Exploring Data Point Interactions: A Dual Representation Approach for Enhanced Machine Learning

Mohamed Karim Belaid[1,2]

[1]IDIADA Fahrzeugtechnik GmbH, Munich, Germany
[2]Dr. Ing. h.c. F. Porsche AG, Stuttgart, Germany

Abstract

In recent years, Explainable Artificial Intelligence (XAI) has attracted significant attention due to the growing complexity and opacity of ML models. While traditional XAI tools have focused on feature interaction analysis, there is a gap in understanding data point interactions and their impact on model performance. This research addresses this gap by studying data point interactions in ML models, specifically KNN, and proposing novel algorithms to enhance prediction performance.

First, we developed **Compare-xAI** [1], a benchmark for direct comparison of popular XAI algorithms across various use cases, confirming the literature's immaturity in leveraging data interactions. Second, we introduced **STI-KNN** [2], the first algorithm to calculate exact pair-interaction Shapley values in polynomial time, improving our understanding of Shapley contributions for data point pairs. Third, we utilized data interactions to create the **Pairwise Difference Classifier** algorithm [3], which solves a binary classification problem on a paired version of the original training data.

Keywords

Supervised learning, Meta-learning, Data Interaction, Explainable AI

1. Related Work

Existing literature in XAI has extensively explored feature interactions, providing insights into how different features contribute to model predictions. Previous works, such as Ribeiro et al.'s LIME [4] and Lundberg and Lee's SHAP [5], have laid the groundwork for feature-based explanations, leading to mature research literature in the field of feature interaction [6, 7, 8, 9]. But the domain of data valuation remains relatively unexplored, besides recent pioneer works using approximation [10] or model specific [11] explanations. Moreover, the interaction between data points themselves has not been thoroughly investigated.

Tynes et al. introduced pairwise difference regressor [12], a novel meta-learner for chemical tasks that enhances prediction performance, compared to random forest and provides robust uncertainty quantification. In computational chemistry, estimating differences between data points helps mitigate systematic errors [12]. In parallel, Wetzel et al. used twin neural network architectures for semi-supervised regression tasks, focusing on predicting differences between target values of distinct data points [13].

2. Research Questions and Challenges

2.0.1. RQ.1: How do different XAI methods, especially data-based, perform in terms of interpretability, accuracy, and computational efficiency?

We question the maturity and usability of explanation methods for the data science community. The abundance of xAI algorithms can be overwhelming, making it hard for practitioners to select the right one for their needs. The differing requirements and implementations of xAI algorithms pose challenges for data scientists in accurately evaluating them and staying current with their development. This

Discovery Science - 27th International Conference, DS 2024, Pisa, Italy, October 14-16, 2024, Proceedings, CEUR
karim.belaid@idiada.com (M. K. Belaid)
0000-0001-9038-9045 (M. K. Belaid)

issue manifests as the illusion of explanatory depth [14] in interpreting xAI results [15], with evidence showing that data scientists often misuse interpretability tools [16].

2.0.2. RQ.2: How can we accurately quantify data point interactions within trained ML models?

Existing literature primarily focuses on feature interactions, lacking methodologies for evaluating data point interactions. How do data points interact in forming patterns? How can we measure this interaction? And how can we leverage this explanation to improve the ML pipeline?

2.0.3. RQ.3: Can data point interactions be leveraged to enhance the performance of ML classifiers?

Exploring whether understanding data point interactions can lead to improved model accuracy and robustness. How can we design and implement algorithms that utilize data point interactions for prediction tasks? Developing and evaluating algorithms that incorporate data point interactions in their predictive mechanisms.

3. Method and Evaluation

3.0.1. RQ.1

Given the unsolved burden of evaluating and correctly choosing xAI algorithms, we propose Compare-xAI that mitigates two issues: non-unified benchmark for xAI algorithms and the illusion of explanatory depth during the interpretation of results. Compare-xAI emerges as a unique and valuable benchmark. Its distinct contributions lie in its simplicity, scalability, ability to integrate any dataset and ML model, and, most importantly, its focus on the user's expected explanation. By addressing the pitfalls highlighted in surveys of xAI algorithms through concrete functional tests, Compare-xAI provides a robust evaluation framework.

3.0.2. RQ.2

We propose, **STI-KNN**, the first algorithm that calculates the exact pair-interaction Shapley values in $\mathcal{O}(tn^2)$ rather than $\mathcal{O}(2^n)$. STI-KNN is the first algorithm that allows studying the exact interaction on large real-world datasets. This research is the first to consider two disjoint fields: Data valuation and Interaction in Explainable AI. Finally, we study various cases of positive and negative data interactions using STI-KNN.

