

# Polar Encoding: A Simple Baseline Approach for Classification with Missing Values (abstract)\*

Oliver Urs Lenz<sup>1,2</sup>, Daniel Peralta<sup>3</sup>, and Chris Cornelis<sup>2</sup>

<sup>1</sup> Leiden Institute of Advanced Computer Science, Leiden University  
o.u.lenz@liacs.leidenuniv.nl

<sup>2</sup> Research Group for Computational Web Intelligence,  
Department of Applied Mathematics, Computer Science and Statistics,  
Ghent University

chris.cornelis@ugent.be <https://cwi.ugent.be>

<sup>3</sup> Department of Information Technology, Ghent University  
daniel.peralta@ugent.be

## 1 Introduction

Missing-indicators generally increase classification performance on real-life datasets with missing values, relative to performing just imputation [3]. However, missing-indicators can only be used in addition to, not instead of imputation. To address this, we introduce a new baseline approach towards missing values called *polar encoding*, which can be used with categorical and  $[0, 1]$ -scaled numerical attributes. Polar encoding represents missing values numerically, in a neutral way, leaving it up to the classification algorithm how to learn from them. It is similar to, and inspired by, the numerical representation of categorical attributes through one-hot encoding.

For categorical attributes, polar encoding corresponds exactly to one-hot encoding, while it converts  $[0, 1]$ -valued attributes into pairs of features with:

$$x \mapsto \langle x, 1 - x \rangle;$$

or, when the the classification algorithm is based on Euclidean distance, with:

$$x \mapsto \left\langle \sin \frac{x \cdot \pi}{2}, \cos \frac{x \cdot \pi}{2} \right\rangle.$$

Missing values are represented as zero vectors for all attribute types.

## 2 Practical advantages

We show that polar encoding satisfies the following four ideals of a good baseline approach towards missing values in the context of classification:

**Modularity.** Polar encoding is self-contained. It results in a complete, numerical dataset that can be used with any classification algorithm. This allows classification algorithms to be agnostic about missing values.

---

\* This is an extended abstract of [4].

**Conservatism.** Polar encoding is a faithful representation of the original dataset. It presupposes as little as possible about how missing values contribute to the learning task. With polar encoding, missing values become equidistant from all non-missing values — the distance is always 1. It achieves this by mapping non-missing values onto the non-negative quadrant of the unit circle, and missing values to the origin. For decision tree algorithms, the two dimensions of a polar-encoded numerical attribute effectively offer a choice as to which side of each split missing values should be grouped with. This is very similar to the *missingness incorporated in attributes* (MIA) proposal [5].

**Simplicity.** Polar encoding is a simple transformation of the data. It can be applied quickly and easily, without the need for any dedicated software. It only requires a minimal amount of computational effort and no parameter choices by the user.

**Performance.** Polar encoding enables good downstream prediction performance. We find that polar encoding generally outperforms the sophisticated imputation strategies MICE [1] and MIDAS [2]. It also performs better than mean/mode imputation with missing indicators, although this difference is less pronounced, and mean/mode imputation may have a slight advantage with Ada-boosted trees when hyperparameter optimisation is applied.

Other approaches towards missing values satisfy these ideals to a lesser degree. This does not mean that they should not be used, only that they are less suited as a baseline approach. Imputation satisfies modularity, but by design it is not conservative. Most imputation algorithms are not simple either, requiring substantial amounts of computation and user input. The missing-indicator approach is not fully conservative, because it still induces classification algorithms to treat missing values like their imputed values. Finally, MIA is conservative, but not modular, since it requires an adaptation of the prediction algorithm itself and only applies to decision tree algorithms.

### 3 Theoretical perspective

Polar encoding can be seen as a representation of *barycentric* attributes, which generalise both categorical and  $[0, 1]$ -valued attributes. It then becomes clear that:

- Polar encoding and one-hot encoding can be identified with, respectively, fuzzy and crisp partitions of a dataset. Thus, polar encoding constitutes a fuzzification of one-hot encoding.
- Polar encoding of  $[0, 1]$ -valued attributes is analogous to encoding binary attributes as two one-hot encoded categorical values  $\langle 1, 0 \rangle$  and  $\langle 0, 1 \rangle$  rather than as simple numerical values 1 and 0.
- This representation is slightly redundant, but it is precisely this redundancy that allows us to represent missing values as zero vectors.

## References

1. Van Buuren, S., Oudshoorn, K.: Flexible multivariate imputation by MICE. Tech. Rep. PG/VGZ/99.054, TNO Prevention and Health, Leiden (1999)
2. Lall, R., Robinson, T.: The MIDAS touch: Accurate and scalable missing-data imputation with deep learning. *Political Anal.* **30**(2), 179–196 (2022)
3. Lenz, O.U., Peralta, D., Cornelis, C.: No imputation without representation. In: *BNAIC/BeNeLearn 2023: Revised Selected Papers from the 35th Joint Benelux Conference on Artificial Intelligence and Machine Learning*. Communications in Computer and Information Science (2024)
4. Lenz, O.U., Peralta, D., Cornelis, C.: Polar encoding: A simple baseline approach for classification with missing values. *IEEE Transactions on Fuzzy Systems* **32**(5), 3084–3093 (2024)
5. Twala, B.E., Jones, M., Hand, D.J.: Good methods for coping with missing data in decision trees. *Pattern Recognition Letters* **29**(7), 950–956 (2008)