

Reducts in Rough Sets: Algorithmic Insights, Open Source Libraries and Applications (Tutorial – Extended Abstract)

Andrzej Janusz*[†], Sebastian Stawicki[†]

*Faculty of Mathematics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland

Email: ajanusz@mimuw.edu.pl

[†]QED Software, Mazowiecka 11/49, 00-052 Warsaw, Poland

Email: sebastian.stawicki@qed.pl

THE theory of rough sets [1] is a powerful mathematical framework for handling imprecise or uncertain information in data analysis and decision-making. At its core, rough set theory introduces the concept of decision reducts [2], which are subsets of attributes or features that preserve the essential information needed to make accurate decisions while eliminating redundant or irrelevant information. By identifying ensembles of decision reducts [3], analysts can simplify complex datasets, improve classification accuracy, and gain valuable insights from noisy or incomplete data [4]. These appealing characteristics make rough sets a valuable tool in various fields, including machine learning, data mining, and expert systems [5].

There have been proposed many extensions to the notion of decision reducts, such as approximate decision reducts [6], dynamic decision reducts [7], DAARs [8], decision bireducts [9], and many others. The key objective of most of them was to prevent the inclusion of illusionary dependencies between attributes and decision values to the reducts. A lot of research was also committed to the problem of algorithms for the efficient computation of diverse reduct sets [10]. This topic is particularly important from the perspective of practical applications of the rough set theory [11].

In this tutorial, we focus on the latter aspect of the decision reduct-related research. We discuss various, both, well-known and relatively new algorithms, and consider their specific advantages. We explain in detail selected implementation aspects that are crucial for the efficient computation of many types of decision reducts. We also overview and demonstrate libraries in popular programming languages that allow easy computation of reducts on real-world datasets, including *RoughSets* library for R [12] and a novel Python language library *scikit-rough*¹ [11]. Finally, we share the results of a study aiming at the comparison of the computational efficiency of various reduct algorithms.

REFERENCES

[1] Z. Pawlak and A. Skowron, “Rudiments of rough sets,” *Institute of Mathematics, Warsaw University, Banacha 2, 02-097 Warsaw, Poland*,

¹<https://github.com/sebov/scikit-rough>

- vol. Information Sciences 177 (2007), pp. 3–27, 2006.
- [2] M. Grzegorowski, “Governance of the redundancy in the feature selection based on rough sets’ reducts,” in *Rough Sets*, V. Flores, F. Gomide, A. Janusz, C. Meneses, D. Miao, G. Peters, D. Ślęzak, G. Wang, R. Weber, and Y. Yao, Eds. Cham: Springer International Publishing, 2016, pp. 548–557.
- [3] J.-H. Zhai, X.-Z. Wang, and H.-C. Wang, “Dynamic ensemble of rough set reducts for data classification,” in *Rough Sets and Knowledge Technology: 9th International Conference, RSKT 2014, Shanghai, China, October 24–26, 2014, Proceedings*. Berlin, Heidelberg: Springer-Verlag, 2023, p. 642–649. [Online]. Available: https://doi.org/10.1007/978-3-319-11740-9_59
- [4] A. Janusz and S. Stawicki, “Applications of approximate reducts to the feature selection problem,” in *Rough Sets and Knowledge Technology - 6th International Conference, RSKT 2011, Banff, Canada, October 9–12, 2011, Proceedings*, ser. Lecture Notes in Computer Science, J. Yao, S. Ramanna, G. Wang, and Z. Suraj, Eds., vol. 6954. Springer, 2011, pp. 45–50. [Online]. Available: https://doi.org/10.1007/978-3-642-24425-4_8
- [5] A. Skowron and S. Dutta, “Rough sets: Past, present, and future,” *Natural Computing: An International Journal*, vol. 17, no. 4, p. 855–876, dec 2018. [Online]. Available: <https://doi.org/10.1007/s11047-018-9700-3>
- [6] H. S. Nguyen and D. Ślęzak, “Approximate reducts and association rules,” in *New Directions in Rough Sets, Data Mining, and Granular-Soft Computing*, N. Zhong, A. Skowron, and S. Ohsuga, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 137–145.
- [7] J. G. Bazan, A. Skowron, and P. Synak, “Dynamic reducts as a tool for extracting laws from decisions tables,” in *Methodologies for Intelligent Systems*, Z. W. Raś and M. Zemankova, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994, pp. 346–355.
- [8] A. Janusz and D. Ślęzak, “Computation of approximate reducts with dynamically adjusted approximation threshold,” in *Foundations of Intelligent Systems - 22nd International Symposium, ISMIS 2015, Lyon, France, October 21–23, 2015, Proceedings*, ser. Lecture Notes in Computer Science, F. Esposito, O. Pivert, M. Hacid, Z. W. Ras, and S. Ferilli, Eds., vol. 9384. Springer, 2015, pp. 19–28. [Online]. Available: https://doi.org/10.1007/978-3-319-25252-0_3
- [9] S. Stawicki, D. Ślęzak, A. Janusz, and S. Widz, “Decision bireducts and decision reducts - a comparison,” *Int. J. Approx. Reason.*, vol. 84, pp. 75–109, 2017. [Online]. Available: <https://doi.org/10.1016/j.ijar.2017.02.007>
- [10] A. Janusz and D. Ślęzak, “Rough set methods for attribute clustering and selection,” *Appl. Artif. Intell.*, vol. 28, no. 3, pp. 220–242, 2014. [Online]. Available: <https://doi.org/10.1080/08839514.2014.883902>
- [11] A. Janusz, D. Ślęzak, S. Stawicki, and K. Stencel, “A practical study of methods for deriving insightful attribute importance rankings using decision bireducts,” *Inf. Sci.*, vol. 645, p. 119354, 2023. [Online]. Available: <https://doi.org/10.1016/j.ins.2023.119354>
- [12] L. S. Riza, A. Janusz, C. Bergmeir, C. Cornelis, F. Herrera, D. Ślęzak, and J. M. Benítez, “Implementing algorithms of rough set theory and fuzzy rough set theory in the R package ‘roughsets,’” *Inf. Sci.*, vol. 287, pp. 68–89, 2014. [Online]. Available: <https://doi.org/10.1016/j.ins.2014.07.029>