Proceedings of the 18th Conference on Computer Science and Intelligence Systems pp. 69–70
DOI: 10.15439/2023F0003
ISSN 2300-5963 ACSIS, Vol. 35

Rough Sets: An Introductory Overview Tutorial – Extended Abstract

Soma Dutta
University of Warmia and Mazury in Olsztyn
Słoneczna 54, Olsztyn, Poland
soma.dutta@matman.uwm.edu.pl

Davide Ciucci
University of Milano-Bicocca
viale Sarca 336/14, 20126, Milano, Italy
davide.ciucci@unimib.it

***Abstract*—The aim of this tutorial is to present a brief overview of the theory of rough sets from the perspective of its mathematical foundations, history of development as well as connections with other branches of mathematics and informatics. The content concerns both the theoretical and practical aspects of applications. The above mentioned target of the tutorial will be covered in two parts. In the first part we would aim to present the introduction to rough sets and the second part will focus on the connections with other branches of mathematics and informatics.**

I. Introduction to Rough Sets

THE THEORY of rough sets, pioneered by Z. Pawlak in 1982 [1], [2], provides a way to formalize imprecise concepts with respect to a given set of attributes. Let us think of a data table, where the rows represent descriptions of the objects from a universe U with respect to a set of (conditional) attributes A, and each column of the data table corresponds to an attribute of A. Each such description is usually known as *information signature* of an object with respect to the set of attributes A, and it can be represented as a vector of values over A. Formally, one can think of a pair (U, A) where for each $a \in A$ and $u \in U$, $a(u)$ denotes the value of the object u for the attribute a in a relevant value set. In rough set literature such a pair is known as *information system* or *information table*. Moreover, an information system along with a designated attribute d for decision, say $(U, A \cup \{d\})$ is called a *decision system*.

Now, given an information system, with respect to any subset X of A the whole data table can be clustered into equivalence classes of objects where each equivalence class contains all those objects from the universe which have the same values with respect to each attribute of X. Thus from (U, A) we obtain a pair (U, R) where R is the respective equivalence relation, usually known as *indistinguishability relation*. In rough set literature (U, R) is known as *approximation space*. A subset $S \subseteq U$, may be called a concept, can be either definable in terms of the union of some equivalence classes, or may fall in the overlapping zone of a few equivalence classes; in the latter case the concept is regarded as imprecise with respect to the considered set of attributes. So, given any set of objects S one can describe the intention of S in terms of rough approximation operators. The lower approximation of S picks up those equivalence classes which are completely contained within S, and thus gives a flavour of 'certainty zone for the concept'. On the other hand, the upper approximation of S selects those equivalence classes which have non-empty intersection with S, and this corresponds to a 'possibility zone for the concept'.

With this mathematical foundation the development of rough set theory goes further to address many useful aspects of data mining. For example, suppose there are two decision classes representing positive and negative decision for the attribute d (i.e., subsets of U representing $d = 1$ and $d = 0$ respectively). Now if these decision classes are not definable in a straightforward manner, one may require to characterize them using rough set approximations. In this presentation, we will try to present some of such aspects which have practical uses in the context of data mining. A few such issues are listed below.

1) To describe a data table in terms of comprehensible rules so that using those rules unseen test examples can be effectively classified.
2) To find out a smaller set of attributes, a *reduct* in RST terminology, that can faithfully classify the decision classes as it is presented with respect to the whole set of attributes.
3) To handle a decision system where two indistinguishable objects may have different decision values.
4) To design decision valuations describing different aspects of decision making by aggregating available information of the training objects (i.e., already available objects).
5) To check similar aspects of decision making when the available data is not clustered into disjoint equivalence classes as the underlying notion of indistinguishability can be based on a relation which is weaker than an equivalence relation, such as a similarity relation.

II. Connection with other branches of mathematics and informatics

Due to its simplicity and effectiveness, the concept of approximations has found applications in various fields, establishing connections with several branches of mathematics and computer science right from its inception. Over the years, we have observed a substantial accumulation of results, which we can only provide a high-level summary of.

A. Mathematics

The main contributions relate to logic and topology, with multiple connections also existing in algebra and graph theory.

a) Logics: Rough sets are associated with modal logics. Indeed, given the standard syntax of the modal system S5, a semantics can be provided by the indiscernibility relation used as the modal accessibility relation. In this manner, the lower and upper approximations coincide with the logic operators of necessity and possibility [3]. It is evident that for rough sets based on weaker forms of relations, there corresponds weaker modal logics. Additionally, by interpreting the lower approximation as positive (or true), the complement of the upper approximation as negative (or false) and the remaining elements of the universe as unknown a correspondence with three-valued logics can be established [4]. Finally, a complete logical framework based on a distinct notion of truth, namely *rough truth*, has been defined where both syntax and semantics are "rough" [5].

b) Topology and Algebra: The lower and upper approximations behave like a topological interior and closure operator on a Boolean algebra. Several links between various types of topological operators and models of rough sets have been established [6]. Moving to a more abstract context, a hierarchy of topological operators can be defined on a lattice structure, each corresponding to a different model based on various generalizations of rough sets [7]. Many authors have taken further steps toward abstraction by defining the approximations in different types of algebraic structures, such as rings and groups.

c) Graph Theory: The connection with graph theory can be interpreted in at least two ways. Firstly, by relating ideas from rough sets to those on graph theory. One significant result in this setting is the equivalence between computing reducts and computing minimal transversal on hypergraphs. The latter is a well-known problem and algorithms that solve it in incremental polynomial time exist [8]. Another approach, consists in applying rough-set ideas to graph theory, such as attempting to define approximations or reducts on graphs.

