Approaching Rough Set Theory via Categories

Federico G. Infusino

Department of Mathematics and Computer Sciences (DeMACS), University of Calabria, Italy, Calabria, Rende (CS), 87036
federico.infusino@unical.it

Extended Abstract - TACL2024

Pawlak's Rough Set Theory (briefly RST) is an elegant and powerful methodology, with applications in numerous research fields, aimed at the extraction and the optimization of the information coming from large amounts of data [9]. RST arose in the context of Pawlak's data tables with the purpose of understanding whether a given subset of objects could be partially or completely determined only on the basis of the information induced by collections of attributes.

The originary approach to RST is said *constructive*: to a suitable kind of relation $\mathcal{R}$ on the set U of the objects of a data table, one assigns a pair $(\mathrm{Lw}_{\mathcal{R}}, \mathrm{Up}_{\mathcal{R}})$ of dual set operators, respectively called $\mathcal{R}$-*lower* and $\mathcal{R}$-*upper approximants*, playing a similar role as the necessity and possibility operators in modal logic [7] and inducing a *constructive set algebra* $(\mathcal{P}(U), \cup, \cap, {}^c, \mathrm{Lw}_{\mathcal{R}}, \mathrm{Up}_{\mathcal{R}})$ on $\mathcal{P}(U)$, that often satisfies different algebraic properties [3, 6]. A further approach is said *algebraic*: one begins with a pair $(\mathbf{L}, \mathbf{H})$ of dual unary set operators defined axiomatically on a ground set U and, next, studies the resulting set algebra $(\mathcal{P}(U), \cup, \cap, {}^c, \mathbf{L}, \mathbf{H})$.

The common way of relating the previous two approaches comes from the characterization of the properties needed for defining an assignment $(\mathbf{L}, \mathbf{H}) \mapsto \mathcal{R}_{\mathbf{L},\mathbf{H}}$, where $\mathcal{R}_{\mathbf{L},\mathbf{H}}$ is a binary relation on U with $\mathbf{L}$ and $\mathbf{H}$ as lower and upper approximants [10]. Evidently, by adding suitable axioms in the definition of $\mathbf{L}$ and $\mathbf{H}$ we get additional properties on $\mathcal{R}_{\mathbf{L},\mathbf{H}}$. The attempts of combining the two approaches fit within the *representation problem*, aimed at the determination of those axiomatizations of $\mathbf{L}$ and $\mathbf{H}$ whose corresponding set algebra turns out to be the constructive set algebra induced by some specific kind of binary relation. For instance, by abstracting the axiomatic properties of $\mathrm{Lw}_{\mathcal{R}}$ and $\mathrm{Up}_{\mathcal{R}}$ when $\mathcal{R}$ is a *Pawlak's indiscernibility relation* [9], we get the so-called *lower* and *upper operators* and, in such a case, it follows that $\mathcal{R}_{\mathbf{L},\mathbf{H}}$ is an equivalence relation, yielding a *cryptomorphism* between all these structures [4].

The previous setting admits a natural categorical-theoretic interpretation - whose development might be useul to provide a unifying framework to RST - as soon as one asks questions about the functoriality of the assignments $\mathcal{R} \mapsto (\mathrm{Lw}_{\mathcal{R}}, \mathrm{Up}_{\mathcal{R}})$ and $(\mathbf{L}, \mathbf{H}) \mapsto \mathcal{R}_{\mathbf{L},\mathbf{H}}$ and of all the other constructions arising when developing the theory. To this end, the first necessity that has arisen concerns the definition of suitable categories to work with. Taking some ideas from the theory of combinatorial species, we use presheaves on the groupoid of sets to get a unifying framework in which to define collections of categories of mathematical structures with objects $\Omega_{\mathcal{X}}$, where Ω is an arbitrary set and $\mathcal{X}$ is either a specific n-ary relation, set operator, set system or data table with Ω as its attribute set.