3.0.3. RQ.3

Leveraging the concept of data point interactions, we introduce the Pairwise Difference Learning (PDL) Classifier. This classifier employs a dual representation of the ML task, achieving better prediction performance by integrating pair interaction data, see Figure 1. The empirical evaluation contains 99 diverse datasets, times 25 CV repetitions. We use the macro F1 metric.

4. Preliminary Results

4.0.1. RQ.1

With 15 post-hoc xAI algorithms, 25 tests, and 50 research papers indexed, Compare-xAI offers a unified benchmark that accurately reproduces experiments. Through a rigorous selection protocol, it highlights the contrast between theoretical foundations and practical implementations, making the limitations of each method transparent. Compare-xAI uses an intuitive scoring method to absorb the vast quantity of

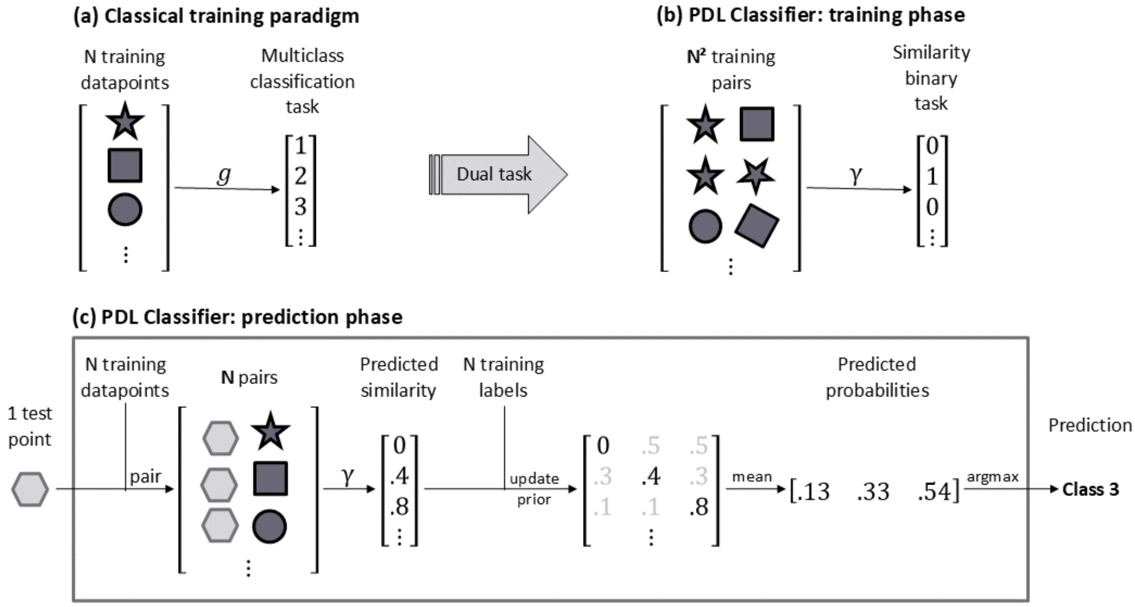

Figure 1: Illustration of the PDL classifier.

xAI-related papers and reduce human errors in interpreting xAI outputs. Its goal is to unify post-hoc xAI evaluation methods into a multi-dimensional benchmark, providing insights into the strengths and weaknesses of different approaches. Link: https://karim-53.github.io/cxai/

4.0.2. RQ.2

Thanks to the STI-KNN algorithm, the data interaction can quickly be visualized using a heatmap of the Shapley interaction values. The matrix shows an example of interaction. We observe, first, a contrast between in-class and out-of-class interactions, second, a reduction in interaction due to data redundancy, and third, an unusual pattern when data contains outliers.

4.0.3. RQ.3

Our benchmark demonstrates that PDL consistently outperforms state-of-the-art ML models, resulting in improved F1 scores in a majority of cases. This highlights PDL's effectiveness in enhancing performance over baseline methods, facilitated through its straightforward integration via our Python package. Link: https://github.com/Karim-53/pdll

