Hybrid Intelligence with Commonality Plots: A First Aid Kit for Domain Experts and a Translation Device for Data Scientists

Nikolas Stege[1,2], Michael H. Breitner[1]

[1] Leibniz Universität Hannover, Information Systems Institute, Hannover, Germany
{stege,breitner}@iwi.uni-hannover.de
[2] Ernst & Young GmbH, Hannover, Germany
{nikolas.stege}@de.ey.com

Abstract. There is a large gap between domain experts capable to identify business needs and data scientists who use insight producing algorithms, but often fail to connect these to the bigger picture. A major challenge for companies and organizations is to integrate practical data science into existing teams and workflows. We are driven by the assumption that efficient data science requires cross-disciplinary teams able to communicate. We present a methodology that enables domain experts and data scientists to analyze and discuss findings and implications together. Motivated by a typical problem from auditing we introduce a visualization method that helps to detect unusual data in a subset and highlights potential areas for investigation. The method is a first aid kit applicable regardless whether unusual samples were detected by manual selection of domain experts or by algorithms applied by data scientists. An applicability check shows how the visualizations facilitate collaboration of both parties.

Keywords: Commonality Plots, Domain Knowledge, Hybrid Intelligence, Visualization, Data Science

1 Introduction and Motivation

Referring to the term data science the spotlight is usually put on the application of state of the art models, machine learning algorithms, on how to tune the algorithms' hyperparameters and on how to optimize scalability and overall performance. Those are all very important aspects in the world of data analytics. However, especially in the context of business intelligence and the overall creation of value there is large potential for information systems researchers and practitioners in the field of data science besides the optimization of algorithms, see [1-2] for an overview. Efficient data science requires a skillset that covers the range from the engineering side (data capturing and processing) to the business side (domain expertise and storytelling) and it comes to no surprise that individuals with such cross-disciplinary skillsets are rare [3]. Because of that shortage, the biggest challenge is to figure out how to efficiently integrate practical data science into existing teams, workflows and processes. We are guided by the

15th International Conference on Wirtschaftsinformatik,
March 08-11, 2020, Potsdam, Germany