Position Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 1–8
DOI: 10.15439/2023F7902
ISSN 2300-5963 ACSIS, Vol. 36

Large Minded Reasoners for Soft and Hard Cluster Validation – Some Directions

1st MANI A
Machine Intelligence Unit
Indian Statistical Institute, Kolkata, India
Email: amani.rough@isical.ac.in, a.mani.cms@gmail.com
ORCID: 0000-0002-0880-1035

2nd SUSHMITA MITRA
Machine Intelligence Unit
Indian Statistical Institute, Kolkata, India
Email: sushmita@isical.ac.in
ORCID: 0000-0001-9285-1117

***Abstract*—In recent research, validation methods for soft and hard clustering through general granular rough clusters are proposed by the first author. Large-minded reasoners are introduced and studied in the context of new concepts of non-stochastic rough randomness in a separate paper by her. In this research, the methodologies are reviewed and new low-cost scalable methodologies and algorithms are invented for computing granular rough approximations of soft clusters for many classes of partially ordered datasets. Specifically, these are applicable to datasets in which attribute values are numeric, vector valued, lattice-ordered or partially ordered. Additionally, new research directions are indicated.**

***Index Terms*—Soft Clustering, General Rough Sets, Cluster Validation, Rough Randomness, Non-Stochastic Randomness, Axiomatic Granules, Large Minded Reasoners, Ontology of Clustering, Low Cost Computing, Tolerances, Clean Rough Randomness**

I. Introduction

BOTH hard and soft clustering with or without additional semi-supervision are popular tools for classification in the AIML literature [1]. The problems of cluster validation is a long-standing one. Numeric measures are known to be unreliable, inconsistent and mathematically unjustifiable [2], [3]. Known proofs often proceed from questionable statistical and topological assumptions [1], [4] about the context associated with a dataset. Recent attempts to solve the problem from an axiomatic granular rough set perspective [5], [6] are proposed by the first author in the paper [7]. Ortho-partitions are related to three-way clusterings in other recent work [8]. In the present paper, the focus is on clustering contexts that can be coerced into granular tolerance frameworks. It is not necessary that the clustering be distance-based.

Suppose a soft clustering $\mathcal{S}$ is defined as a finite sequence of ordered pairs of cores and their fringes. The underlying philosophy of the invented methodology is that $\mathcal{S}$ is valid relative to a granular rough model $\mathcal{R}$ if and only if the components of $\mathcal{S}$ are definite objects or are very closely approximated in $\mathcal{R}$.

In the earlier work mentioned [7], a rough model is essential for cluster validation. The meaning associated with the construction of the approximations is the basis of the framework. For example, blocks of a tolerance [9]–[11] can be interpreted as maximal sets of mutually similar objects, and approximations formed as unions of blocks have a simple disjunctive meaning (over higher order concepts of similarity). However, if no rough models seem to be reasonable, then can they be discovered/constructed? A solution for this problem is proposed through a slightly lengthy process involving concepts of clean rough randomness, large minded reasoners, and invented algorithms based on recent advances in the theory of tolerances. Research directions are additionally provided.

The first author's work is supported by grant no. SR/WOS-A/PM-22/2019 of DST, India

A. Structure of this Paper

Some background is provided in the next section. The concept of rough randomness is explained from a different perspective, the associated polysemy is fixed, and large minded reasoners that embody clean rough randomness are formalized in the third section. The overall strategy for validation of soft and hard clustering is formulated next. In the fifth, algorithms for validation methods for truly unsupervised clustering are invented. Related directions and applications are discussed in the last section.

II. Background

A *distance function* on a set S is a function $\rho : S^2 \longmapsto \Re_+$ that satisfies

$$(\forall a)\rho(a,a) = 0 \qquad \text{(distance)}$$

The collection $\mathcal{B} = \{B_\rho(x,r) : x \in S \& r > 0\}$ of all r-spheres generated by ρ is a weak base for the topology τ_ρ defined by

$$V \in \tau_\rho \text{ if and only if } (\forall x \in V \exists r > 0) B_\rho(x,r) \subseteq V$$

Any $\epsilon > 0$ and a distance function ρ determines a tolerance T defined by

$$Tab \text{ if and only if } \rho(a,b) + \rho(b,a) \leq \epsilon.$$

One can define other tolerances through conditions such as

$$\frac{\rho(a,b) + \rho(b,a)}{1 + \rho(a,b) + \rho(b,a)} \leq \epsilon.$$

The point is that a function much weaker than a semimetric suffices for defining a tolerance relation. More complex definitions are often possible.

Thematic track: Rough Sets: Theory and Applications

Proposition 1. *For a numeric complete information table $\mathcal{I}$, the following holds:*

1) *Valuations for each attribute are totally ordered by $\leq$,*
2) *$\mathcal{O}$ is totally ordered relative to the induced lexicographic order.*
3) *$\mathcal{O}$ is lattice ordered relative to $\leq$ defined by $(a_1, \ldots, a_n) \preceq (b_1, \ldots, b_n)$ if and only if $\&_i a_i \leq b_i$ with $a_i, b_i \in Ran(\nu(, At_i)$.*

However, a numeric table is not necessary for any of the three properties to hold.

A. Tolerances

For more details, the reader is referred to the works [9], [12], [13].

If T is a tolerance on a set S, then a *pre-block* of T is a subset $K \subseteq S$ that satisfies $K^2 \subseteq T$. The set of all pre-blocks of T is denoted by $p\mathcal{B}(T)$. Maximal pre-blocks of T with respect to the inclusion order are referred to as *blocks*. The set of all blocks of T is denoted by $\mathcal{B}(T)$. If $S = \langle underline S,\, f_1,\, f_2,\, \ldots,\, f_n, (r_1,\, \ldots,\, r_n)\rangle$ ($\underline{S}$ being a set and f_i being r_i-place operation symbols interpreted on it) is an algebra, then a tolerance T is said to be *compatible* if and only if for each $i \in \{1, 2, \ldots n < \infty\}$,

$$\&_{j=1}^{r_i} T a_j b_j \longrightarrow T f_i(a_1, a_2, \ldots a_{r_i}) f_i(b_1, b_2, \ldots b_{r_i}).$$

When S is a lattice, every tolerance is the image of a congruence by a surjective morphism : $S \longmapsto S$. Further, if $A, B \in \mathcal{B}(T)$, then $\{a \vee b : a \in A \& b \in B\}, \{a \wedge b : a \in A \& b \in B\} \in p\mathcal{B}(T)$. The smallest blocks containing these are unique, and the resulting lattice of blocks is denoted by $S|T$. The set $\mathbf{UBD}(S) = \{\mathcal{B}(T) : T \in Tol(S)\}$ will be referred to as the *universal block distribution* (UBD) of S. It can be assigned the same algebraic lattice order on $Tol(S)$.

