Granular Computing for Machine Learning: Pursuing New Development Horizons

Witold Pedrycz, *Life Fellow, IEEE*

***Abstract*—Undoubtedly, machine learning (ML) has demonstrated a wealth of far-reaching successes present both at the level of fundamental developments, design methodologies and numerous application areas, quite often encountered in domains requiring a high level of autonomous behavior. Over the passage of time, there are growing challenges of privacy and security, interpretability, explainability, confidence (credibility), and computational sustainability, among others. In this study, we advocate that these quests could be addressed by casting them both conceptually and algorithmically in the unified environment augmented by the principles of granular computing. It is demonstrated that the level of abstraction, delivered by granular computing plays a pivotal role in the interpretation by quantifying the level of credibility of ML constructs. The study also highlights the principles of granular computing and elaborates on its landscape. The original idea of a comprehensive and unified framework of data-knowledge environment of ML is introduced along with a detailed discussion on how data and knowledge are used in a seamless fashion by invoking granular embedding and producing relevant loss functions. Key categories of knowledge-data integration realized at the levels of data and model (involving symbolic/qualitative models and physics-oriented models) and investigated.**

***Index Terms*—Gaussian process (GP), granular computing, granular embedding, information granules, knowledge-data design environment, machine learning (ML).**

I. Introduction

MACHINE learning (ML) constructs have assumed a central role in the design and analysis of intelligent systems. Along with success stories, there have been a rising awareness about challenges that become more apparent with the application areas.

Granular computing is regarded as a unified environment of processing information granules established at a certain level of abstraction information granules emerge as a sound conceptual and algorithmic vehicle owing to their way of delivering a more general view at data, ignoring irrelevant details and supporting a suitable level of abstraction aligned with the nature of the problem at hand. It is worth exploring granular computing as a contributor to the methodology of ML. While some investigations and promising results have been reported, there is a visible need to explore and establish main directions as well as build a coherent environment to cope with the above-stated challenges.

The objective of this study is to deliver a general overview of granular computing, identify the main pursuits forming a comprehensive research agenda and identify its far-reaching and impactful and synergistic role in addressing vital challenges in ML and computational intelligence. We aim to foster new and far-reaching directions of ML in forming a synergistic granular computing-ML platform of or granular ML for brief. There are a number of key objectives which associate with a significant level of originality.

1) Development a conceptual and algorithmic framework of ML augmented by the principles of granular computing (in particular, by forming a cohesive way of constructing granules through the principle of justifiable granularity and augmenting models through the mechanisms of granular embedding).
2) Conceptualization of data and knowledge facets in the construction of ML models.
3) Identification of crucial roles of information granules and specification of levels of information granularity required to realize a unified framework of data-knowledge environment of ML.

ML and granular computing have been developing independently. When some challenges of ML started to surface more vividly, in particular those concerning interpretability and transparency, advocated under the banner of explainable AI (XAI), it becomes noticeable that concepts of information granularity and information granules (especially fuzzy sets and rough sets) could support the agenda of XAI. Granular computing revolves around interpretability which has been for a long time vigorously emphasized in fuzzy rule-based models. The quality of ML models becomes also expressed in terms of nonnumeric outcomes and quantified in terms of granular results, such as intervals or fuzzy sets. It is likely that considering the recently growing interest in knowledge-data ML, the role of granular computing in these pursuits of ML will attract attention and trigger further investigations.

Serving as a position paper, this study identifies focal points and highlights future potential directions. To structure our discussion in a coherent fashion and highlight the main trends as well as deliver a self-contained material, the study is structured in a top-down manner. Some introductory material offering some motivating insights and main concepts and

Received 28 August 2024; revised 15 October 2024; accepted 23 October 2024. Date of publication 18 November 2024; date of current version 20 December 2024. This article was recommended by Associate Editor F. J. Cabrerizo.

The author is with the Faculty of Automatic Control, Electronics and Computer Science, Division of Measurements and Control Systems, Silesian University of Technology, 44-100 Gliwice, Poland, also with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6R 2V4, Canada, and also with the Faculty of Engineering and Natural Sciences, Department of Computer Engineering, Istinye University, 34010 Sariyer/Istanbul, Türkiye (e-mail: wpedrycz@ualberta.ca).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2024.3487934.

