Fast and accurate regional effect plots for automated tabular data analysis

TaDA Workshop @ VLDB 2024

Vasilis Gkolemis[1,2] Christos Diou[2] Eirini Ntoutsi[3] Theodore Dalamagas[1]

[1]ATHENA Research and Innovation Center

[2]Harokopio University of Athens

[3]University of the Bundeswehr Munich

August 2024

Program

1. ML + XAI $\rightarrow$ a good Data Analysis Pipeline (5')
2. RegionalRHALE: a good XAI choice (4')
3. `Effector` - a Python Package for Feature Effect (1')

Problem Statement

Tabular data

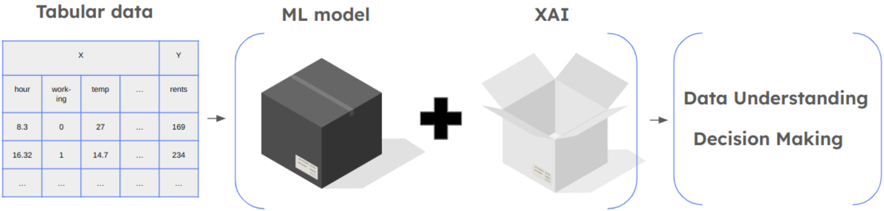
X				Y
hour	work-ing	temp	...	rents
8.3	0	27	...	169
16.32	1	14.7	...	234
...	...	...	...	...

Data Understanding Pipeline

Data Understanding

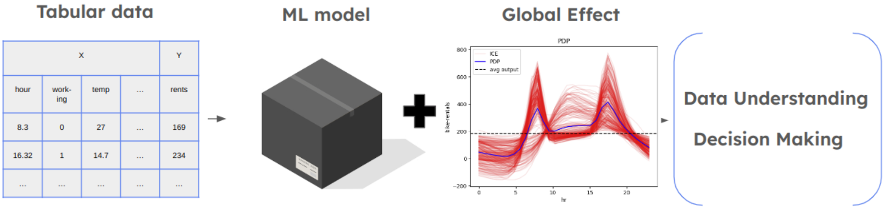
Decision Making

Idea 1



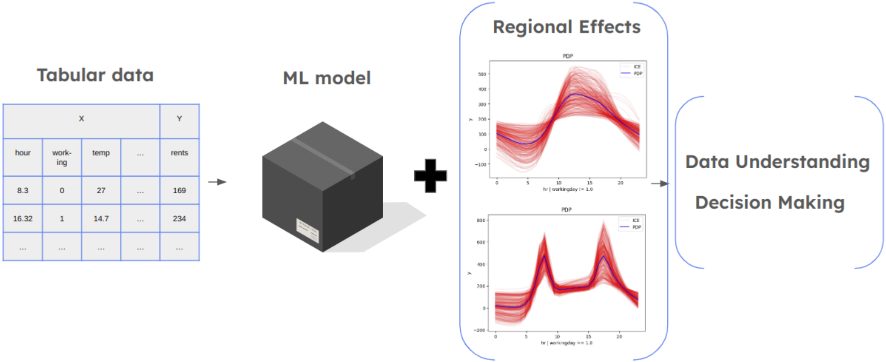
Black box ML model + XAI = a Data Analysis pipeline!

Idea 2



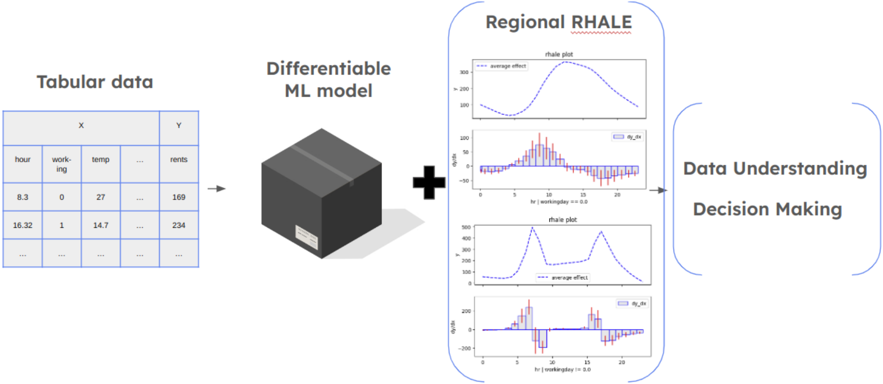
Global effects is a good XAI choice!

Idea 3



Regional effects is a better XAI choice!

Idea 4



Use RegionalRHALE if the black box model is differentiable!