A sublattice Z of a lattice S is called a *convex sublattice* if and only if it satisfies $(\forall x, b \in Z)(x \leq a \leq b \longrightarrow a \in Z)$. The blocks of a lattice are all convex sublattices. If C is a subset of S then $\downarrow C$, and $\uparrow C$ will respectively denote the lattice-ideal and lattice-filter generated by C. The following result [12], [14], [15] is not usable for a direct computational strategy:

Theorem 1. *For a finite lattice L, a collection $\mathcal{C}$ of nonempty subsets is the set of all blocks of a tolerance $T \in Tol(L)$ if and only if it is a collection of intervals of the form $\{[a_i, b_i] : i \in I\}$, and*

- $\bigcup_{i \in I} [a_i, b_i] = L$
- *For all $i, j \in I$, $(a_i = a_j \longrightarrow b_i = b_j)$.*
- $(\forall i, j \in I)(\exists k \in I)\, a_k = a_i \vee a_j \,\&\, b_i \vee b_j \leq b_k$.

Theorem 2. *In the context of Theorem 1,*

1) $(\forall C, E \in \mathcal{C})\, (\downarrow C = \downarrow E \Longleftrightarrow \uparrow C = \uparrow E)$.
2) *For any two elements $C, A \in \mathcal{C}$ there exist E, F such that $(\downarrow C \vee \downarrow A) = \downarrow E$, $(\uparrow C \vee \uparrow A) \leq \uparrow E$, $\downarrow F \leq (\downarrow A \wedge \downarrow C)$, and $(\uparrow C \wedge \uparrow A) = \uparrow F$).*

For finite chains, the following holds [16]

Theorem 3. 1) *A collection $\mathcal{C}$ of subsets of the chain $L_n = \langle \{0, 1, 2, \ldots n-1\}, \leq \rangle$ is the set of all blocks of a tolerance $T \in Tol(L)$ if and only if $\mathcal{C}$ is of the form $\{[n_i, m_i] : i = 1, \ldots k\}$ for some $1 \leq k \leq n-1$, with $n_1 = 0$, $m_k = n-1$, and $n_i < n_{i+1} \leq m_i + 1$, and $m_i < m_{i+1}$ for all $i = 1, \ldots k$.*

2) *A collection $\mathcal{C}$ of subsets of the chain $L_n = \langle \{0, 1, 2, \ldots n-1\}, \leq \rangle$ is the set of all blocks of a glued tolerance $T \in Glu(L)$ if and only if $\mathcal{C}$ is of the form $\{[n_i, m_i] : i = 1, \ldots k\}$ for some $1 \leq k \leq n-1$, with $n_1 = 0$, $m_k = n-1$, and $n_i < n_{i+1} \leq m_i < m_i + 1$, and $m_i < m_{i+1}$ for all $i = 1, \ldots k$.*

3) *A collection $\mathcal{C}$ of subsets of the chain $L_n = \langle \{0, 1, 2, \ldots n-1\}, \leq \rangle$ is the set of all blocks of a congruence $R \in Con(L)$ if and only if $\mathcal{C}$ is of the form $\{[n_i, m_i] : i = 1, \ldots k\}$ for some $1 \leq k \leq n-1$, with $n_1 = 0$, $m_k = n-1$, and $n_i < n_{i+1} = m_i + 1$, and $m_i < m_{i+1}$ for all $i = 1, \ldots k$.*

The next theorem is a combination of special cases of known results.

Theorem 4.

- *Tolerances on a product of finite lattices are directly decomposable [17].*
- *If a lattice L is a direct product $\Pi_{i=1}^n L_i$ of the lattices L_i, then $Tol(L) \simeq Tol(L_1) \times Tol(L_2) \times \ldots Tol(L_n)$*

So every $T \in Tol(L)$ can be written as a direct product of tolerances $T_i \in Tol(L_i)$ $(i = 1, 2, \ldots, n)$. That is

$$T = \Pi_{i=1}^n T_i = \{(a, b);\, (e_i a, e_i b) \in T_i, \text{ for } i = 1, 2 \ldots n\}.$$

Further,

Theorem 5. *Let S_1 and S_2 be two lattices with compatible tolerances T_1 and T_2 respectively. If $T(a_1, a_2, \ldots a_n)(b_1, b_2, \ldots b_n)$ if and only if $T_1 a_1 b_1 \& T_2 a_2 b_2 \& \ldots T_n a_n b_n$, then the blocks of T are direct products of the blocks of the component tolerances.*

Proof. Suppose $\mathcal{S}_\infty = \{B_i i \in [1, n]\}$ and $\mathcal{S}_\in = \{F_i i \in [1, m]\}$ are two distinct normal covers for the same tolerance T on an algebra S. If Tab for any $a, b \in S$, then it is necessary that $a, b \in F_i$, and $a, b \in B_j$ for some i, j. This means $\mathcal{S}_\infty$ must be a mere rearrangement of $\mathcal{S}_\in$. In other words, normal covers of a tolerance are unique. □

The above proof works for finite direct products as well. It means that one needs to create all permutations of the blocks on components in general.