Digital Object Identifier 10.1109/TCYB.2024.3487934

objectives of granular computing are presented in Section II. The formal frameworks of information granules are covered in Section III. Section IV is devoted to the design of information granules carried out with the use of the principle of justifiable granularity. We also show linkages of this principle with unsupervised learning and clarify a role of clusters as data-driven information granules. The key conceptual and design challenges emerging in ML are highlighted in Section V. We also offer some useful insights as to the roles of information granules in addressing these quests. A new and emerging direction of unified knowledge-data ML is identified in Section VI. A knowledge-data integration aiming at the development of ML models is covered in Section VII. The realization of such integration process is discussed for rule-based models. Main conclusions are covered in Section IX.

II. Granular Computing—Principles, Information Granules, and Information Granularity

Granular computing is about information granules. More specifically, it offers a coherent environment aimed at 1) formalizing information granules via one among existing formal settings (which have been developed independently of each other and include sets, intervals, fuzzy sets, rough sets, probabilistic information granules, among others); 2) eliciting information granules; 3) establishing processing mechanisms; and 4) delivering interpretation of results and communicating obtained findings carried out at levels of abstraction. As a discipline, granular computing constitutes a well-rounded body of knowledge and establishes a collection of principles that apply across the diversity of the existing formal frameworks. Humans perceive the world, reason, and communicate at some level of abstraction. Abstraction comes hand in hand with *nonnumeric* constructs.

Information granules are intuitively appealing constructs, which play a paramount role in human cognitive and decision-making activities [1], [2], [3], [4]. We perceive complex phenomena by organizing existing knowledge along with any available experimental evidence and structuring them into a collection of some meaningful, semantically sound entities. Those become central to ensuing activities of describing the world, reasoning about the environment, classification, prediction, and decision-making methods. Zadeh [4] coined an informal, yet highly descriptive and compelling concept of information granules. Generally, by information granules one regards a collection of elements drawn together by their closeness (resemblance, proximity, functionality, etc.) articulated in terms of some useful spatial, temporal, or functional relationships. The concept of information granules is omnipresent, and this becomes well documented through a series of applications, one may refer to some selected studies [5], [6], [7], [8], [9], [10], [11], [12], [13], [14].

No matter which problem is taken into consideration, we usually position it in a certain conceptual framework composed of some generic and conceptually meaningful entities - information granules. They are relevant to the problem formulation, further problem solving, and a way in which the findings are communicated to the external world and recipients of the solution. Information granules realize a framework in which we formulate generic concepts by adopting a certain level of generality. In this sense, information granules are examples of abstractions. As such, they naturally give rise to hierarchical structures: the same data, problem or system can be perceived at different levels of specificity (detail) depending upon the complexity of the problem, available computing resources, and particular needs to be addressed.

Granular computing fully acknowledges a notion of variable granularity, whose range could cover detailed numeric entities and very abstract and general information granules. It looks at the aspects of compatibility of such information granules and ensuing communication mechanisms of the granular worlds. Granular computing gives rise to processing that is less time demanding than the one required when dealing with detailed numeric processing.

III. Frameworks of Information Granules

There are numerous formal frameworks of information granules; for illustrative purposes, we recall and contrast some commonly considered alternatives by looking at the underlying concepts and associated abilities to cope with the real-world problems.

A. *Diversity of Formal Settings of Information Granules*

The landscape of formal environments of information granules is remarkably diversified. Numerous alternatives fall under the realm of granular computing. There are convincing arguments motivating their conceptualization. We also recall some systematic ways of elicitation of the granules themselves.

Sets realize a concept of abstraction by introducing a notion of dichotomy. We admit that a certain element either belongs to a given information granule or is excluded from it. The notion of interval information granules falls under the same rubric whereas these constructs are defined over the space of real numbers leading to discipline of interval analysis (interval mathematics) [15], [16], [17].

