Social Ranking for Feature Selection

Extended Abstract

Laurent Gourvès
LAMSADE, CNRS, Université Paris-Dauphine, Université PSL
75016, Paris, France
laurent.gourves@lamsade.dauphine.fr

Stefano Moretti
LAMSADE, CNRS, Université Paris-Dauphine, Université PSL
75016, Paris, France
stefano.moretti@lamsade.dauphine.fr

Satya Tamby
LAMSADE, CNRS, Université Paris-Dauphine, Université PSL
75016, Paris, France
satya.tamby@cnrs.fr

ABSTRACT

In this paper, we focus on limitations in the use of the Shapley value within the field of eXplainable AI (XAI) through the lens of the axiomatic analysis and its implications in the realm of machine learning. As an alternative to the Shapley value, we analyse the properties of the lex-cel, a social ranking solution introduced in the recent literature at the intersection between coalitional games and social choice theory, showing that axioms characterizing the lex-cel, under certain circumstances, are more suitable for ranking features in machine learning models, compared to those satisfied by the Shapley value. Via experiments conducted on public datasets, we also show that the lex-cel outperforms a commonly employed feature selection algorithm based on the Shapley value, in particular with respect to the capacity of selecting less redundant features.

KEYWORDS

Cooperative Games; Social Ranking; Feature Selection; Shapley value; Axioms; Explainability

ACM Reference Format:
Laurent Gourvès, Stefano Moretti, and Satya Tamby. 2025. Social Ranking for Feature Selection: Extended Abstract. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 3 pages.

1 INTRODUCTION

During the last decade, the Shapley value [16] for coalitional games has seen considerable success as a method to convert the information about the quality of a ML model's performance (over all possible subsets of features) into a single-feature numerical attribution of importance for the model's prediction [6, 13, 14]. Nevertheless, some recent studies have raised important concerns about the interpretation of the Shapley value regarding its ability to rank features based on their relevance in constructing simplified models [7]. In this paper, we further investigate this issue with a twofold objective: (1) identifying properties of the Shapley value that seem less suited for feature selection, and alternative foundational properties more adapted to provide meaningful rankings of features; (2) using the new identified axioms for guiding the design of a novel feature selection method, comparing it with the Shapley value on basic instances and on more articulated experiments.

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

We first explore, on numerical examples, the effects of the *additivity* axiom on feature selection. Then, we identify a possible alternative to the use of additivity by means of the combination of other properties borrowed from the theory of social ranking [4, 8]. To be more specific, a social ranking aims to rank single elements of a finite set according to their position in a ranking over coalitions. Thanks to some recent results shown in [1], we argue that two properties specific for social rankings (namely, *Coalitional Anonymity (CA) and Independence of the Worst Set (IWS) [4]*) together with other properties that are also satisfied by the Shapley value (namely, *Symmetry* [16] and *Strict Desirability* [10, 15]) lead to a compelling method for feature selection, already known in the literature of social ranking as *lexicographic-excellence* (lex-cel in short [1, 4]; see also [2, 3, 17] for related lexicographic social rankings). Finally, we provide an experimental analysis showing that the method based on the lex-cel to construct simplified models (called the *LeXAI* method) outperforms the SHAP method on several data-sets, using the standard approach introduced in [14] to evaluate the prediction performance of submodels. We argue that the LeXAI method allows to select sets of features showing a lower level of correlation than the features selected by the SHAP method, making it more effective for constructing simplified models. To address the computational issues of LeXAI, we also discuss an approximate version based on a limited number of coalitions.

We start in Section 2 with a short introduction to our property-driven analysis. We continue in Section 3 with a summary of some experimental results. Section 4 concludes with some future research directions. A more comprehensive version of this paper can be obtained from the authors upon request.