B. Approximations

Let $S = \langle \underline{S}, T \rangle$ be a general approximation space, with $\underline{S}$ being a set, and T a tolerance relation on it. If $\mathcal{G}$ is the set of all blocks of T, and $A \subseteq S$, then the following granular approximations [9] will be used in the main algorithms (the semantics and history of the approximations are described in the mentioned reference).

$$A^l = \bigcup\{H : H \subseteq A,\, H \in \mathcal{G}\} \quad \text{(Lower)}$$
$$A^u = \bigcup\{H : A \cap H \neq \emptyset,\, H \in \mathcal{G}\} \quad \text{(Upper)}$$
$$A^{u_b} = A^u \setminus A^{cl} \quad \text{(Bited Upper)}$$

Further, the model $\mathbb{S} = \left\langle \underline{\wp(\mathbb{S})}, \gamma, l, u, \subseteq, \cup, \cap, \emptyset, S \right\rangle$ generated by the granular approximations on the power set $\underline{\wp(\mathbb{S})}$ will be used to discuss cluster validation. It is additionally a set HGOS in the sense of the first author [6].

C. Validation Indices – a Brief Critique

The process or concept of *cluster validation* generally refers to exploring the quality of one or more clustering methods and possibly comparing them. In almost all cases, true class information is not available (that is if one avoids looking at anything apart from the dataset) and validation methods are inherently not rigorous even in comparison to statistical methods used in supervised learning. For example, in a regression modeling context, it is possible to say something concrete about model fit relative to a set of statistical assumptions (that may be invalid). Clustering contexts are difficult to investigate from similar perspective. In this subsection, some issues faced are mentioned.

A number of indices for quantifying a clustering's quality are known. Typically, they are used to simply assess the quality of a single clustering or to select the most appropriate clustering method and related parameters (like number of optimal clusters). For datasets of the form $\{x_i\}_{i=1}^n$ over a Euclidean space with standard norm or some distance ρ, common indices such as the Davies-Bouldin (DB) index, Calinski–Harabasz(CH) index and variants thereof start with measures of within-cluster variation and compare them with measures of between-cluster variation from a numerical perspective over the real number field. Many adaptations to rough clustering are also known [18], [19]. In the case of CH-index, the shape of associated clusters are assumed to be spherical, with data points concentrated around the cluster means. Even if points close to the boundary of the sphere are close to points on the boundary of another cluster, the distance between the two clusters will be the distance between the means. Further, the index is naturally connected with the hard k-means algorithm. These can be used in determining the appropriateness of the index in a specific application context. From this *it should be clear that indices carry very little information about ontology*.

Latent class model-based approach (LCC) is sometimes used for clustering multivariate categorical data. There each cluster is assumed to be a mixture component and the whole is a mixture of probability distributions [20]. These are not well-related to distance based approaches, though the *average silhouette width* (ASW) measure that emphasizes the separation between clusters and their neighbors is known to be useful. ASW is also heavily used in fuzzy clustering.

External Criteria: A simple example of an external criterion for hard clustering is the *quality index* Q_+. If N_+, N_- and N are respectively the number of correctly assigned, incorrectly assigned and total number of clusters.

$$Q_+ = \frac{N_+}{N} \,\&\, Q_- = \frac{N_-}{N}.$$

This measure can be generalized to rough clustering [19], [21], and other soft approaches (with cores and fringes). Obviously these indices avoid most of the complexity and semantics involved in the clustering process. However, they allow gradation of boundaries, and are less controversial than other indices because of the minimal number of assumptions.

In general, the following remarks about indices may be noted:

- Cluster validation is sometimes done from a statistical test perspective. The null hypothesis is taken to be the statement that the data homogeneous and unclustered according to a null model. Reasonable clusterings are expected to be significantly better than what is expected relative to its performance on the null model. This is usually done relative to specific cluster performance indices. Both the reality of the statistical scheme of things assumed and indices used remain very questionable.
- Clusters with complicated shapes are common in application contexts like image processing. For example, in many photos of natural scenarios, similar leaves can be in different parts [22] (these are handled with descriptive proximities and related functions). Indices for clusters in such contexts are not well-developed.
- Combining multiple validation indices for the purpose of measuring multiple characteristics has limited scope and the act of combining does not go beyond forming a set of indices [3].

In hierarchical clustering in particular, indices such as partial R-squared monotonically change with number of clusters. Strong decrease (or increase) followed by weak decrease (or increase) of the index relative to the number of clusters correspond to their optimal values. Further, the applicability of associated indices is limited.

It can therefore be asserted that *most indices assume some heuristics that are not well understood and in some cases even the values produced may not be clear (see [2], [3]) for details*. From a mathematical point of view, a few rigorous studies on indices in the context of semimetric based clustering are known.

III. Clean Rough Randomness and LMR

Many types of randomness are known in the literature. Stochastic randomness, often referred to as randomness, is often misused without proper justification. In the paper [23], a phenomenon is defined to be *stochastically random* if it has probabilistic regularity in the absence of other types of regularity. In this definition, the concept of regularity may be understood as *mathematical regularity* in some sense. Generalizations of mathematical probability theory through hybridization with rough sets from a stochastic perspective are explained in the book [24]. This approach is not ontologically

consistent with pure rough reasoning or explainable AI as its focus is on modeling the result of numeric simplifications in a measure-theoretic context.

A *rare property* in the theory of computation is an effectively testable property that is valid over a set of measure zero. A finite or infinite sequence $\mathbf{x}$ is said to be *algorithmically random* if and only if no computational agent recognizes $\mathbf{x}$ as possessing some rare property (for details see [23], [25]). While associations with subjective probability are known, connections of such ideas with rough sets are not known in the literature.

Empirical studies show that humans cannot estimate measures of stochastic randomness and weakenings thereof in real life properly [26]. This is consistent with the observation that connections in the rough set literature between specific versions of rough sets and subjective probability theories (Bayesian or frequentist) are not good approximations. In fact, rough inferences are grounded in some non-stochastic comprehension of attributes (their relation with the approximated object in terms of number or relative quantity and quality) [27], [28].