Fuzzy sets deliver an important conceptual and algorithmic generalization of sets [3], [18], [19] by admitting a notion of partial membership. This quantification of membership becomes instrumental in a vast array of situations where the principle of dichotomy is neither justified. Furthermore, owing to the smooth (differentiable) nature of membership functions, fuzzy sets become useful in realizing optimization tasks in which they are present. This is of particular usage when considering gradient-based optimization schemes.

Shadowed sets [20] deliver an interesting and practically sound description of information granules by distinguishing among three categories of elements of the universe of discourse. Those are the elements, which 1) fully belong to the concept; 2) are excluded from it; and 3) their belongingness is completely *unknown* [21], [22].

The idea of *hesitant fuzzy sets* [23], [24] along their generalizations admit nonnumeric membership grades. A hesitant fuzzy set is defined by membership function A which for any element in the space produces a subinterval of [0,1]. In this sense, a shadowed set can exhibit some resemblance to this construct with a visible difference coming from the conceptualization of the construct and its construct.

Rough sets [25], [26] are concerned with a roughness phenomenon, which arises when an object (pattern) is described in terms of a limited vocabulary of certain granularity. The description of such nature gives rise to so-called lower and upper bounds of concepts forming the essence of a rough set.

Information granules with concept of membership and nonmembership A certain class of useful information granules is aimed at the representation of concepts of membership and nonmembership. Intuitionistic fuzzy set A [27], [28] is described by a pair of membership $A(x)$ and nonmembership functions $A^{\sim}(x)$ with the constraint satisfied for each element x of the space stating that $A(x) + A^{\sim}(x) \leq 1$. The byproduct is a hesitation index 1-$A(x)$-$A^{\sim}(x)$. The modifications come in the form of the Pythagorean fuzzy sets in which one has $A^2(x) + A^{\sim 2}(x) \leq 1$ [29] and Fermatean fuzzy sets with the constrain on the membership and nonmembership expressed as $A^3(x) + A^{\sim 3}(x) \leq 1$ [30].

The landscape of formal frameworks is quite diversified; as interesting examples, one can recall here probabilistic sets, a slew of probabilistic-fuzzy hybrid models of information granules, and axiomatic fuzzy sets.

At the practical end, one should emphasize that while the existing approaches come with some conceptual motivation pointing at their relevance, it is of paramount importance to have them equipped with sound development mechanisms and estimation procedures; in several cases this is not the case. Design processes along with the evaluation mechanisms becomes of critical relevance implying successful deployments.

B. Information Granules of Higher Type and Higher Order

From the structural perspective, information granules are generalized along two directions giving rise to higher order and higher-type information granules. Higher-type granules are granules whose membership grades of individual elements are information granules themselves. The well-known and intensively studied constructs are type-2 and interval-valued fuzzy sets. In this case, the membership grades are fuzzy sets defined over [0, 1] or intervals contained in the unit interval. Likewise, intervals generalize to granular intervals where the bounds are information granules, such as fuzzy sets. The membership grades could be of probabilistic nature and described by probability density functions defined over [0, 1]; this construct and its generalizations fall under the umbrella of probabilistic sets. This category of generalizations is well aligned with the real-world phenomena where one admits that the precision of numeric membership grades is neither feasible nor attainable.

Type-2 information granules are generalized to type-n granules where the membership grades are expressed recursively on a basis of information granules formed at the lower level of hierarchy. It is worth noting that type-0 information granules are just numbers. Conceptually, this generalization looks promising however to take full advantage of such information granules one needs to establish a systematic way of elicitation of granular membership grades.

Order-2 and order-n, in general, information granules associate with the generalizations of spaces over which granules are defined. In contrast to order-1 information granules, order-2 information granules are used to describe hierarchically complex concepts based on a collection of reference information granules. For instance, the term *comfortable* temperature is described over the family of terms {*low* temperature, *average* temperature, *high* temperature, *very high* temperature} producing the vector of membership grades [0.4/*low* temperature 0.9/*average* temperature 0.5/*high* temperature 0.3/*very high* temperature].

IV. Design of Information Granules and Numeric-Granular Compatibility

A task of building information granules constitutes a central item on the agenda of granular computing with far-reaching implications on its applications.

