Contextual Importance and Utility: a Theoretical Foundation

Kary Främling
Umeå University, Department of Computing Science
kary.framling@cs.umu.se

UMEÅ UNIVERSITY
DEPARTMENT OF COMPUTING SCIENCE

Abstract

The paper provides new theory to support the eXplainable AI (XAI) method Contextual Importance and Utility (CIU). CIU arithmetic is based on the concepts of Multi-Attribute Utility Theory, which gives CIU a solid theoretical foundation. The novel concept of *contextual influence* is also defined, which makes it possible to compare CIU directly with so-called additive feature attribution (AFA) methods for model-agnostic outcome explanation. One key takeaway is that the 'influence' concept used by AFA methods can be inadequate for outcome explanation purposes even for simple models to explain. Experiments with simple models show that explanations using contextual importance (CI) and contextual utility (CU) produce explanations where influence-based methods fail. It is also shown that CI and CU guarantees explanation faithfulness towards the explained model.

Introduction

Contextual Importance and Utility (CIU) was originally proposed by Kary Främling in 1995. CIU is a model-agnostic post-hoc explanation method and provides uniform explanation concepts most black-box models f ranging from linear models such as the weighted sum, to rule-based systems, decision trees, fuzzy systems, neural networks and any machine learning-based models.

Main Objectives

- Present a solid mathematical theory for CIU.
- Provide distinct definitions of the concepts *influence*, *importance* and *utility*.
- Define the new concept of *contextual influence* derived from CIU.
- Situate CIU within the latest state-of-the-art in XAI and show that it performs better than current main-stream XAI methods.

Materials and Methods

Additive Feature Attribution (AFA) Methods

AFA methods use an *explanation model* g of the original model f:

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i, \tag{1}$$

where $z' \in \{0,1\}^M$, M is the number of simplified input features, and $\phi \in \mathbb{R}$ is the influence of feature i. *Shapley value* and *LIME* are AFA methods.

Decision Theory and Multi-Attribute Utility Theory

Decision Theory proposes a set of quantitative methods for reaching optimal, or at least rational, decisions. An optimal decision is one that maximizes the expected utility. An additive n-attribute utility function is expressed as:

$$u(x_1, \ldots, x_n) = \sum_{i=1}^{n} k_i u_i(x_i) \tag{2}$$

where u and the u_i are normalized to the range [0,1], and the k_i are normalization constants [**?**].

Contextual Importance and Utility (CIU)

CIU estimates the values k_i and $u_i(x_i)$ in Equation 2 for one or more input features $\{i\}$ in a specific context C and any black-box model f, where the context is defined by the instance to be explained.

Definition 1 (Contextual Importance)**.**

$$CI_j(C, \{i\}, \{I\}) = \frac{umax_j(C, \{i\}) - umin_j(C, \{i\})}{umax_j(C, \{I\}) - umin_j(C, \{I\})}, \tag{3}$$

where $\{i\} \subseteq \{I\}$ and $\{I\} \subseteq \{1, \ldots, n\}$. C is the instance/context to be explained and defines the values of input features that do not belong to $\{i\}$ or $\{I\}$.

The *Contextual Utility (CU)* corresponds to the factor $u_i(x_i)$ in Equation 2. CU expresses to what extent the current value of a given input feature contributes to obtaining a high output utility u_j.

Definition 2 (Contextual Utility)**.**

$$CU_j(C, \{i\}) = \frac{u_j(C) - umin_j(C, \{i\})}{umax_j(C, \{i\}) - umin_j(C, \{i\})} \tag{4}$$

Contextual influence

$$\phi = (rmax - rmin) \times CI \times (CU - neutral.CU) \tag{5}$$

where '$_j(C, \{i\})$' has been omitted from all three terms ϕ, CI, and CU for easier readability.