The idea of *rough randomness* is expressed by the first author [29] as follows: *a phenomenon is roughly random if it can be modeled by general rough sets or a derived process thereof*. To avoid the resulting polysemy (as the term is used in a different sense in the monograph [24]), it is useful to rename it as *clean rough randomness* (or C-rough randomness for short). In concrete situations, such a concept should be realizable in terms of C-roughly random functions or predicates defined below (more variations will be part of future work):

Definition 1. *Let $\mathcal{A}_\tau$ be a collection of approximations of type τ, and E a collection of rough objects [9] defined on the same universe S, then by a* C-rough random function of type-1 *(CRRF1) will be meant a partial function*

$$\xi : \mathcal{A}_\tau \longmapsto E.$$

Definition 2. *Let $\mathcal{A}_\tau$ be a collection of approximations of type τ, $\mathcal{S}$ a subset of $\wp(S)$, and $\Re$ the set of reals, then by a* C-rough random function of type-2 *(CRRF2) will be meant a function*

$$\chi : \mathcal{A}_\tau \times \mathcal{S} \longmapsto \Re.$$

Definition 3. *Let $\mathcal{A}_\tau$ be a collection of approximations of type τ, and F a collection of objects defined on the same universe S, then by a* C-rough random function of type-3 *(CRRF3) will be meant a function*

$$\mu : \mathcal{A}_\tau \longmapsto F.$$

Definition 4. *Let $\mathcal{O}_\tau$ be a collection of approximation operators of type τ_l or τ_u, and E a collection of rough objects defined on the same universe S, then by a* C-rough random function of type-H *(CRRFH) will be meant a partial function*

$$\xi : \mathcal{O}_\tau \times \wp(S) \longmapsto E.$$

It is obvious that a CRRF1 and CRRF2 are independent concepts, while a total CRRF1 is an CRRF3, and CRRFH is distinct (though related to CRRF3). The set of all such functions will respectively be denoted by $CRRF1(S, E, \tau)$, $CRRF2(S, \Re, \tau)$, $CRRF3(S, F, \tau)$, and $CRRFH(S, E, \tau)$. Examples that show the semantic nature of the associations are mentioned below:

Examples: CRRF

Example 1. *Let S be a set with a pair of lower (l) and upper (u) approximations satisfying (for any $a, b, x \subseteq S$)*

$$x^l \subseteq x^u \tag{int-cl}$$
$$x^{ll} \subseteq x^l \tag{l-id}$$
$$a \subseteq b \longrightarrow a^l \subseteq b^l \tag{l-mo}$$
$$a \subseteq b \longrightarrow a^u \subseteq b^u \tag{u-mo}$$
$$\emptyset^l = \emptyset \tag{l-bot}$$
$$S^u = S \tag{u-top}$$

The above axioms are minimalist, and most general approaches satisfy them.

In addition, let

$$\mathcal{A}_\tau = \{x : (\exists a \subseteq S)\, x = a^l \text{ or } x = a^u \tag{1}$$
$$E_1 = \{(a^l, a^u) : a \in S\} \tag{E1}$$
$$F = \{a : a \subseteq S \& \neg\exists b b^l = a \vee b^u = a\} \tag{E0}$$
$$E_2 = \{b : b^u = b \& b \subseteq S\} \tag{E2}$$
$$\xi_1(a) = (a, b^u) \text{ for some } b \subseteq S \tag{xi1}$$
$$\xi_2(a) = (b^l, a) \text{ for some } b \subseteq S \tag{xi2}$$
$$\xi_3(a) = (e, f) \in E_1 \,\&\, e = a \text{ or } f = a \tag{xi3}$$

E_1 in the above is a set of rough objects, and a number of algebraic models are associated with it [9]. A partial function $f : \mathcal{A}_\tau \longmapsto E_1$ that associates $a \in \mathcal{A}_\tau$ with a minimal element of E_1 that covers it in the inclusion order is a CRRF of type 1. For general rough sets, this CRRF can be used to define algebraic models and explore duality issues [13], and for many cases associated these are not investigated. A number of similar maps with value in understanding models [27] can be defined. Rough objects are defined and interpreted in a number of other ways including F or E_2.

Conditions xi1-xi3 may additionally involve constraints on b, e and f. For example, it can be required that there is no other lower or upper approximation included between the pair or that the second component is a minimal approximation covering the first. It is easy to see that

Theorem 6. *ξ_i for $i = 1, 2, 3$ are CRRF of type-1.*

Example 2. *In the context of the above example, rough inclusion functions, membership, and quality of approximation functions [30], [31] can be used to define CRRF2s. An example is the function ξ_5 defined by*

$$\xi_5(a, b) = \frac{Card(b \setminus a)}{Card(b)} \tag{III.1}$$

In the paper [29], it is additionally proved that

Theorem 7. *A rough random variable [24] in the sense of Liu, is not a rough random function of any type.*

IV. Hard and Soft Clustering Validation Strategies

The considerations of this section will be restricted to validation of soft clustering defined through ortho-pairs [8]. In the mentioned paper, the authors do not explicitly say that their universe is finite, and it is not clarified whether it is a semimetric set (a set with a semimetric) or a semimetric space. The former does not always define a semimetric topology. Connections with proximities [32] are additionally not mentioned. *However, these assumptions are not required for obtaining three-way clusters or rough clusters in their sense*.

An ortho pair is a pair of disjoint subsets (of a universe S) of the form $O = (C, F)$ with C being the core and F being the boundary or fringe that satisfies $C \cap F = \emptyset$. An ortho-partition $\mathcal{O}$ is a collection of ortho pairs of the form

$$\{(C_1, F_1),\, (C_2, F_2),\, (C_3, F_3),\, \ldots, (C_n, F_n)\}$$

that satisfies O0, O1, O2, and O3

$$\text{For all } i\; C_i \neq \emptyset \qquad \text{(T1)}$$

$$\text{For all } i\; C_i \cap F_i = \emptyset \qquad \text{(O0)}$$

$$\text{If } i \neq j \text{ then } C_i \cap C_j = C_i \cap F_j = C_j \cap F_i = \emptyset \qquad \text{(O1)}$$

$$\bigcup (C_i \cup F_i) = S \qquad \text{(O2)}$$

$$(\forall x)(x \in F_i \longrightarrow (\exists j) j \neq i \,\&\, x \in F_j) \qquad \text{(O3)}$$

$$\text{If } \forall i\; x \notin C_i \text{ then } \exists i, j\; i \neq j \,\&\, x \in F_i \cap F_j \qquad \text{(R2)}$$

A *rough clustering* is a collection of ortho pairs that satisfies O0, O1, and R2. While a *three-way clustering* is a collection of orthopairs that satisfies O1, O2, and T1. However, it is interpreted as a soft clustering $\mathcal{K}$ in which each cluster K_i is associated with three regions C_i, F_i and $E_i = (C_i \cup F_i)^c$. The last region being interpreted as the *certainly not that* region.

