticscholar.org/CorpusID:268502796
2. Chen, Y., Wu, L., Zaki, M.J.: Toward subgraph-guided knowledge graph question generation with graph neural networks. IEEE Transactions on Neural Networks and Learning Systems pp. 1–12 (2023). https://doi.org/10.1109/TNNLS.2023.3264519
3. Chen, Z., Li, D., Zhao, X., Hu, B., Zhang, M.: Temporal knowledge question answering via abstract reasoning induction. In: Ku, L., Martins, A., Srikumar, V. (eds.) Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024. pp. 4872–4889. Association for Computational Linguistics

(2024). https://doi.org/10.18653/V1/2024.ACL-LONG.267, https://doi.org/10.18653/v1/2024.acl-long.267
4. Guo, S., Zhang, J., Wang, et al, Y.: Dsm: Question generation over knowledge base via modeling diverse subgraphs with meta-learner. In: Conference on Empirical Methods in Natural Language Processing (2022)
5. Jia, R., Liang, P.: Data recombination for neural semantic parsing. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers. The Association for Computer Linguistics (2016). https://doi.org/10.18653/v1/p16-1002, https://doi.org/10.18653/v1/p16-1002
6. Liang, Y., Wang, J., Zhu, H., Wang, L., Qian, W., Lan, Y.: Prompting large language models with chain-of-thought for few-shot knowledge base question generation. In: Bouamor, H., Pino, J., Bali, K. (eds.) Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, EMNLP 2023, Singapore, December 6-10, 2023. pp. 4329–4343. Association for Computational Linguistics (2023). https://doi.org/10.18653/V1/2023.EMNLP-MAIN.263, https://doi.org/10.18653/v1/2023.emnlp-main.263
7. Mulla, N., Gharpure, P.: Automatic question generation: a review of methodologies, datasets, evaluation metrics, and applications. Prog. Artif. Intell. **12**(1), 1–32 (2023). https://doi.org/10.1007/S13748-023-00295-9, https://doi.org/10.1007/s13748-023-00295-9
8. Orogat, A., El-Roby, A.: Maestro: Automatic generation of comprehensive benchmarks for question answering over knowledge graphs. Proc. ACM Manag. Data **1**(2), 177:1–177:24 (2023). https://doi.org/10.1145/3589322, https://doi.org/10.1145/3589322
9. Seyler, D., Yahya, M., Berberich, K.: Knowledge questions from knowledge graphs. In: Kamps, J., Kanoulas, E., de Rijke, M., Fang, H., Yilmaz, E. (eds.) Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval, ICTIR 2017, Amsterdam, The Netherlands, October 1-4, 2017. pp. 11–18. ACM (2017). https://doi.org/10.1145/3121050.3121073, https://doi.org/10.1145/3121050.3121073
10. Suchanek, F.M., Alam, M., Bonald, T., Chen, L., Paris, P., Soria, J.: YAGO 4.5: A large and clean knowledge base with a rich taxonomy. In: Yang, G.H., Wang, H., Han, S., Hauff, C., Zuccon, G., Zhang, Y. (eds.) Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024. pp. 131–140. ACM (2024). https://doi.org/10.1145/3626772.3657876, https://doi.org/10.1145/3626772.3657876
11. Suchanek, F.M., Kasneci, G., Weikum, G.: Yago: a core of semantic knowledge. In: Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007. pp. 697–706. ACM (2007)
12. Zhao, R., Tang, J., Zeng, W., Chen, Z., Zhao, X.: Zero-shot knowledge graph question generation via multi-agent llms and small models synthesis. In: Serra, E., Spezzano, F. (eds.) Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, CIKM 2024, Boise, ID, USA, October 21-25, 2024. pp. 3341–3351. ACM (2024). https://doi.org/10.1145/3627673.3679805, https://doi.org/10.1145/3627673.3679805
13. Zhao, R., Tang, J., Zeng, W., Guo, Y., Zhao, X.: Towards human-like questioning: Knowledge base question generation with bias-corrected reinforcement learning from human feedback. Inf. Process. Manag. **62**(3), 104044 (2025). https://

doi.org/10.1016/J.IPM.2024.104044, https://doi.org/10.1016/j.ipm.2024.104044