Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

SHAPLEY-SHUBIK EXPLANATIONS OF FEATURE IMPORTANCE

Gayane Grigoryan

Department of Engineering Management and Systems Engineering
Old Dominion University
5115 Hampton Blvd
Norfolk, VA, 23529, USA

ABSTRACT

Explaining feature importance values in models is a central concern in the realm of explainable artificial intelligence (XAI). While the Shapley value has garnered significant attention, there are other promising cooperative game theory (CGT) solutions, such as the Shapley-Shubik, that have not received the same amount of attention. In this paper, we explore the potential of the Shapley-Shubik method for elucidating feature importance values in simulations and machine learning models.

1 INTRODUCTION

Assessing how a feature affects a model's prediction is a fundamental question in simulation and machine learning. This analysis provides valuable insights into the model's behavior. Various methods have been developed to measure a feature's contribution to the model's predictions (Lundberg and Lee 2017).

A cooperative game-theoretic approach, Shapley value, has been instrumental in providing a principled and rigorous framework for measuring the importance of each feature in a machine learning model (Lundberg and Lee 2017). Despite its popularity, Shapely value has several limitations, as shown by (Kumar, Venkatasubramanian, Scheidegger, and Friedler 2020), such as mathematical and human-centric issues, or the additivity constraint associated with Shapley-value-based explanations. In this paper, we suggest Shapley-Shubik as an alternative method for gauging feature importance in simulations and machine learning models

2 BACKGROUND

Cooperative game theory (CGT) models scenarios where participants cooperate for mutual benefit, a concept applicable in simulation and machine learning (Myerson 1997). The cooperative game involves the model and its feature space, with coalitions representing subsets of features. This theory has been used in explainable AI (XAI) to allocate feature importance fairly based on their contributions to model performance.

3 METHODS

In this paper, we introduce the Shapley-Shubik feature importance values and compare the results with commonly used Shapley value methods.

Shapley values are computed using the Eq. 1, where N is the total number of features, S is the subset of features and $(v(S\cup\{i\})-v(s)$ is the value of the model with and without the feature i.

$$\phi_i(v)=\sum_{s\subseteq N\setminus\{i\}}\frac{|S|!(|N|-|S|-1)!)}{|N|!}(v(S\cup\{i\})-v(s)) \tag{1}$$

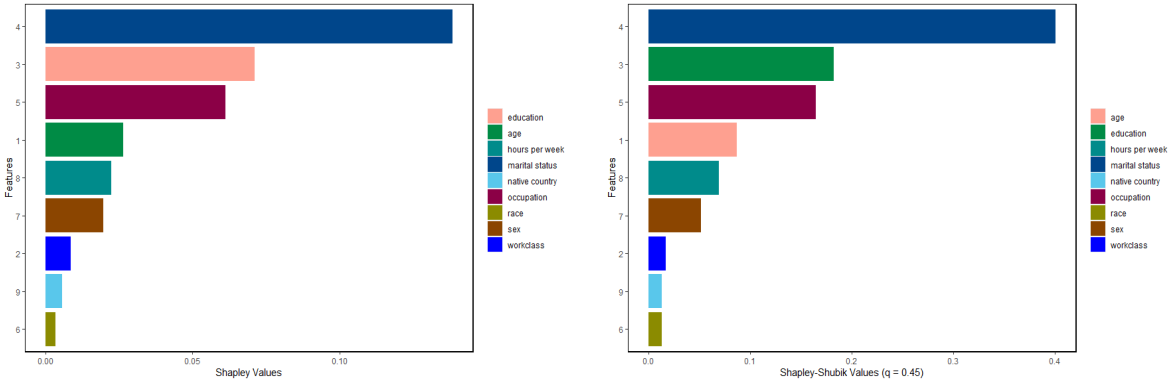
The Shapley-Shubik power index can be expressed as:

$$\phi_i = \sum \frac{(s-1)!(n-s)!}{n!} \tag{2}$$

with $\sum_{i \in N} \phi_i = 1$, and $\phi = (\phi_1, ..., \phi_n)$, where $s = |S|$ is the number of voters in set S. The summation is taken over all winning coalitions S for which S without i, $S - \{i\}$ is losing. The Shapley-Shubik determines the number of sequences in which player i is pivotal over all possible orderings of n players. A player i is pivotal or swing for a coalition S if the player i turns S from losing $v(S) = 0$ to a winning coalition $v(S) = 1$ by joining that coalition.

4 RESULTS

The aim of this study was to examine the importance of various features in a model. We consider the US adult dataset from the 1994 US Census Bureau database and conducted a regression model to determine feature importance values. The results of the analysis are presented in Figure 1.



(a) Shapley values feature importance

(b) Shapley-Shubik feature importance

Figure 1: Shapley values vs Shapley-Shubik feature importance analysis results

5 CONCLUSION

In conclusion, considering other cooperative game theory techniques such as Shapley-Shubik can help overcome limitations associated with the Shapley value. Also, Shapley-Shubik can be more interpretable, as they allow for experimentation with different threshold values to observe how feature importance values vary with adjustments. Nevertheless, further experimentation is necessary to validate these observations.

REFERENCES

Kumar, I. E., S. Venkatasubramanian, C. Scheidegger, and S. Friedler. 2020. "Problems with Shapley-value-based Explanations as Feature Importance Measures". In *International Conference on Machine Learning*, 5491–5500. PMLR.

Lundberg, S. M., and S.-I. Lee. 2017. "A Unified Approach to Interpreting Model Predictions". *Advances in Neural Information Processing Systems* 30.

Myerson, R. B. 1997. *Game Theory: Analysis of Conflict*. Harvard University Press.