While rough clusterings can be interpreted as ortho-partitions, three-way clusterings are collections of ortho-pairs that do not satisfy O3 in general. However, it is possible to collect the elements not satisfying O3 and create a new cluster – the resulting clustering satisfies O3. Therefore, ortho-partitions suffice for representing semi metric based rough, and three-way clustering, and have a few arguably nice properties (of scale invariance, generalized richness and consistency). If $\mathfrak{D}(S)$ is the set of all semimetrics on S, and $\Pi(S)$ the set of all partitions of S, and $\mathcal{O}(S)$ the set of all ortho partitions on S, then an algorithm is a computable function $c_{tw} : \mathfrak{D}(S) \longmapsto \mathcal{O}(S)$.

Here we are concerned with validation techniques for the clustering. Our basic principle for validation that *if the interpretation of the cores and exteriors are almost the same as their respective approximations in a granular rough semantics, then they are valid relative to the semantics*. This is useful because granular rough semantics in the sense of the first author [6], [9] can explain the meaning of the clusters. Formally,

Definition 5. *Let* $\mathbb{S} = \left\langle \underline{\wp(\mathbb{S})}, \gamma, l, u, \subseteq, \leq, \cup, \cap, \emptyset, S \right\rangle$ *be the set HGOS generated by a tolerance and its granular approximations. Further, let* $\mathcal{Z} = \{(C_i, F_i) i = 1, \ldots r\}$ *be a soft clustering on* S.

- *The* lower deficit *of a soft cluster* $(C, F) \in \mathcal{Z}$ *will be the pair* $((C \setminus C^l)^u, (F \setminus F^l)^u)$,
- *While its* upper deficit *will be the pair* $((C^u \setminus C)^u, (F^u \setminus F)^u)$

The lower and upper deficit of (C, F) *will respectively be denoted by* $(C^\flat, F^\flat))$ *and* $(C^\eth, F^\eth)$. *For hard clustering, it suffices to restrict attention to the core alone.*

Definition 6. *In the context of Definition 5, a soft cluster* $(C, F) \in \mathcal{Z}$ *will be said to*

- lu-valid *if and only if* $C^l = C^u = C$ *and* $F^l = F^u = F$
- l-pre-valid *if and only if* $(\exists V, W \in \mathbb{S}) V^l = C \& W^l = F$.
- u-pre-valid *if and only if* $(\exists V, W \in \mathbb{S}) V^u = C \& W^u = F$.
- l-traceable *if and only if* $(\exists V, W \in \mathbb{S}) V = C^l \& W = F^l$.
- u-traceable *if and only if* $(\exists V, W \in \mathbb{S}) V = C^u \& W = F^u$.

In addition, if all soft clusters in $\mathcal{Z}$ *are* l*-pre-valid (resp.* lu*-valid,* u*-pre-valid,* l*-traceable,* u*-traceable) then* $\mathcal{Z}$ *will itself be said to be* l-pre-valid *(resp.* lu*-valid,* u*-pre-valid,* l*-traceable,* u*-traceable).*

Proposition 2. *In the context of Definition 5, if the* l*-deficit (resp.* u*-deficit) of a hard cluster* C *is computable, then it must necessarily be* l*-traceable (resp.* u*-traceable).*

Proof. If a cluster C has l-deficit A, then it is necessary that $A = (C \setminus C^l)^u$. However, for this C^l should be an element of $\mathbb{S}$. The proof for the u-deficit is similar. □ □

Proposition 3. *In the context of Definition 5, if the* l*-deficit (resp.* u*-deficit) of a soft cluster* (C, F) *is computable, then it must necessarily be* l*-traceable (resp.* u*-traceable).*

The central idea of lu-validity (and weakenings thereof) is that of representability in terms of granules and approximations. These do not test the key predicate δ for validation, and the aspect is left to the process of construction of rough approximations. By contrast, the $*$-deficits are an internal measure of what is lacking or what is in excess.

In relation to the framework of minimal soft clustering system (MSS) invented in the paper [7], it is possible to define the ternary predicate δ through the above concepts. Intended meanings of δabc are a *is closer to* b *than* c in some sense, a *is more similar to* b *than* c *in some sense* and variants thereof. This predicate covers the intent of using metrics, similarities, dissimilarities, proximities, descriptive proximities, kernels and other functions for the purpose.

V. Algorithms and LMR Algorithms

In this section, improved algorithms for the computation of blocks that avoid weakenings are invented. These improve earlier work of the first author [29].

A. Direct Algorithms-1,2

The following two algorithms are used in a forthcoming paper on satellite remote sensing by the first author. They are resource intensive, as the computational ease is limited by the max-clique algorithms. Their complexities are directly defined by that of the similarity matrix computation and the maximal clique algorithms. In the paper, a low-cost implementation could be used through supplementary measures.

Suppose a hard clustering $\{C_i\}_{i=1}^k$ or a soft clustering $\{(C_i, E_i)\}_{i=1}^k$ [1], [8] obtained through any method is given.

Algorithm-1:

Distance: Specify distinct distance function σ_i on the i th column (attribute) for each i

Tolerance: Define a similarity (tolerance relation) T_i on the ith column.

Conjunction: Combine to a single tolerance relation over objects on the table through conjunction of instances across columns.

Relation: Compute the similarity matrix through parallelized methods.

Granules: Compute the blocks of the tolerance by a maximal clique algorithm (for example the modified Bronkerbosch algorithm [33]).

Approximations: Compute granular rough approximations of C_i, and E_i for each i and estimate the closeness of the cluster core or exterior to decide on validation.