Bike-sharing dataset

- hourly count of bike-rentals (2011, 2012)
- Design-matrix X:
  - year, month, day, hour
  - working day vs. non-working day
  - temperature
  - humidity
  - windspeed
- Target variable Y:
  - bike-rentals per hour
    - $Y_\mu = 189.5$
    - $Y_\sigma = 181.4$
- Decision Making: decide a discount policy
- Data Understanding: how bike rental market works

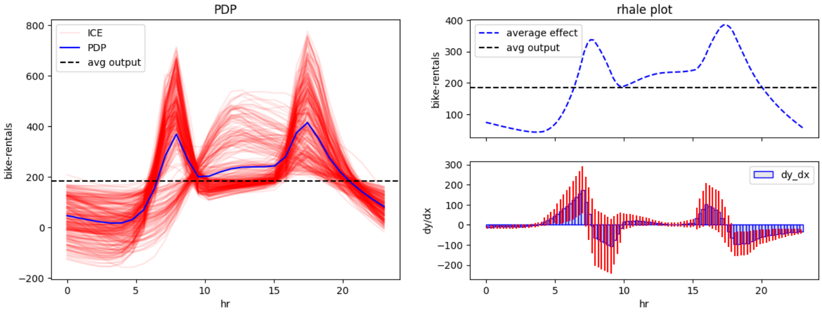
Proposed pipeline: Fit and Explain

- decide a discount policy
  - which hour of the day to apply the discount
  - how the feature $x_{\texttt{hour}}$ relates to $y_{\texttt{bike\_rentals}}$
- Step 1: Fit a black-box ML model
  - Could be any ML model
  - a Neural Network achieves $\texttt{RMSE} \approx 45.35$ counts $(0.25 Y_\sigma)$
- Step 2: Use feature effect
  - Global effect: $x_{\texttt{hour}}$ vs $y_{\texttt{bike\_rentals}}$ globally
  - Regional effect: $x_{\texttt{hour}}$ vs $y_{\texttt{bike\_rentals}}$ regionally

---

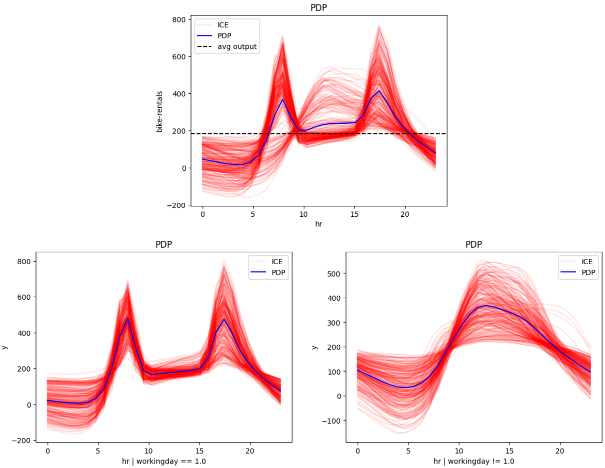
Let's see!

Global Effect: PDP and RHALE



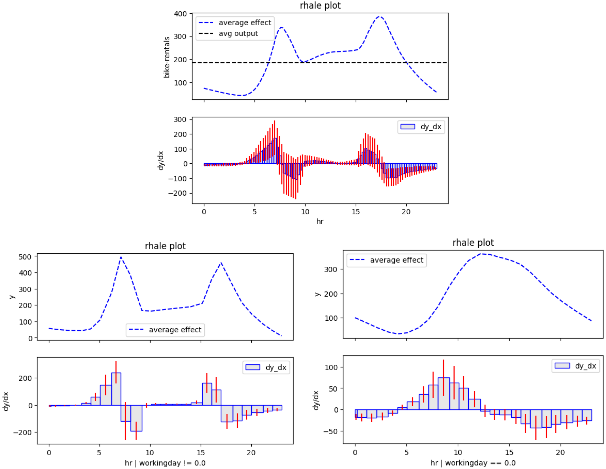
PDP and RHALE (Gkolemis et al., 2023b) are global effect methods

Regional Effect: Regional-PDP



Regional PDP (Herbinger, Bischl, and Casalicchio, 2022)

Regional Effect: Regional-RHALE



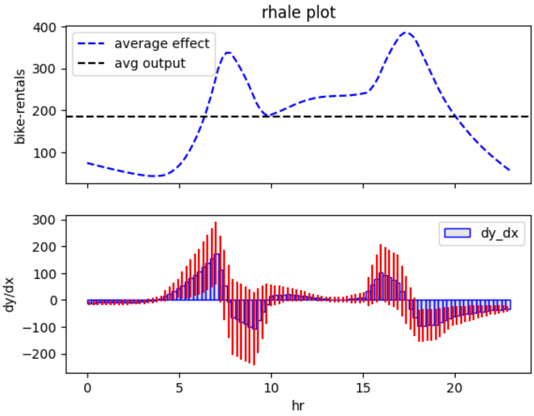
Regional RHALE - our proposal!