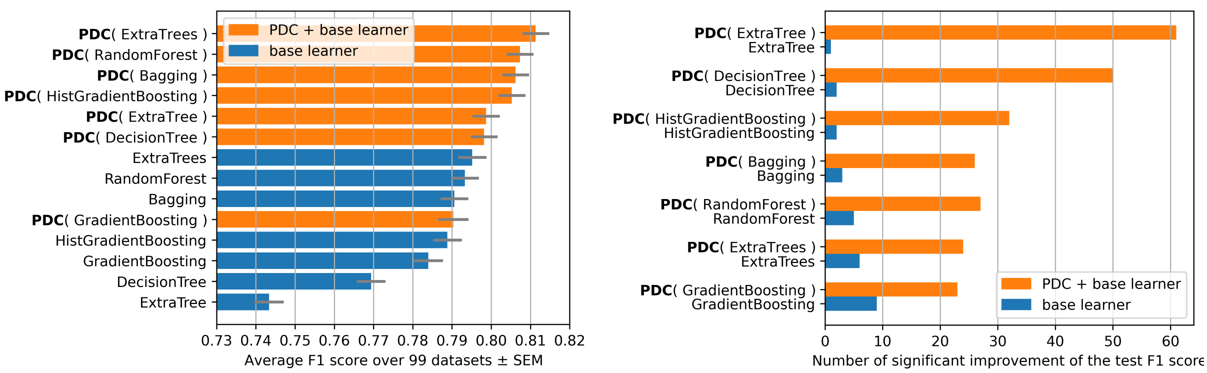

Figure 2: Evaluating PDL classifier on 99 datasets. Significant wins use 25 CV repetitions on each dataset.

5. Intermediary Conclusions

Our research indicates that data point interactions play a crucial role in the performance of ML models. By shifting the focus from feature interactions to data interactions, we have opened up new avenues for enhancing model interpretability and accuracy. For more detailed results, refer to the following papers[1, 2, 3].

6. Planned Next Steps

Confirming the efficiency of the PDL algorithm by studying its calibration and uncertainty estimation. By continuing to explore the interactions between data points, we hope to contribute significantly to the field of xAI and ML, ultimately leading to more transparent, accurate, and robust models.

References

[1] M. K. Belaid, E. Hüllermeier, M. Rabus, R. Krestel, Do we need another explainable ai method? toward unifying post-hoc xai evaluation methods into an interactive and multi-dimensional benchmark, arXiv preprint arXiv:2207.14160 (2022).

[2] M. K. Belaid, D. El Mekki, M. Rabus, E. Hüllermeier, Optimizing Data Shapley Interaction calculation from $\mathcal{O}(2^N)$ to $\mathcal{O}(TN^2)$ for KNN models, arXiv preprint arXiv:2304.01224 (2023).

[3] M. K. Belaid, M. Rabus, E. Hüllermeier, Pairwise difference learning for classification, arXiv preprint arXiv:2406.20031 (2024).

[4] M. T. Ribeiro, S. Singh, C. Guestrin, " why should i trust you?" explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135–1144.

[5] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Advances in neural information processing systems 30 (2017).

[6] M. Muschalik, F. Fumagalli, B. Hammer, E. Hüllermeier, Beyond treeshap: Efficient computation of any-order shapley interactions for tree ensembles, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 38, 2024, pp. 14388–14396.

[7] F. Fumagalli, M. Muschalik, P. Kolpaczki, E. Hüllermeier, B. Hammer, Shap-iq: Unified approximation of any-order shapley interactions, Advances in Neural Information Processing Systems 36 (2024).

[8] F. Fumagalli, M. Muschalik, P. Kolpaczki, E. Hüllermeier, B. Hammer, Kernelshap-iq: Weighted least-square optimization for shapley interactions, arXiv preprint arXiv:2405.10852 (2024).

[9] P. Kolpaczki, M. Muschalik, F. Fumagalli, B. Hammer, E. Hüllermeier, Svarm-iq: Efficient approximation of any-order shapley interactions through stratification, arXiv preprint arXiv:2401.13371 (2024).

[10] A. Ghorbani, J. Zou, Data Shapley: Equitable valuation of data for ML, in: International Conference on ML, PMLR, 2019, pp. 2242–2251.

[11] R. Jia, D. Dao, B. Wang, F. A. Hubis, N. M. Gurel, B. Li, C. Zhang, C. J. Spanos, D. Song, Efficient task-specific data valuation for nearest neighbor algorithms, arXiv preprint arXiv:1908.08619 (2019).

[12] M. Tynes, W. Gao, D. J. Burrill, E. R. Batista, D. Perez, P. Yang, N. Lubbers, Pairwise difference regression: A ML meta-algorithm for improved prediction and uncertainty quantification in chemical search, Journal of Chemical Information and Modeling 61 (2021) 3846–3857.

[13] S. J. Wetzel, R. G. Melko, I. Tamblyn, Twin neural network regression is a semi-supervised regression algorithm, ML: Science and Technology 3 (2022) 045007.

[14] L. Rozenblit, F. Keil, The misunderstood limits of folk science: An illusion of explanatory depth, Cognitive science 26 (2002) 521–562.

[15] M. Chromik, M. Eiband, F. Buchner, A. Krüger, A. Butz, I think i get your point, ai! the illusion

of explanatory depth in explainable ai, in: 26th International Conference on Intelligent User Interfaces, 2021, pp. 307–317.
[16] H. Kaur, H. Nori, S. Jenkins, R. Caruana, H. Wallach, J. Wortman Vaughan, Interpreting interpretability: understanding data scientists' use of interpretability tools for machine learning, in: Proceedings of the 2020 CHI conference on human factors in computing systems, 2020, pp. 1–14.