2 THE MODEL AND ITS PROPERTIES

An *evaluation function* (e.f.) on a set N of n features is a map $v : 2^N \to \mathbb{R}$ assigning to each $S \in 2^N$ (2^N denotes the set of all subsets of N), a real number $v(S) \in \mathbb{R}$. A subset of features $S \in 2^N$ is also called a *coalition* and $v(S)$ represents the performance (measured according to predefined metrics) of the prediction provided by the ML submodel restricted to features in S. Different approaches to compute an evaluation function v have been proposed in the literature (see, for instance, [5, 7, 12] for a discussion on this topic). In the following, we will denote by $\mathcal{E}^N$ the class of all e.f.s on the set N. We define a *ranking solution* as a function $R : \mathcal{E}^N \to \mathcal{R}(N)$ that maps any evaluation function $v \in \mathcal{E}^N$ into a ranking $R^v \in \mathcal{R}(N)$ on N, where $\mathcal{R}(N)$ denotes the set of all possible rankings on N (a ranking on N is a binary relation in $N \times N$ that is also total and transitive). For any evaluation function $v \in \mathcal{E}^N$, define the partition $\Sigma_1^v, \ldots, \Sigma_m^v \subseteq 2^N \setminus \{\emptyset\}$ of

coalitions (i.e., elements of 2^N) such that the following two conditions hold: (i) $v(S) = v(T)$ for all $S, T \in \Sigma_k^v$ with $k \in \{1, \ldots, m\}$; (ii) $v(S) > v(T)$ for all $S, T \in 2^N$ with $S \in \Sigma_h^v, T \in \Sigma_k^v$ and $h, k \in \{1, \ldots, m\}$ such that $h < k$. Since the coalitions in the partition's elements $\Sigma_1^v, \ldots, \Sigma_m^v$ are arranged in descending order according to v we will also write $\Sigma_1^v > \ldots > \Sigma_m^v$. We denote by i_k^v the number of sets in Σ_k^v that contain the element i, with $k = 1, \ldots, m$. Let $\boldsymbol{\theta}^v(i)$ be the m-dimensional vector $\boldsymbol{\theta}^v(i) = (i_1^v, \ldots, i_m^v)$ associated with v. The *lexicgraphic excellence (lex-cel)* [4] is the map $R_{le} : \mathcal{E}^N \to \mathcal{R}(N)$ such that

$$i \, R_{le}^v \, j \Leftrightarrow \boldsymbol{\theta}^v(i) \; \geq_L \; \boldsymbol{\theta}^v(j)$$

for any $v \in \mathcal{E}^N$ and $i, j \in N$, where $\geq_L$ is the lexicographic order among vectors. Consider, for instance, the following e.f. $v \in \mathcal{E}^N$ similar to the one in Example 11 of [7], with $N = \{1, 2, 3\}$ and such that $v(1, 2, 3) = v(1, 2) = v(1, 3) = 10, v(2, 3) = v(2) = v(3) = 7, v(1) = v(\emptyset) = 0$. The evaluation function v can be seen as the result of an ML model applied to a dataset where features 2 and 3 are strongly correlated. The redundancy between 2 and 3, together with the criticality of 1 in reaching the best prediction with at least one of the other two features, advocates in favor of the decision of ranking feature 1 above features 2 or 3. In fact, the only way to obtain the best evaluation 10 with a smallest set of features is via the selection of either 1 and 2 together, or 1 and 3 together. Moreover, considering that 2 and 3 cannot be distinguished in a ranking because of their symmetry, the unique ranking of features that put forward one of these two best sets of features, is the one where 1 is the top feature. This ranking is precisely the one provided by the lex-cel, according to the lexicographic comparison of vectors $\theta^v(1) = (3, 0, 1)$, $\theta^v(2) = \theta^v(3) = (2, 2, 0)$. One can check that the Shapley value of v yields the opposite ranking where 2 or 3 have the highest rank.