A. Principle of Justifiable Granularity

As noted, information granules are designed through the process of abstracting available experimental evidence. Let us start with an illustrative example. Consider a collection of numbers representing readings of temperature over the past month. By eyeballing them and without any computing, we conclude that we witnessed high temperature. What has been done, we build an information granule by arranging the existing data in a form of a single descriptor located at the higher level of abstraction. The principle of justifiable granularity guides a construction of an information granule based on available experimental evidence [11]. The resulting information granule becomes a summarization of data (viz. the available experimental evidence).

This principle mimics this process by admitting that the constructed information granule is meaningful (semantically sound) and at the same time well justified in light of the existing data. Formally, these two intuitively appealing criteria are expressed by the coverage and specificity measures. The first one expresses an extent to which the granule "covers" (includes) the data, viz. it is supported by available experimental evidence. The specificity measure quantifies an extent to which the granule is specific (precise) and stresses semantics (meaning) of the granule.

In case of 1-D numeric data $\{x_1, x_2, \ldots, x_N\}$ as witnessed in the above example and an information granule given in the form of a numeric interval $A = [a, b]$, these two criteria are expressed as $cov(A) = card\{x_k | \; x_k \in [a, b])\}$ and $spec(A) = g(length(A))$ with g standing for the monotonically decreasing function, $length(A) = b\text{-}a$, respectively. The criteria are in conflict; with the objective to maximize both of them, the optimization problem could be posed as the maximization of the product of coverage and specificity completed with respect to the bounds a and b, namely, arg $max_{a,b}$ $[cov(A)spec(A)]$.

The principle can be extended to multidimensional data and as a result of optimization, we arrive at the optimal parameters of the granule describing its geometry; for instance $A(\boldsymbol{x}; \boldsymbol{m})$ with $\boldsymbol{m}$ denoting the parameters (geometry) of the granule. In the above case, $\boldsymbol{m} = [a, b]$.

Based on the above stated example, two design strategies are encountered: 1) both parameters (*a, b*) are optimized, and 2) a two-phase process where one starts with a numeric representative of the numeric data (that could be mean, median, modal) and then optimizes a and *b* separately by maximizing the product of coverage and specificity where the data lower or larger than the numeric representative are involved in the calculations. This makes up two optimization tasks that are 1-D.

The generic design process can be extended in different ways. In general, the principle can be developed and applied to the construction of information granules expressed in different formal settings. Furthermore, interesting considerations involve the design scenarios where information contents are also involved that stems from the ML context: 1) data belonging to different classes and the design is intended to build the granule of data belonging to a single class (homogeneous granule); in this construction, inhibitory information is considered; 2) building granule in the presence of weighted data; and 3) building granules of the lowest variability with respect to some auxiliary variable. In these cases, not only geometric characteristics are formed but the information content I becomes quantified in the obtained granule $A(\boldsymbol{x}; \boldsymbol{m}, I)$. Several representative design scenarios are listed in Appendix.

The principle is also useful in expressing equivalency between two information granules being expressed in the presence of available experimental data. An information granule A is equivalent to B in terms of the principle when the equality $cov(A)spec(A) = cov(B)spec(B)$ is satisfied. This helps us determine the parameters of A. Likewise, one can build an information granule expressed in different formalisms, such as building a fuzzy set equivalent with probabilistic information granule B. If some underlying probabilistic characteristics of data are available, the above requirement can be read as $cov(A)spec(A) = {}^{p}cov(B)spec(B)$ with p denoting the underlying probability density function.

