A Characterization of Choice Functions Representable by Pareto-Embedding of Alternatives

Karlson Pfannschmidt[1] and **Eyke Hüllermeier**[2]

Abstract. We address the problem of learning choice functions, that is, functions that map sets of objects (each of which is characterized in terms of a feature vector) to subsets of favored objects. To this end, we leverage the recently proposed notion of Pareto-embedding. The idea is to map the objects into a multi-dimensional utility space in such a way that the chosen subset corresponds to the Pareto set in that space. We deepen the theoretical understanding of the approach by showing the characteristic property of choice functions that can be represented in this way, and prove the suitability of the proposed surrogate losses.

1 Introduction

Beginning with the inception of the social sciences, an important endeavor has been the study of how humans express preferences and make decisions [1]. *Choice functions* were established by Arrow [2] in his seminal monograph, which shaped subsequent research in economic theory. In the choice problem, sets of objects are presented and preferences are expressed by choosing corresponding subsets. Contrary to the classical approaches from the literature, we view this problem in the context of machine learning. Since choices are a way to indicate preferences, this problem is part of the field of *preference learning* [3], which most notably encompasses learning to rank and recommender systems. A choice function maps sets of objects to subsets thereof, and we aim to infer such functions from data. To this end, we consider a learner that observes data in the form of subset choices from candidate sets, from which it induces a predictive model in the form of a choice function.

This problem has received some recent attention with work by Benson, Kumar, and Tomkins [4], who present a generalization of a multinomial logit model which in addition allows correction terms for some subsets of objects; the authors show, however, that selecting the optimal set is NP-hard. Pfannschmidt et al. [5] propose to generalize the utility function instead to include more of the set-context, with the downside that they have to tune a separate threshold with which to compute choice sets.

In this paper, we study an interesting alternative, which is based on the extension of a *one-dimensional* (real-valued) utility function to a *multi-dimensional* ($\mathbb{R}^d$-valued) utility function [6, 7]. A function of this kind can be seen as an embedding of objects in a higher-dimensional utility space, in which choice sets are identified quite naturally with Pareto-optimal points. In contrast to the one-dimensional case, a Pareto-embedding does not induce a total order on the objects. Instead, the order is only partial, and the sets of Pareto-optimal points can be of any size. See Fig. 1 for a visualization of a two-dimensional Pareto-embedding.

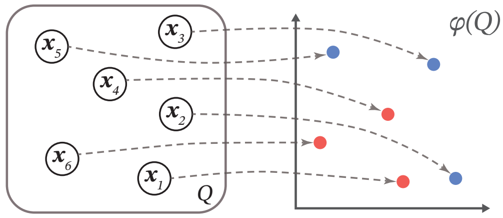

Figure 1. A Pareto-embedding $\varphi(\cdot)$ maps a given set of objects Q into a higher-dimensional utility space $\mathcal{Z}$. The Pareto-optimal points in this space (here depicted in blue) we define to be the choice set C. Figure replicated from original paper [6].

This idea was first sketched in a short paper by Pfannschmidt and Hüllermeier [6]. We extend on the work by making the following contributions: We identify and prove the characteristic property of choice functions that can be represented by Pareto-embeddings. We also show the suitability of the (differentiable) surrogate loss function for learning an embedding (Section 2).

2 Theoretical Results

We assume that a reference set of objects (or choice alternatives) $\mathcal{X} \subseteq \mathbb{R}^d$ exists, from which subsets are drawn. For ease of exposition, let this reference set be finite. Each object $\mathbf{x} \in \mathcal{X}$ is a vector $(x_1, \dots, x_d)$ comprised of real-valued feature values. A subset $Q = \{\mathbf{x}_1, \dots, \mathbf{x}_m\} \subseteq \mathcal{X}$ of arbitrary but finite size m is called a *choice task*. Preferences are expressed in the form of a *choice set* $C \subseteq Q$. We call a function $c\colon 2^{\mathcal{X}} \to 2^{\mathcal{X}}$ a *choice function* if $c(Q) \subseteq Q$ for all $Q \subseteq \mathcal{X}$ and $|c(Q)| \geq 1$, i. e., choice sets are always subsets and cannot be empty.

We illustrate the main idea of Pareto-embeddings in Fig. 1. A set of objects $Q \subseteq \mathcal{X}$ is shown on the left side. The function $\varphi\colon \mathcal{X} \to \mathcal{Z}$ maps each object into the utility space $\mathcal{Z} \subseteq \mathbb{R}^{d'}$. This mapping is done such that the choice set $C \subseteq Q$ forms the Pareto-set in this d'-dimensional utility space, i. e., the set of points that are not dominated by any other point. Such a mapping is subsequently called a *Pareto-embedding*.