Program

1. ML + XAI $\rightarrow$ a good Data Analysis Pipeline (5')
2. RegionalRHALE: a good XAI choice (4')
3. `Effector` - a Python Package for Feature Effect (1')

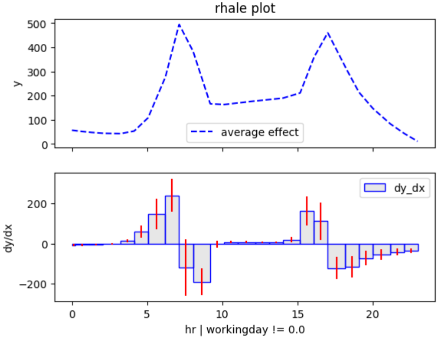
RegionalRHALE - How it works (a)

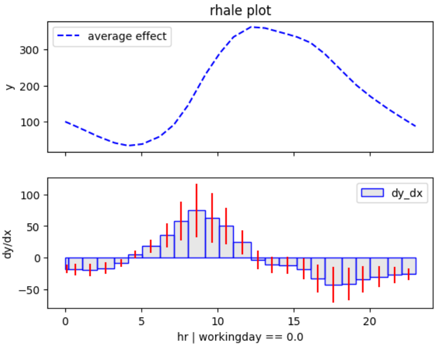
- RHALE plot (Gkolemis et al., 2023b)
- red bars express the heterogeneity



Regional RHALE - How it works (b)

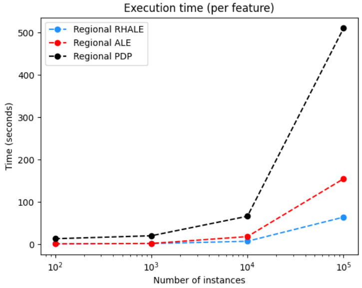
- iterate over all other features
- select the split with the maximum heterogeneity reduction

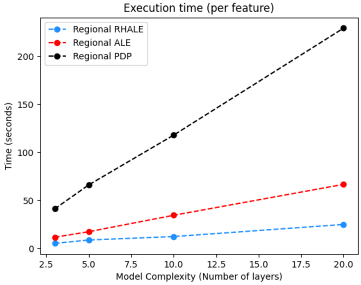




Regional RHALE is fast

- iterating over all other features $\rightarrow$ is slow
- needs fast evaluation of the heterogeneity
- if model is differentiable, regional RHALE is very fast
- regional RHALE treats well cases with correlated features





Program

1. ML + XAI $\rightarrow$ a good Data Analysis Pipeline (5')
2. RegionalRHALE: a good XAI choice (4')
3. `Effector` - a `Python` Package for Feature Effect (1')

Effector - a Python package for feature effect

- Implements:
  - many global effect methods (PDP, RHALE, SHAP-DP)
  - many regional effect methods (regionalPDP, regionalRHALE, regionalSHAP-DP)
- Work in progress
- If you are interested, please use it and give feedback
- Source: https://github.com/givasile/effector
- Documentation: https://xai-effector.github.io/

References I

Apley, Daniel W. and Jingyu Zhu (2020). “Visualizing the effects of predictor variables in black box supervised learning models”. In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 82.4, pp. 1059–1086. ISSN: 14679868. DOI: 10.1111/rssb.12377. arXiv: 1612.08468.

Friedman, Jerome H and Bogdan E Popescu (2008). “Predictive learning via rule ensembles”. In: *The annals of applied statistics.* Publisher: JSTOR, pp. 916–954.

Gkolemis, Vasilis, Theodore Dalamagas, and Christos Diou (Oct. 2022). “DALE: Differential Accumulated Local Effects for efficient and accurate global explanations”. In: *Asian Conference on Machine Learning (ACML).*

Gkolemis, Vasilis et al. (2023a). “Regionally Additive Models: Explainable-by-design models minimizing feature interactions”. In: *arXiv preprint arXiv:2309.12215.*

References II

Gkolemis, Vasilis et al. (2023b). "RHALE: Robust and Heterogeneity-Aware Accumulated Local Effects". In: *ECAI 2023*. IOS Press, pp. 859–866.

Goldstein, Alex et al. (Mar. 2014). *Peeking Inside the Black Box: Visualizing Statistical Learning with Plots of Individual Conditional Expectation*. en. arXiv:1309.6392 [stat]. URL: http://arxiv.org/abs/1309.6392 (visited on 01/23/2023).

Herbinger, Julia, Bernd Bischl, and Giuseppe Casalicchio (Feb. 2022). *REPID: Regional Effect Plots with implicit Interaction Detection*. arXiv:2202.07254 [cs, stat]. DOI: 10.48550/arXiv.2202.07254. URL: http://arxiv.org/abs/2202.07254 (visited on 06/11/2023).

— (2023). "Decomposing Global Feature Effects Based on Feature Interactions". In: *arXiv preprint arXiv:2306.00541*.

References III

Lundberg, Scott M and Su-In Lee (2017). “A unified approach to interpreting model predictions”. In: *Advances in neural information processing systems* 30.