If the rough model can explain the soft/hard clustering, then the latter is meaningful and valid.

Algorithm-2:

Distance: Specify a single distance function σ between objects

Tolerance: Define a similarity (tolerance relation) T on the basis of descriptive statistics relative to σ.

Relation: Compute the similarity matrix through parallelized methods.

Granules: Compute the blocks of the tolerance by a maximal clique algorithm.

Approximations: Compute granular rough approximations of C_i, and E_i for each i and estimate the closeness of the cluster core or exterior to decide on validation.

Theorem 8. *The direct algorithm-2 has a computational complexity of* $\mathsf{O}(dk^{3^{d/3}} + N^3)$*, where* d *is the degeneracy of the* n *-vertex graph corresponding to the similarity relation,* k *the number of rows, and* N *is the maximum of number of rows and columns in the dataset.*

Proof. $\mathsf{O}(dk^{3^{d/3}})$ is the complexity of computing max-cliques, while $\mathsf{O}(N^3)$ is that of computing the distance matrix. List operations are assumed to be of linear complexity. □

B. Improved AGRSSA (IAGRSSA)

In an earlier preprint of the first author [29], the AGRSSA (Axiomatic Granular Reversed Similarity Based Semi-Supervised) algorithm(s) was proposed. This is improved below through relaxed assumptions, and stricter constraints on the decision steps involved. It is assumed that each column (attribute) is totally ordered, and that an order-compatible distance (as opposed to a metric) is defined on them. Specifically, it applies to all numeric (real valued) datasets.

IAGRSSA:

Distance: Specify distinct distance functions on each column (attribute).

Quantiles: Identify f-quantiles at a suitable level of precision on each column. Let these be $\{q_{i1}, q_{i2}, \ldots q_{if}\}$ on the ith column based on the distance specified earlier.

Interval Boundaries: Interval boundaries are specifiable by the sequence $\perp_i, q_{i1} - e_{i1}, q_{i1} + e_{i1}, q_{i2} - e_{i2}, q_{i2} + e_{i2}, \ldots, q_{if} - e_{if}, q_{if} + e_{if}, \top_i$. The quantities $e_{i1}, e_{i2}, \ldots e_{if}$ need to computed as a fraction of the measures of variation or other heuristics.

Decision on Blocks: Assume that the intervals on each column are exactly the set of blocks.

Blocks: Form all possible products of the sequence of blocks on each column to form the set of admissible blocks. That is if $\{B_{ij}\}_{j=1}^f$ is the set of blocks on the ith column, the blocks of the whole dataset would have the form $B_{1j_1} \times B_{2j_2} \times \ldots \times B_{nj_k}$, with k being the number of columns and j_i taking values from $1, 2, \ldots f$.

Approximations: Compute granular rough approximations by Subsection II-B and perform decision-making. If a set of objects H are to be approximated, then

1) The lower approximation of H is the union of blocks included in it.
2) The lower approximation of H is the union of blocks that have some common elements with H.

Meaning: This can be specified directly from blocks, or through its associated tolerance.

IAGRSSA does not require any decision column on the dataset, and yet its computational complexity is far below that of the direct algorithms. Ideally, the block construction process should involve supervision as it requires an understanding of the attributes. AGRSSA-M [29] is based on reducing the blocks required for decisions.

Theorem 9. *IAGRSSA computes the blocks of the tolerance constructed as a direct product of the tolerances on each column.*

Proof. First, chains and partial orders on a set are equivalent to specific groupoidal operations [9], [34], and the compatibility is assumed with respect to these. Additionally, direct products

of groupoids are groupoids. The rest follows from Theorems 5 and 3. □

The next example illustrates the product.

Example 3. *Let T_1 be the tolerance on Q defined by*

$$T_1 ab \text{ if and only if } |a-b| \leq 1,$$

and T_2 be the tolerance on $\Re$ defined by

$$T_2 ab \text{ if and only if } |a-b| \leq e.$$

On the product set $Q \times \Re$ with the induced lattice order, the product tolerance T is defined by the condition

$$T(a_1, a_2)(b_1, b_2) \text{ if and only if } T_1 a_1 b_1 \& T_2 a_2 b_2.$$

The blocks of the tolerance T_1 are of the form $\{x : |x-q| \leq 0.5\}$ for distinct $q \in Q$ The blocks of the tolerance T_2 on the other hand are of the form $\{x : |x-a| \leq 0.5e\}$ for distinct $a \in \Re$. The blocks of the product need to be formed by taking a direct product of these as the components are independent. It is therefore the set

$$\{(x, w) : |x-q| \leq 0.5 \& |w-a| \leq 0.5e\} \text{ for } q \in Q \,\&\, a \in \Re.$$

Exhaustive Tolerance Discovery Algorithm (ETDA-LMR)

A *large-minded reasoner* is so-named because it is essentially about discovering suitable similarities. It selects the more reasonable collections of blocks. The exhaustive tolerance discovery algorithm that involves LMRs is invented in the paper [29]. However, it is relatively opaque as it leaves out the crucial steps of selection to the dynamics of the context. A natural question is: can the simplicity of the structure of blocks on chains be exploited to improve the meta algorithm.

Definition 7. *A* large-minded reasoner *(LMR) is a partial function* $\psi : \mathbf{UBD}(A_1) \times \mathbf{UBD}(A_2) \times \ldots \mathbf{UBD}(A_n) \longmapsto \mathbf{UBD}(A)$.

The ETDA algorithm applies to all kinds of information tables including decision tables, and in clearer terms is given below.

EDTA Algorithm:

Step 1: Define sequences of q number of quantiles – this by itself means a certain understanding of categories associated with attributes.
Step 2: Using the quantiles form intervals (with or without intersections) under the conditions of Theorem 3.
Step 3: Optionally, some intervals may be fused together in relation to relative changes in decisions. This amounts to removing interval boundaries.
Step 4: Specify the large minded reasoner ψ. This is the same as defining a number of compatible tolerances using the intervals.
Step 5: Identify the defined tolerances in ψ.
Step 6: Compute relevant lower, upper, and bited approximations and optionally the associated decisions for each normal cover.
Step 7: In case of soft cluster validation, compute the approximations of the cores and exteriors, and evaluate their closeness to the evaluated.
Step 8: Select relevant tolerances (or normal covers) specified in ψ.
Step 9: Explain the data context on the basis of the associated tolerance(s) (or normal covers).

