

Strong Robust Similarities: A new family of similarity measures

Pedro Huidobro¹[0000-0002-0170-0426], Noelia Rico²[0000-0002-6169-4523], Sara S. Fernández¹[0009-0003-1397-2036], and Agustina Bouchet¹[0000-0003-1516-8976]

¹ Department of Statistics. University of Oviedo, 33003 Oviedo, Spain
`{huidobropedro,uo264389,bouchetagustina}@uniovi.es`

² Department of Computer Science. University of Oviedo, 33003 Oviedo, Spain
`noeliarico@uniovi.es`

Abstract. Similarity measures are a constant research topic among researchers as they provide a very useful tool for quantifying the degree of similarity between two objects. The most well-known and widely used definition of similarity involves three axioms, the symmetry of the operator, the monotonicity concerning interval inclusion, and the operator taking the value of 1 if and only if the intervals are equal ([1], [3]).

When dealing with interval data, which represents uncertain or imprecise information, the conventional similarity measures may not suffice. To address this challenge, robust similarity measures are necessary, which can handle the uncertainty inherent in interval data. Kabir et.al. [2] proposed the definition of robust similarities by adding five additional axioms. The most important axiom has to do with the propensity for aliasing. When two pairs of intervals that have different widths but the same union and intersection widths, you can see aliasing.

Our proposal centers on characterizing similarity through pre-inclusion functions, which forms the basis for defining a family of similarity measures. Well-known measures like Jaccard and Dice similarities arise as particular cases within this framework. Furthermore, we introduce a new family of similarities, called strong robust similarity measures, by adding three axioms to the robust similarity measures, and characterize them based on pre-inclusions. We conduct a comprehensive analysis of our proposed similarity measure, examining its behavior across various scenarios and comparing it with established measures. Additionally, we demonstrate its applicability in real-world contexts by employing weather data obtained from AEMET (the National Weather Service in Spain) as an illustrative example. By applying our measure to analyze this data, we showcase its effectiveness in capturing similarities between interval-based datasets, highlighting its potential for practical applications.

Keywords: Similarity · Intervals · Robust similarity measures.

Acknowledgments. Authors would like to express their gratitude to the Spanish Ministry of Science and Innovation (fund reference: PREP2022-000355. Project: MCINN-23-PID2022-139886NB-I00) and to the European Project UE-23-AI4RA-101132914.

References

1. S. Kabir, C. Wagner, T. Havens, D. Anderson, and U. Aickelin. Novel similarity measure for interval-valued data based on overlapping ratio. In 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 07 2017.
2. S. Kabir, C. Wagner, T. C. Havens, and D. T. Anderson. A similarity measure based on bidirectional subsethood for intervals. *IEEE Transactions on Fuzzy Systems*, 28(11):2890–2904, 2020.
3. S. Ontañón. An overview of distance and similarity functions for structured data. *Artificial Intelligence Review*, 53:5309–5351, 2020.