Results

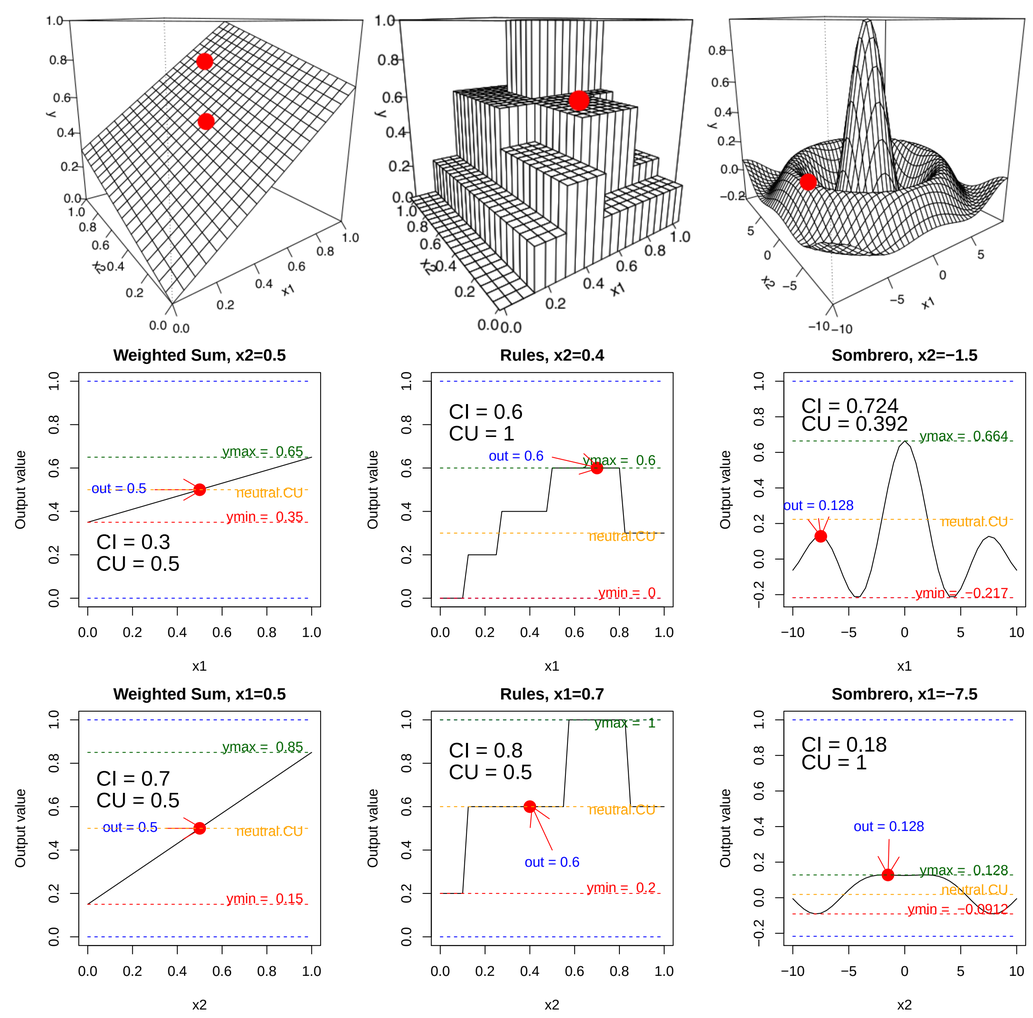

Figure 1: Illustration of how CI and CU are calculated for different functions.