Example 4. *The first three steps of the ETDA algorithm are illustrated in this example. Let $\{1, 5, 6, 9, 10\}$ be a sequence of equally spaced quantiles. Some sets of intervals that can be formed with these are*

$$\mathcal{B}_1 = \{[1,6],\ [5,9], [6,10]\}$$
$$\mathcal{B}_2 = \{[1,5],\ [5,6], [6,9], [9,10]\}$$
$$\mathcal{B}_3 = \{[1,9],\ [5,10]\}.$$

Definition 8. *By an* interpreted large-minded reasoner *associated with ψ of Def. 7 will be meant a partial function $\psi^* : \mathbf{UB}(A_1) \times \mathbf{UB}(A_2) \times \ldots \mathbf{UB}(A_n) \times \wp(S) \longmapsto \mathbf{UB}(A)$, that indicates the granular components or parts of approximations of subsets.*

The function is intended to represent the compositionality of approximations in terms of blocks of components. These can be quite complex (see [5], [9]), and so *granular components or parts of approximations* are referred to.

Theorem 10. *ψ^* is a CRRF of type H.*

VI. Problems and Directions

The invented algorithms appear to be well-suited for low cost computing. A detailed investigation is necessary to confirm the same. Additionally, it is necessary to formulate post-processing techniques for seamless interpretation. If a context results in a hundred blocks, then a description of the blocks, and the approximations generated is essential for keeping track of meaning. How does one solve this *problem of meaningful empirical representation*?

One way is to encode the blocks lexicographically on the basis of its position on components (or columns). Next, the extent of expression of these blocks (encoding) in relatively important clusters can be computed. The combination of these expressions can be expressed in natural language with limited or no involvement of numeric estimates. A full solution of this problem will appear separately.

The logic of decision-making on the basis of set-theoretic measures for cluster validation requires additional work and will appear in a forthcoming paper. A substantial amount of the machinery required is invented in the papers [35], and a forthcoming three part paper by the first author.

In similarity-based clustering [36], clusters are formed from data supplemented with similarity grades (usually with values in the real interval $[0, 1]$ or the rationals Q) between data points. For a set of n data points, the associated similarity matrix formed by these similarity grades is a symmetric square matrix $K = (s_{ij})_{n \times n}$ with s_{ij} being the similarity between the ith and jth data point. These can as well be approached

through spectral clustering methods. Can the EDTA algorithm be extended to these contexts?

A. Problems of Medical Imaging and Beyond

Big datasets associated with medical images (obtained through MRI, FMRI and CT scans) are mostly patterns formed by products of finite totally ordered subsets of the reals. Rough sets combined with clustering techniques are used to identify brain tumors in the presence of bias field and noise in recent work [37]. However, additional methods need to be employed to possibly rectify the results. Specifically, the CoLoRS segmentation algorithm does not clearly provide the reasons for inclusion or exclusion of tumors or healthy tissues. It is of interest to use the transparent algorithms invented in this research to these problem contexts, and additionally in those for identification of lesions in lung CT [38]. These characteristics are typical of a number of other application contexts of AIML, and therefore the application contexts are boundless.

REFERENCES

[1] C. Hennig, M. Meila, F. Murtagh, and R. Rocci, *Handbook of Cluster Analysis*, 1st ed. Edited Volume. Chapman and Hall: CRC Press, 2016.

[2] M. Kim and R. S. Ramakrishna, "New Indices For Cluster Validity Assessment Pattern," *Pattern Recognition Letters*, vol. 26, pp. 2353–2363, 2005. doi: 10.1016/j.patrec.2005.04.007

[3] C. Hennig and T. F. Liao, "Comparing Latent Class and Dissimilarity Based Clustering for Mixed Type Variables..." *Journal of the Royal Statistical Society*, pp. 309–369, 2013. doi: 10.1111/j.1467-9876.2012.01066.x

[4] C. Bouveyron, G. Celeux, B. Murphy, and A. Raftery, *Model-Based Clustering and Classification for Data Science: With Applications in R*. Cambridge University Press, 2019.

[5] A. Mani, "Dialectics of Counting and The Mathematics of Vagueness," *Transactions on Rough Sets*, vol. XV, no. LNCS 7255, pp. 122–180, 2012. doi: 10.1007/978-3-642-31903-7_4

[6] ——, "Comparative Approaches to Granularity in General Rough Sets," in *IJCRS 2020*, ser. LNAI, R. Bello *et al.*, Eds. Springer, 2020, vol. 12179, pp. 500–518.

[7] ——, "General Rough Modeling of Cluster Analysis," in *Rough Sets: IJCRS-EUSFLAT 2021*, ser. LNAI 12872, S. Ramanna *et al.*, Eds. Springer Nature, 2021.

[8] A. Campagner and D. Ciucci, "A formal learning theory for three-way clustering," in *SUM 2020*, J. Davis and K. Tabia, Eds. Springer, 2020, vol. LNAI 12322, pp. 128–140.

[9] A. Mani, "Algebraic Methods for Granular Rough Sets," in *Algebraic Methods in General Rough Sets*, ser. Trends in Mathematics, A. Mani, I. Düntsch, and G. Cattaneo, Eds. Birkhauser Basel, 2018, pp. 157–336.

[10] ——, "Algebraic Semantics of Similarity-Based Bitten Rough Set Theory," *Fundamenta Informaticae*, vol. 97, no. 1-2, pp. 177–197, 2009. doi: 10.3233/FI-2009-196

[11] P. Wasilewski and D. Ślęzak, "Foundations of Rough Sets from Vagueness Perspective," in *Rough Computing: Theories, Technologies and Applications*, A. Hassanien *et al.*, Eds. IGI, Global, 2008, pp. 1–37.