We study four properties for ranking solutions:

- *Symmetry* [16]: if two features i and j are "perfect substitutes", in the sense that for all coalitions $S \in 2^{N \setminus \{i,j\}}$ we have $v(S \cup \{i\}) = v(S \cup \{j\})$, then i and j should be ranked indifferent.
- *Strict desirability* [10, 15]: if feature i performs systematically not worse than a feature j, in the sense that for all coalitions $S \in 2^{N \setminus \{i,j\}}$ we have $v(S \cup \{i\}) \geq v(S \cup \{j\})$, and there exists $T \in 2^{N \setminus \{i,j\}}$ with $v(T \cup \{i\}) > v(T \cup \{j\})$, *i.e.* the features in the submodel on T together with i have a strictly better prediction performance than the submodel on T together with j, then i should be ranked strictly better than j.
- *Coalitional anonymity* [4]: when comparing two features i and j, a ranking solution must focus on the (ordinal) position of coalitions containing only one of them (and not those containing both i and j, or neither of them); moreover, as the number of selected features is not imposed *a priori*, a solution should not pay attention to the size of coalitions containing those features.
- *Independence from the worst set* [4]: suppose that, according to a ranking solution, a feature i is declared strictly better than j on an e.f. $v \in \mathcal{E}^N$. Take a another e.f. $v' \in \mathcal{E}^N$ obtained from v by partitioning the elements of the worst class Σ_m^v. Then i should be declared strictly better than j also in v'.

Based on an analogous result in the ordinal framework of social ranking studied in [1], we have shown that R_{le} is the unique ranking solution fulfilling properties (1), (2), (3) and (4).

3 COMPUTATIONAL EXPERIMENTS

To analyze the performance of LeXAI (the feature selection method based on lex-cel) against SHAP [14], we computed the average error of both methods to select the k most relevant features for each possible k as follows. First, for each point in a dataset, we computed an order on the features according to both approaches. Then, given a number k, we evaluated the performance of each approach by averaging the $F1$-score or the mean square error, depending if the problem is either a classification or regression task, respectively. This error was averaged on the prediction of 1000 perturbations of every point in the dataset where the k most important features according to each approach remained unchanged. We observed that LeXAI outperforms SHAP in the vast majority of the experiments we carried out, and in particular for regressions tasks. We also noticed that LeXAI is at least as good as SHAP for small number of k features; however, for lqrger k, LeXAI consistently outperforms SHAP in terms of prediction performance.

In terms of computation time, LeXAI usually performs significantly faster than SHAP for relatively small k (< 15). The time performances of LeXAI, however, seem to decrease significantly when the number of features is larger. This is probably due to the fact that our approach enumerates all the coalitions, while the Kernel-SHAP method estimates features' attributions based on a limited sample of coalitions [11]. Therefore, we adopted a heuristic approach approximating the LeXAI that only requires considering coalitions of size $|N| - 1$, with the objective to approximate the LeXAI linearly with the number of individuals. Furthermore, the approximation (referred to as *LeXAI approx*) assigns to each feature $i \in N$ and every e.f. $v \in \mathcal{E}^N$ the value $\bar{M}_i(v) = M_i(v) + \frac{1}{|N|}(v(N) - \sum_{i \in N} M_i(v))$, where where $M_i(v) = v(N) - v(N \setminus \{i\})$ is the *marginal index* [9]. This value can be interpreted as a numerical representation of features' importance, similar to the importance evaluation provided by the Shapley value. Although the LeXAI approx does not achieve the same level of performance as LeXAI, we observed it still outperforms SHAP. We also proved that LeXAI approx coincides with the exact LeXAI when coalitions of size $|N| - 1$ have distinct values, and the evaluation function v is monotonic w.r.t. set inclusion.

4 FUTURE WORK

In this paper we have proposed and studied the LeXAI method for feature selection and we have compared it with the SHAP method. We have shown that LeXAI outperforms SHAP both on an axiomatic basis and from an experimental perspective. Looking for a compelling numerical representation of a ranking provided by the lex-cel which could help to measure the explanation power of each feature, is an interesting direction for future research, that we are currently exploring.

We are also exploring more sophisticated approximation strategies, such as iterative procedures that elicit the lex-cel ranking taking advantage, at each step of the iteration, of the partial ranking elicited at previous steps (so, filtering out uninformative coalitions at each step).

