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 illusory 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† [11]. Finally, we share the results of a study aiming at the comparison of the computational efficiency of various reduct algorithms.

REFERENCES


Tutorials

Andrzej Janusz†, Sebastian Stawicki†

†Faculty of Mathematics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland
Email: ajanusz@mimuw.edu.pl

†QED Software, Mazowiecka 114/9, 00-052 Warsaw, Poland
Email: sebastian.stawicki@qed.pl

IEEE Catalog Number: CFP2385N-ART ©2023, PTI