Therefore, if $Q = \{\mathbf{x}_1, \dots, \mathbf{x}_m\} \subseteq \mathcal{X}$ is a choice task, then let $Z = \varphi(Q) \coloneqq \{\varphi(\mathbf{x}_1), \dots, \varphi(\mathbf{x}_m)\} = \{\mathbf{z}_1, \dots, \mathbf{z}_m\}$ be its embedding. In the embedding space, a vector $\mathbf{z}_i$ is *dominated* by another vector $\mathbf{z}_j$, if $z_{i,k} \leq z_{j,k}$ for all $k \in [d']$, and if $z_{i,k} < z_{j,k}$ for at least one k. The usual notation for this relation will be $\mathbf{x}_i \succ_\varphi \mathbf{x}_j$. If it holds that a vector $\mathbf{z}_i$ is not dominated by any other point, i. e.,,

[1] Department of Computer Science, Paderborn University, email: kiudee@mail.upb.de
[2] Institute of Informatics, LMU Munich, email: eyke@lmu.de

$\mathbf{z}_j \in Z$, $1 \leq j \neq i \leq m$, then we call such a vector *Pareto-optimal*. With this we can introduce the set of Pareto-optimal points $\mathrm{Par}_\varphi(Q) \subseteq Q$ of Q, which contains all points in Q such that their embedding is Pareto-optimal.

The Pareto-embeddings, as introduced above, define a subset of the set of all possible choice functions. It is natural to ask if Pareto-embeddings are able to express any choice function, or if their expressive power is restricted. It is straightforward to show, that there are choice functions which cannot be expressed by a Pareto-embedding, regardless of the dimension d'.

Proposition 2.1 *Let $\mathcal{X}$ be a set of objects with $|\mathcal{X}| \geq 3$. There exists a choice function $c\colon 2^{\mathcal{X}} \to 2^{\mathcal{X}}$ that cannot be represented by a Pareto-embedding.* □

Having established that the set of Pareto-embeddings is a strict subset of all possible choice functions, we may wonder what the characteristic property of this class of choice function is, *without* having to invoke the specific vector-based representation of Pareto-embeddings. We begin by stating a few useful definitions. Let $c\colon 2^{\mathcal{X}} \to 2^{\mathcal{X}}$ be a choice function. We will say that $x \in \mathcal{X}$ *is preferred to* $y \in \mathcal{X}$ by a choice function c (denoted $x \succ_c y$), if and only if $\exists Q \in 2^{\mathcal{X}}$ with $y \in Q, x \notin Q, y \in c(Q), y \notin c(Q \cup \{x\})$ and $x \in c(Q \cup \{x\})$. In words, if the addition of an object x causes another object y to be no longer in the set of chosen objects $c(Q \cup \{x\})$, then $x \succ_c y$. Note, that for arbitrary choice functions c, it can happen that both $x \succ_c y$ and $y \succ_c x$ hold.

With that we will define the key property used in our main result: A choice function $c\colon 2^{\mathcal{X}} \to 2^{\mathcal{X}}$ satisfies the *Pareto-property*, if and only if: (i) $\forall x, y \in \mathcal{X}$: If $x \succ_c y$, then $\forall Q \in 2^{\mathcal{X}}$ with $Q \supseteq \{x, y\}$: $y \notin c(Q)$, and (ii) $\succ_c$ is transitive. The first property (i) requires that a preference $x \succ_c y$ needs to hold for all $Q \in 2^{\mathcal{X}}$, making it impossible that $x \succ_c y$ and $y \succ_c x$ hold at the same time. Since cycles of preferences would still be possible, e. g., $x \succ_c y$, $y \succ_c z$ and $z \succ_c x$, we also have to require that transitivity holds for $\succ_c$. We will now see that the set of choice functions having the Pareto-property is exactly equal to the set of choice-functions which can be represented by Pareto-embeddings. Remember that the dominance relation also defines an order on objects.