ACKNOWLEDGMENTS

Financial support from the ANR project THEMIS (ANR-20-CE23-0018) is gratefully acknowledged.

REFERENCES

[1] Michele Aleandri, Felix Fritz, and Stefano Moretti. 2024. Desirability and social rankings. arXiv:2404.18755 [cs.GT] https://arxiv.org/abs/2404.18755

[2] Encarnación Algaba, Stefano Moretti, Eric Rémila, and Philippe Solal. 2021. Lexicographic solutions for coalitional rankings. *Social Choice and Welfare* 57, 4 (2021), 817–849. https://doi.org/10.1007/s00355-021-01340-z

[3] Sylvain Béal, Eric Rémila, and Philippe Solal. 2022. Lexicographic solutions for coalitional rankings based on individual and collective performances. *Journal of Mathematical Economics* 102 (2022), 102738.

[4] Giulia Bernardi, Roberto Lucchetti, and Stefano Moretti. 2019. Ranking objects from a preference relation over their subsets. *Social Choice and Welfare* 52, 4 (2019), 589–606. https://doi.org/10.1007/s00355-018-1161-1

[5] Thomas W Campbell, Heinrich Roder, Robert W Georgantas III, and Joanna Roder. 2022. Exact Shapley values for local and model-true explanations of decision tree ensembles. *Machine Learning with Applications* 9 (2022), 100345.

[6] Ian Covert, Scott Lundberg, and Su-In Lee. 2021. Explaining by removing: A unified framework for model explanation. *Journal of Machine Learning Research* 22, 209 (2021), 1–90.

[7] Daniel Fryer, Inga Strümke, and Hien Nguyen. 2021. Shapley values for feature selection: The good, the bad, and the axioms. *Ieee Access* 9 (2021), 144352–144360.

[8] Adrian Haret, Hossein Khani, Stefano Moretti, and Meltem Öztürk. 2018. Ceteris paribus majority for social ranking. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, Jérôme Lang (Ed.). ijcai.org, California, USA, 303–309.

[9] Yan-An Hwang and Yu-Hsien Liao. 2010. Consistency and dynamic approach of indexes. *Social Choice and Welfare* 34, 4 (2010), 679–694.

[10] John R Isbell. 1958. A class of simple games. *Duke Mathematical Journal* 25 (1958), 423–439.

[11] Gwladys Kelodjou, Laurence Rozé, Véronique Masson, Luis Galárraga, Romaric Gaudel, Maurice Tchuenté, and Alexandre Termier. 2024. Shaping Up SHAP: Enhancing Stability through Layer-Wise Neighbor Selection. In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2014, February 20-27, 2024, Vancouver, Canada*, Michael J. Wooldridge, Jennifer G. Dy, and Sriraam Natarajan (Eds.). AAAI Press, Washington, DC, 13094–13103.

[12] I. Elizabeth Kumar, Suresh Venkatasubramanian, Carlos Scheidegger, and Sorelle A. Friedler. 2020. Problems with Shapley-value-based explanations as feature importance measures. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event (Proceedings of Machine Learning Research, Vol. 119)*. PMLR, Virtual, 5491–5500.

[13] Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. 2020. From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence* 2, 1 (2020), 56–67.

[14] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30 (2017), 4765–4774.

[15] Michael Maschler and Bezalel Peleg. 1966. A characterization, existence proof and dimension bounds for the kernel of a game. *pacific Journal of Mathematics* 18, 2 (1966), 289–328.

[16] Lloyd S. Shapley. 1953. A value for n-person games. In *Contributions to the Theory of Games II*, Kuhn H. and Tucker A.W. (Eds.). Princeton University Press, Princeton, New Jersey, USA, 307–317.

[17] Takahiro Suzuki and Masahide Horita. 2024. Consistent social ranking solutions. *Social Choice and Welfare* 62 (2024), 549–569.