B. Computer Science

We highlight the main contributions of rough sets to Artificial Intelligence and Theoretical Computer Science.

a) Knowledge Representation: Of course, the first link with computer science and artificial intelligence concerns the ability of rough sets to represent and handle uncertainty due to the granularity of the universe. Particularly fruitful in this domain has been the connection with other tools to manage different forms of uncertainty, mainly fuzzy sets [9] and belief functions [10].

b) Machine Learning and Data Mining: From an application stanpoint, the main contribution of Rough sets is in Machine Learning and Data Mining, where they have been used in several tasks [11]. In particular, in Machine Learning their success can be seen in feature selection and classification by means of reducts and decision rules and in clustering where the idea of approximations has been applied to obtain new soft clustering methods. In Data Mining, a major contribution is the use of rough sets to perform *approximate queries* in relational databases by means of standard SQL statements, this approach also lead to a successful industrial application [12].

c) Theoretical Computer Science: The concept of entropy has been used to evaluate the uncertainty of a given information table with an equivalence relation. Extended approaches to missing values and generalized relation has also been provided [13] as well as applications in computing approximate reducts [14]. Another connection with TCS concerns the use of rough sets in dealing with uncertainty in discrete dynamical systems, such as cellular automata, reaction systems and Petri nets [15].

REFERENCES

[1] Z. Pawlak, "Rough sets," *Int. J. Inform. Comput. Sci.*, vol. 11, pp. 341–356, 1982.

[2] ——, *Rough sets - theoretical aspects of reasoning about data*, ser. Theory and decision library : series D. Kluwer, 1991, vol. 9.

[3] E. Orlowska, "A logic of indiscernibility relations," ser. Lecture Notes in Computer Sciences. Berlin: Springer-Verlag, 1985, no. 208, pp. 177–186.

[4] D. Ciucci and D. Dubois, "Three-valued logics, uncertainty management and rough sets," *Trans. Rough Sets*, vol. 17, pp. 1–32, 2014. [Online]. Available: https://doi.org/10.1007/978-3-642-54756-0_1

[5] M. Banerjee, "Logic for rough truth," *Fundam. Informaticae*, vol. 71, no. 2-3, pp. 139–151, 2006.

[6] P. K. Singh and S. Tiwari, "Topological structures in rough set theory: A survey," *Hacettepe Journal of Mathematics and Statistics*, vol. 49, no. 4, pp. 1270 – 1294, 2020.

[7] G. Cattaneo and D. Ciucci, "Lattices with interior and closure operators and abstract approximation spaces," *Trans. Rough Sets*, vol. 10, pp. 67–116, 2009.

[8] G. Chiaselotti, D. Ciucci, T. Gentile, and F. Infusino, "Generalizations of rough set tools inspired by graph theory," *Fundam. Informaticae*, vol. 148, no. 1-2, pp. 207–227, 2016.

[9] M. Inuiguchi, W.-Z. Wu, C. Cornelis, and N. Verbiest, *Fuzzy-Rough Hybridization*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, pp. 425–451.

[10] A. Campagner, D. Ciucci, and T. Denoeux, "Belief functions and rough sets: Survey and new insights," *Int. J. Approx. Reason.*, vol. 143, pp. 192–215, 2022.

[11] R. Bello and R. Falcon, *Rough Sets in Machine Learning: A Review*. Cham: Springer International Publishing, 2017, pp. 87–118.

[12] D. Slezak, P. Synak, A. Wojna, and J. Wroblewski, "Two database related interpretations of rough approximations: Data organization and query execution," *Fundam. Informaticae*, vol. 127, no. 1-4, pp. 445–459, 2013.

[13] D. Bianucci and G. Cattaneo, "Information entropy and granulation co-entropy of partitions and coverings: A summary," *Trans. Rough Sets*, vol. 10, pp. 15–66, 2009.

[14] D. Slezak, "Approximate entropy reducts," *Fundam. Informaticae*, vol. 53, no. 3-4, pp. 365–390, 2002.

[15] A. Campagner, D. Ciucci, and V. Dorigatti, "Uncertainty representation in dynamical systems using rough set theory," *Theor. Comput. Sci.*, vol. 908, pp. 28–42, 2022.