[12] I. Chajda, *Algebraic Theory of Tolerance Relations*. Olomouc University Press, 1991. [Online]. Available: https://www.researchgate.net/publication/36797871

[13] A. Mani, "Representation, Duality and Beyond," in *Algebraic Methods in General Rough Sets*, ser. Trends in Mathematics, A. Mani, I. Düntsch, and G. Cattaneo, Eds. Birkhauser Basel, 2018, pp. 459–552.

[14] H. J. Bandelt, "Tolerance relations of a lattice," *Bulletin Austral.Math. Soc.*, vol. 23, pp. 367–381, 1981. doi: 10.1017/S0004972700007255

[15] G. Czedli and L. Klukovits, "A note on tolerances of idempotent algebras," *Glasnik Matematicki (Zagreb)*, vol. 18, pp. 35–38, 1983. [Online]. Available: https://web.math.pmf.unizg.hr/glasnik/18.1/18103.pdf

[16] A. Górnicka, J. Grygiel, and I. Tyrala, "On the lattice of tolerances for a finite chain," *Czestochowa Mathematics*, vol. XXI, pp. 25–30, 2016. doi: 10.16926/m.2016.21.03

[17] J. Niederle, "A note on tolerance lattices of products of lattices," *Casop. Pest. Matem.*, vol. 107, pp. 114–115, 1982. doi: 10.21136/CPM.1982.118112

[18] S. Mitra, "An Evolutionary Rough Partitive Clustering," *Pattern Recognition Letters*, vol. 25, no. 12, pp. 1439–1449, 2004. doi: 10.1016/j.patrec.2004.05.007

[19] G. Peters, "Rough Clustering Utilizing the Principle of Indifference," *Information Sciences*, vol. 277, pp. 358–374, 2014. doi: 10.1016/j.ins.2014.02.073

[20] J. Vermunt and J. Magidson, "Latent Class Cluster Analysis," in *Applied Latent Class Analysis*. Cambridge: Cambridge University Press, 2002, pp. 89–106.

[21] I. Düntsch and G. Gediga, "Rough Set Clustering," in *Handbook of Cluster Analysis*, C. Hennig, M. Meila, and F. Murtagh, Eds. CRC Press, 2016, ch. 25, pp. 575–594.

[22] A. D. Concilio, C. Guadagni, J. Peters, and S. Ramanna, " Descriptive Proximities. Properties and Interplay Between Classical Proximities and Overlap," *Mathematics in Computer Science*, vol. 12, no. 1, pp. 91–106, 2018. doi: 10.1007/s11786-017-0328-y

[23] A. N. Kolmogorov, "On the logical foundations of probability theory," in *Selected Works of A. N. Kolmogorov*, A. N. Shiryayev, Ed. Kluwer Academic, Nauka, 1986, vol. 2, ch. 53, pp. 515–519.

[24] B. Liu, *Uncertainty Theory*, ser. Studies in Fuzziness and Soft Computing. Springer, 2004, vol. 154.

[25] T. Steifer, "A note on learning theoretic characterizations of randomness and convergence," *Review of Symbolic Logic*, vol. 15, no. 3, pp. 807–822, 2022. doi: 10.1017/S1755020321000125

[26] L. Beach and G. Braun, "Laboratory studies of subjective probability: a status report," in *Subjective Probability*, G. Wright and P. Ayton, Eds. John Wiley, 1994, pp. 107–128.

[27] A. Mani, I. Düntsch, and G. Cattaneo, Eds., *Algebraic Methods in General Rough Sets*, ser. Trends in Mathematics. Birkhauser Basel, 2018. ISBN 978-3-030-01161-1

[28] P. Pagliani and M. Chakraborty, *A Geometry of Approximation: Rough Set Theory: Logic, Algebra and Topology of Conceptual Patterns*. Berlin: Springer, 2008.

[29] A. Mani, "Rough randomness and its application," *Journal of the Calcutta Mathematical Society*, pp. 1–15, 2023. doi: 10.5281/zenodo.7762335. [Online]. Available: https://zenodo.org/record/7762335

[30] J. Stepaniuk, *Rough-Granular Computing in Knowledge Discovery and Data Mining*, ser. Studies in Computational Intelligence,Volume 152. Springer-Verlag, 2009. ISBN 978-3-540-70800-1

[31] A. Gomolinska, "Rough Approximation Based on Weak q-RIFs," *Transactions on Rough Sets*, vol. X, pp. 117–135, 2009. doi: 10.1007/978-3-642-03281-3_4

[32] M. Gagrat and S. Naimpally, "Proximity approach to semi-metric and developable spaces," *Pacific Journal of Mathematics*, vol. 44, no. 1, pp. 93–105, 1973. doi: 10.2140/pjm.1973.44-1

[33] O. Cheong *et al.*, Eds., *Listing All Maximal Cliques in Sparse Graphs in Near-Optimal Time*, ser. LNCS 6506. Berlin: Springer, 2010. doi: 10.1007/978-3-642-17517-6_36

[34] R. Freese, J. Jezek, J. Jipsen, P. Markovic, M. Maroti, and R. Mckenzie, "The Variety Generated by Order Algebras," *Algebra Universalis*, vol. 47, pp. 103–138, 2002. doi: 10.1007/s00012-002-8178-z

[35] A. Mani, "Granularity and Rational Approximation: Rethinking Graded Rough Sets," *Transactions on Rough Sets*, vol. XXIII, no. LNCS 13610, pp. 33–59, 2022. doi: 10.1007/978-3-662-66544-2_4

[36] M. Meila, "Spectral Clustering," in *Handbook of Cluster Analysis*, C. Hennig, M. Meila, and F. Murtagh, Eds. CRC Press, 2016, ch. 7, pp. 125–144.

[37] S. Roy and P. Maji, "Tumor delineation from 3-d mr brain images," *Signal, Image and Video Processing*, pp. 1–9, 2023. doi: 10.1007/s11760-023-02565-4

[38] P. Dutta and S. Mitra, "Composite deep network with feature weighting for improved delineation of covid infection in lung ct," *MedRxiv*, pp. 1–22, 2023. doi: 10.1101/2023.01.17.23284673