In our main result, we show the equivalence of the Pareto-property and the existence of an embedding:

Theorem 2.2 *Let $\mathcal{X}$ be finite set. A choice function $c\colon 2^{\mathcal{X}} \to 2^{\mathcal{X}}$ satisfies the Pareto-property if and only if there exists an embedding $\varphi\colon \mathcal{X} \to \mathbb{R}^{d'}$, such that $c(Q) = \mathrm{Par}_\varphi(Q)$ for all $Q \in 2^{\mathcal{X}}$.* □

The main idea of the proof is that we can reduce the Pareto-embedding problem to the general embedding problem of a partial order. This allows us to leverage a few interesting theoretical results from the literature: It is known that, for partial orders of dimension 2, there exists an efficient algorithm to compute the corresponding embedding [8, 9]. For higher-dimensional partial orders, it becomes NP-complete to decide that the dimension is at most k [10]. Characterizing the properties of higher-dimensional partial orders is a recent endeavor and could lead to new insights [11].

Another question concerns the suitability of the surrogate loss function proposed by Pfannschmidt and Hüllermeier [6] for the task of learning Pareto-embeddings. The following theorem establishes that if the surrogate loss function is 0, we can be sure that our embedding is a valid Pareto-embedding.

Theorem 2.3 *Let $c\colon 2^{\mathcal{X}} \to 2^{\mathcal{X}}$ be a choice function with the Pareto-property, and let its dimension be d'. If $\varphi\colon \mathcal{X} \to \mathbb{R}^{d'}$ achieves a loss $L(\varphi(Q), c(Q)) = 0$ for all $Q \subseteq \mathcal{X}$, then φ is a Pareto-embedding of c.* □

We have already seen that Pareto-embeddings are intimately connected to a certain pairwise preference relations $\succ_c$ defined on the objects. It stands to reason that a Pareto-embedding could be constructed solely by considering choices on pairs of objects, which, as shown by the following proposition, is indeed the case.

Proposition 2.4 *Assume for a given $\varphi\colon \mathcal{X} \to \mathbb{R}^{d'}$ and c a choice function with the Pareto-property, that $L(\varphi(Q), c(Q)) = 0$ holds for any subsets $Q \subseteq \mathcal{X}$ with $|Q| = 2$. Then it holds that $L(\varphi(Q), c(Q)) = 0$ for all subsets $Q \subseteq \mathcal{X}$.* □

3 Conclusion

We elaborated on a recent approach to view the problem of learning choice functions as an embedding problem. We advance the theory, by showing that choice functions resulting from Pareto-embeddings are equivalent to a family of choice functions consistent with a certain partial order relation.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the European Regional Development Fund (ERDF) and the valuable feedback provided by the industry partners of the Smart-GM research project.

References

[1] C. Domshlak et al., “Preferences in AI: An overview”, *Artificial Intelligence* **175**.7-8 (2011), pp. 1037–1052.

[2] K. J. Arrow, *Social Choice and Individual Values*, John Wiley & Sons, 1951.

[3] J. Fürnkranz and E. Hüllermeier, eds., *Preference Learning*, Springer, 2010.

[4] A. R. Benson, R. Kumar, and A. Tomkins, “A Discrete Choice Model for Subset Selection”, in: *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, ed. by Y. Chang et al., ACM, 2018, pp. 37–45.

[5] K. Pfannschmidt et al., “Learning Context-Dependent Choice Functions”, *International Journal of Approximate Reasoning* **140** (2022), pp. 116–155.

[6] K. Pfannschmidt and E. Hüllermeier, “Learning Choice Functions via Pareto-Embeddings”, in: *KI 2020: Advances in Artificial Intelligence*, ed. by U. Schmid, F. Klügl, and D. Wolter, Cham: Springer International Publishing, 2020, pp. 327–333.

[7] A. Benavoli, D. Azzimonti, and D. Piga, *Choice functions based multi-objective Bayesian optimisation*, version 1, Oct. 15, 2021, arXiv: 2110.08217 [stat.ML].

[8] B. Dushnik and E. W. Miller, “Partially Ordered Sets”, *American Journal of Mathematics* **63**.3 (1941), pp. 600–610.

[9] A. Pnueli, A. Lempel, and S. Even, “Transitive Orientation of Graphs and Identification of Permutation Graphs”, *Canadian Journal of Mathematics* **23**.1 (1971), pp. 160–175.

[10] M. Yannakakis, “The Complexity of the Partial Order Dimension Problem”, *SIAM Journal on Algebraic Discrete Methods* **3**.3 (Sept. 1982), pp. 351–358.

[11] S. Qi, “A Characterization of the N-Agent Pareto Dominance Relation”, *Social Choice and Welfare* **46**.3 (Mar. 1, 2016), pp. 695–706.