Probabilistic information granules can be converted into intervals or fuzzy sets by using the principle of justifiable granularity. As the probabilistic information granules are commonly governed by normal distribution, the calculations are carried out as follows. Let $N(0, \sigma)$ denote a Gaussian probability function $p(x)$ with the zero mean and standard deviation σ. Assume T is an interval distributed around zero with a spread d to be optimized. The coverage is computed in the form $cov(T) = \int_{-d}^{d} p(x)dx$, whereas the specificity $sp(T)$ is $1-([2d]/[6\sigma]) = 1-(d/[3\sigma])$. d_{opt} is produced by maximizing the product of coverage and specificity; $d_{opt} = \max_d[cov(T)sp(T)]$. In the same way, one obtains the parameters of the fuzzy set. The corresponding coverage and specificity are computed as $cov(T) = \int_{-d}^{d} T(x)p(x)dx$ and $sp(T) = \int_0^1 1-(|b_\alpha - a_\alpha|/[6\sigma])dx$, where a_α and b_α are the lower and upper bound of the α-cut for fuzzy set $T(x)$. The optimal spread d for interval and triangular fuzzy set could be determined by maximizing the product of coverage and specificity. Through straightforward computation, the relationship between the spread d and standard deviation σ for interval granule is $d = 1.15715\sigma$. For a triangular fuzzy set, one has $d = 1.64395\sigma$.

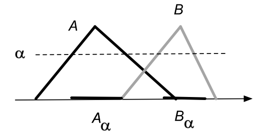

Fig. 1. Computing distance between fuzzy sets A and B based on the use of a family of α-cuts.

The principle of justifiable granularity highlights an important facet of elevation of the type of information granularity: the result of capturing a number of pieces of numeric experimental evidence comes as a single abstract entity-information granule. As numeric data can be thought as information granule of type-0, the result of their concise description becomes a single information granule of type-1. This is an interesting phenomenon of elevation of the type of information granularity. The increased level of abstraction is a direct consequence of the diversity present in the originally available granules. This elevation effect is of a general nature and can be emphasized by stating that when dealing with experimental evidence composed of a family of information granules of type-n, the result becomes a single information granule of type $(n + 1)$. Thus, we can express this finding in the form: a number of data of type-n are abstracted to a single type-$(n + 1)$ information granule.

For instance, numeric data (which are type-0) give rise to a single information granule (type-1), type-1 data (e.g., fuzzy sets) give rise to type-2 information granule (viz. type-2 fuzzy set), ..., etc.

To Recap: as a way of constructing information granules, the principle of justifiable granularity exhibits a significant level of generality in two essential ways. First, given the underlying requirements of coverage and specificity, different formalisms of information granules can be engaged. Second, experimental evidence could be expressed as information granules articulated in different formalisms and on this basis certain information granule is being formed.

Clustering methods deliver a sound way of building information granules. Depending on the way clustering method is formulated, set-based or fuzzy information granules are built. In many cases, like those involving K-means [31] or fuzzy c-means (FCMs) [32], information granules are formed around the obtained prototypes using the principle of justifiable granularity. Some discussion is covered in [33]. This two-phase procedure helps facilitate the design process of information granules. Refer also to detailed algorithmic considerations covered in Appendix.

An important concept studied in granular computing and used in its application concerns defining a distance between information granules expressed in $\boldsymbol{R}$. An example of computing the distance for fuzzy sets in case of a 1-D space is illustrated in Fig. 1.

Following [34], the distance between the intervals $[a,b]$ and $[c, d]$ is computed by taking the lower and upper bounds, $(a\text{-}c)^2 + (b\text{-}d)^2$. For fuzzy sets, see Fig. 1, the construct

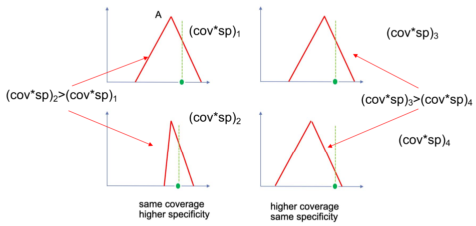

Fig. 2. Distribution of numeric datum and information granule visualized in the coverage-specificity space.

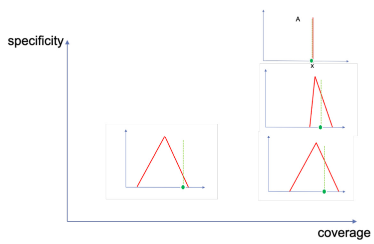

Fig. 3. Optimization of compatibility—A display in the coverage-specificity space; and the highest value is achieved when A and x are the same numeric entities.