However, in this context, the choice of the morphisms is not uniquely determined: for instance, when dealing with equivalence relations we can assume that the morphisms should preserve lower or upper approximants as required in [1, 8], or that the morphisms just preserve the equivalences as for the category **EqR**. The possibility of choosing the morphisms in completely different ways turns out to be fundamental in the attempt of making functorial various constructions on objects: as an example, some natural transformations among the presheaves that define the ambient categories within which to select the needed structures may be used to construct a non-trivial chain of categorical isomorphisms and embeddings involving suitable categories of equivalence relations, set partitions, upper and lower operators, in such a way to get a categorical counterpart for the cryptomorphism between these structures [4]. The previous result allows us to transport categorical properties from one category to another. To this end, to study the category of lower operators we made use of a particular category of set partitions, that has pullbacks and is regular, though it does not admits basic constructions as products, coproducts and coequalizers. In the same spirit, we made use of **EqR** to get informations on a category of upper operators and *continuos* functions: **EqR** becomes a specific case of *proper Moore-subcategory* of the category **Rel** of binary relations and relation-preserving maps. The investigation of proper Moore-subcategories of a given concrete category led to general results [5], holding for **EqR**: it is a reflective modification of **Rel** and inherits its limits and co-limits; it is **Set**-topological and **Set**-solid, has extremal subobject classifier but it is not regular [4].

Finally, to enrich our categorical framework for RST, after comparing possible definitions [2, 4] we introduce a category **PR** of Pawlak's data tables, obtained by dropping out the finiteness condition on its ground objects in view of possible theoretical applications from algebra and topology and assuming a compatibility condition on morphisms with interesting interpretations in applied contexts. There are at least three convincing reasons for working with **PR**. First, we proved that it is complete, balanced, exact, regular, Heyting, it admits (RegEpi, Mono-Source)-factorizations but, in general, not coproducts [4]. Secondly, being inspired by the existence of an embedding of **EqR** into **PR** that formalizes the fact that different subsets of attributes may induce the same Pawlak's indiscernibility, we can easily define convenient subcategories and functors through which to reinterpret in our categorical-theoretic setting various constructions of RST such as functional dependence or attribute reduction [4]. Third, **PR** becomes a specific instance of a further mathematical generalization, susceptible of an advanced study, by replacing sets with objects of an arbitrary category equipped with a symmetric monoidal structure.

References

[1] R. A. Borzooei, A.A. Estaji, and M. Mobini. On the category of rough sets. Soft Computing 21: 2201–2214, 2017.

[2] M. Diker. Categories of rough sets and textures. Theoretical Computer Science 488: 46–65, 2013.

[3] M. Gehrke and E. Walker. On the structure of rough sets. Bulletin of the Polish Academy of Sciences, Mathematics 40, pp. 235–255, 1992.

[4] F.G. Infusino. Categorification of a set relation geometry induced by rough operators. Journal of Algebra and Its Applications 22(10): 2350203., 2023.

[5] F.G. Infusino. Prekernels of topologically axiomatized subcategories of concrete categories. Journal of Algebra 622: 469–505, 2023.

[6] T.B. Iwinski. Algebraic approach to rough sets. Bulletin of the Polish Academy of Sciences, Mathematics 35: 673–683, 1987.

[7] M. Ma and M.K. Chakraborty. Covering-based rough sets and modal logics. part i. International Journal of Approximate Reasoning 77: 55–65, 2016.

[8] A.K. More. A study of algebraic structures and logics based on categories of rough sets. Transactions on Rough Sets XXIII. Springer Berlin Heidelberg, 288–507, 2023.

[9] Z. Pawlak. Rough sets: Theoretical aspects of reasoning about data. Springer Sciences & Business Media, Vol. 9, 2012.

[10] Y.Y. Yao. Constructive and algebraic methods of the theory of rough sets. Information Sciences, 109(1-4): 21–47, 1998.