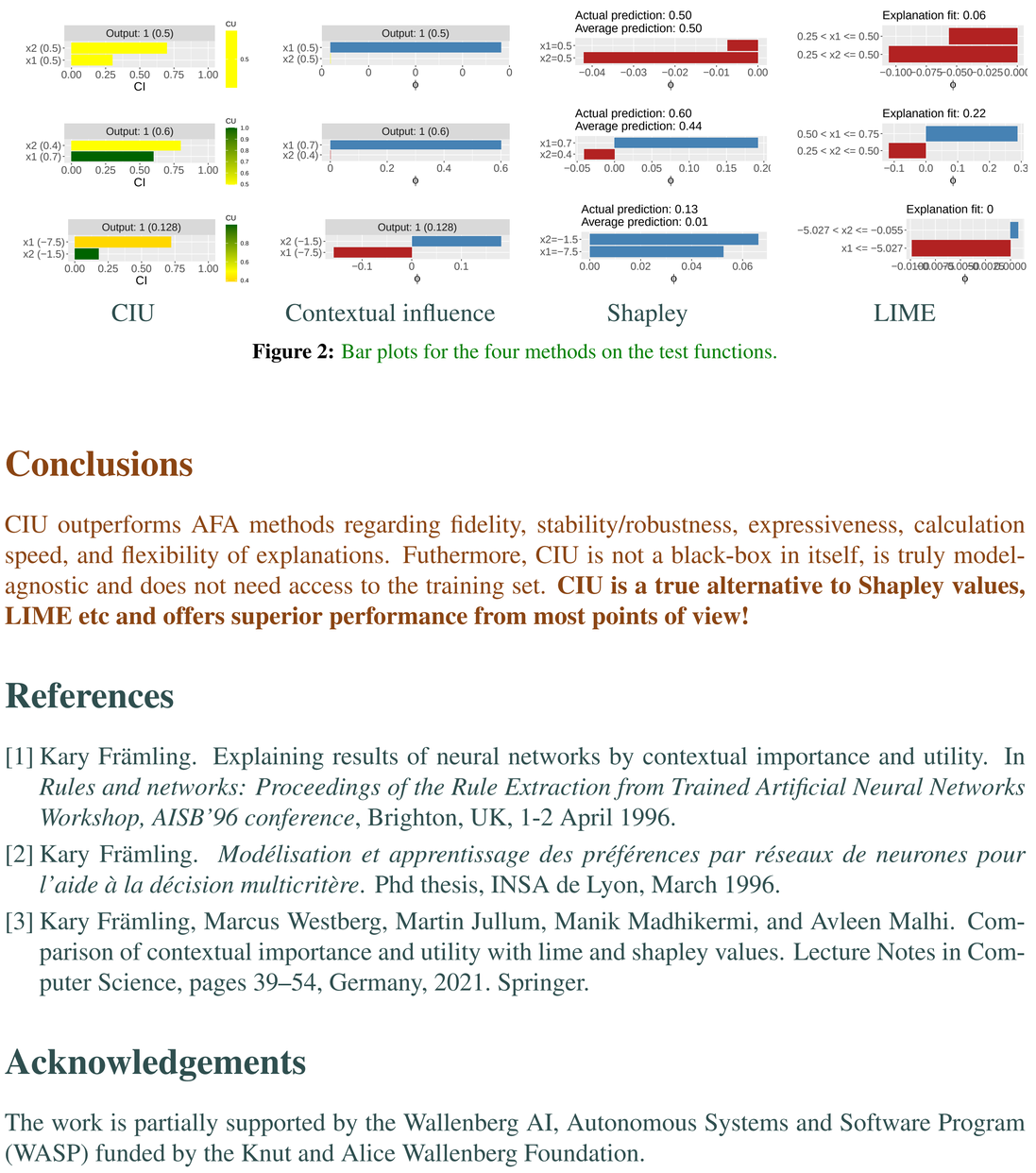

Figure 2: Bar plots for the four methods on the test functions.

Conclusions

CIU outperforms AFA methods regarding fidelity, stability/robustness, expressiveness, calculation speed, and flexibility of explanations. Futhermore, CIU is not a black-box in itself, is truly model-agnostic and does not need access to the training set. **CIU is a true alternative to Shapley values, LIME etc and offers superior performance from most points of view!**

References

[1] Kary Främling. Explaining results of neural networks by contextual importance and utility. In *Rules and networks: Proceedings of the Rule Extraction from Trained Artificial Neural Networks Workshop, AISB'96 conference*, Brighton, UK, 1-2 April 1996.

[2] Kary Främling. *Modélisation et apprentissage des préférences par réseaux de neurones pour l'aide à la décision multicritère*. Phd thesis, INSA de Lyon, March 1996.

[3] Kary Främling, Marcus Westberg, Martin Jullum, Manik Madhikermi, and Avleen Malhi. Comparison of contextual importance and utility with lime and shapley values. Lecture Notes in Computer Science, pages 39–54, Germany, 2021. Springer.

Acknowledgements

The work is partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.