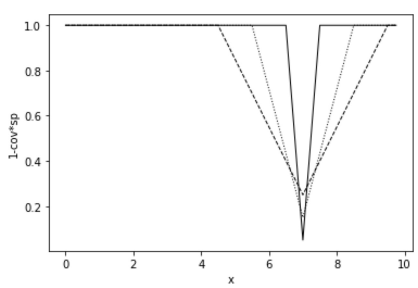

Fig. 4. Plot of expression 1-*cov*sp* shown for selected values of the spread of the triangular fuzzy set $A = T(x;\ 7\text{-}s,\ 7,\ 7 + s)$, $s = 0.5$ (solid line), 1.5 (dotted line), and 2.5 (dashed line) with x moving along [0,10].

is realized by taking α-cuts A_α and B_α and integrating the distances $d(A_\alpha, B_\alpha) = (a_\alpha - c_\alpha)^2 + (b_\alpha - d_\alpha)^2$ over the threshold values, $d(A, B) = \int_0^1 d(A_\alpha, B_\alpha)d\alpha$.

B. Compatibility of Numeric Evidence and Information Granule

The use of the coverage and specificity criteria provide an interesting insight into a way in which a location of information granule and numeric datum is characterized as illustrated in Fig. 2. The numeric datum and information granule (in this case we take a triangular fizzy set) show the highest level of compatibility (agreement) when both coverage and specificity are made the highest, namely, the product of these two measures attains its maximal value, see Fig. 3.

The visualization of the compatibility between a triangular information granule A with a modal value and bounds and numeric value sliding across the input space [0, 10] with the specificity regarded as a linearly decreasing function (with $range = 10$) is shown in Fig. 4. Here, plots for As with several values of spread s are covered.

As will be discussed in Section VI-F, the derivative of the expression 1-*cov*sp* will be used in the computing the loss function. For the higher values of a, the plot shows broader ranges of the linearly changing sections of the expression. This becomes instrumental in the optimization of the loss function with the usage of the gradient-based methods.

V. Challenges of Machine Learning and Perspective of Granular Computing

With the rapid advances in ML, it becomes apparent that there are a number of evident challenges. Addressing them is required to further foster the progress in the area and expand a spectrum of application areas. We identify several of them and point out that information granules offer some interesting, conceptually and algorithmically viable solutions.

A. Confidence/Credibility of ML Results

As data are central to all ML models, the quality produced by the models are directly implied by the location of a new situation vis-à-vis data being used for training purposes. Obviously, going beyond some boundaries of acquired "expertise" of the model associated with substantial risks that have to be assessed. Questions emerge: given $\boldsymbol{x}$, what confidence/credibility level could one associated with the produced result? Is the result credible? Could the result trigger some legitimate action or the system so that actionability becomes secured.

Information granularity plays a pivotal role with this regard. Instead of a numeric result produced by the model, information granule produced by the ML can offer a useful and quantifiable solution to the problem through raising awareness as to the possible need to acquire more evidence before proceeding with making any sound decision. The specificity of the output information granule can serve as a sound measure describing the quality of the result.

Different formal approaches could be envisioned here. We recall the two of them.

1) Granular Embedding: Numeric parameters, $\boldsymbol{a}$, of the ML model $M(\boldsymbol{x}; \boldsymbol{a})$ are elevated to their granular counterparts by forming information granules distributed around them. The transformation G, $\boldsymbol{A} = G(\boldsymbol{a}, \varepsilon)$ denotes the mapping of $\boldsymbol{a} \in R^p$ (with p standing for the number of parameters of the model) to a p-dimensional vector of information granules $\boldsymbol{A}$. For interval embedding, a numeric parameter a is transformed into an interval spanned over a, $[\min(a(1\text{-}\varepsilon), a(1 + \varepsilon)), \max(a(1\text{-}\varepsilon), a(1 + \varepsilon))]$, $\varepsilon \in [0, 1]$. The hyperparameter ε denotes a level of information granularity and its value is optimized through the maximization of the product of coverage and specificity of results of the model $M(\boldsymbol{x}; G(\boldsymbol{a},\varepsilon))$ with the coverage being computed for the data $\mathcal{D}$ used to learn the numeric model $M(\boldsymbol{x}; \boldsymbol{a})$. Granular embedding can be applied to input data. To describe their granularity (as encountered in regression with uncertain data), instead of numeric $\boldsymbol{x}$, we admit its granular counterpart $\boldsymbol{X}$ associated with some level of information granularity δ, $\boldsymbol{X} = G(\boldsymbol{x}, \delta)$. Then we have $M(G(\boldsymbol{x}, \delta), \boldsymbol{a})$. Also, granular embedding can be completed at the parameter and data side resulting in $M(G(\boldsymbol{x}, \delta), G(\boldsymbol{a},\varepsilon))$.

2) Gaussian Process: model [35], [36] produces results that come in the form of probabilistic information granules described by Gaussian distributions. The process GP($\boldsymbol{m}$, $\kappa(\boldsymbol{x}, \boldsymbol{x}')$) is characterized by some mean function $\boldsymbol{m}$ and covariance function $\kappa(\boldsymbol{x}, \boldsymbol{x}')$. For any $\boldsymbol{x}$, the GP model $GP(y| \boldsymbol{x}, \mathcal{D})$ gives rise to the corresponding output viewed in the form of a Gaussian distribution, $N(m, \sigma)$ inferred on a basis of training data $\mathcal{D}$ where the standard deviation σ describes the level of granularity of the obtained result for given input $\boldsymbol{x}$. Several design alternatives are sought with regard to the selection of kernels realizing the covariance function. A two-phase design process is invoked: a numeric model is built first and then a representative subset of training data is used to construct the GP model thus augmenting the numeric results with the probabilistic granules.

While the formats of information granules arising from 1) and 2) are different, the results could be treated uniformly by engaging the principle of justifiable granularity. Another general observation, which is self-explanatory, is worth making here: there are no ideal ML models. While type-n information granules are used to build the model, to be in rapport with reality, the results produced by the model have to be type-($n + 1$) information granules. In granular embedding, we advocate granular format of parameters (assuming numeric data and optimal numeric parameters) which entails granular results. In the GP model, eventually built on a basis of a numeric model, the results arise in the form of probabilistic information granules.

The above augmentations can endow the ML design methodology with a useful self-awareness mechanism, where the quality (granularity) of the result not only quantifies the confidence (and eventually offer some visualization mechanism) but helps the model to become aware of its limitations and potentially invoke further learning through the mechanisms of active learning [38], [39], [40].

B. Interpretability and Explainability

These two features of ML architectures [41], [42] is one of the crucial requirements well-articulated in XAI. Information granules are conceptually and algorithmically viable constructs as lucidly demonstrated in [43]. Concepts described by fuzzy sets, rough sets, relationships in the form of rule-based models and associations have been intensively studied in the literature.

In fuzzy rule-based models, information granularity establishes a suitable level of abstraction (depending upon a particular application and the needs of the recipient of the ML model). More rules coming with more specific condition and conclusion parts capture more details and potentially produce higher accuracy (at the numeric level), however the interpretability and interpretability become adversely impacted. One usually admits that any quantification of interpretability (which is multifaceted) includes in its formulation the number of rules (and this could be controlled by the number of clusters produced by the commonly used FCM clustering algorithm). Its interesting aspect of interpretability is a stability of rules. Intuitively, as knowledge conveyed (or elicited) by the rules should not be volatile, we anticipate that the rules observed at a level of information granules and expressing associations among input and output variables (for instance, causes and effects) are stable so they do not change their semantics in the presence of some variations in the data. The rule "if x is small then y is medium" remains the same even data change slightly. Obviously, for such distorted data, the linguistic terms (small, medium) could come with varying membership functions (different parameter values) but the related meaning (conveyed by the symbolic contents of the terms) remains the same. Here, we capitalize on the duality of information granules, that could be treated either as symbols or numeric constructs with the underlying parameters.

VI. Unified Knowledge-Data Environment of Design of ML

Data are a holy grail of ML. They are behind the undisputable successes of the discipline. They also come with challenges that are imminent, in particular in light of the diversity and the nature of data and their volume. The key observations concern some essential characteristics of the multitude of ML constructs, their design and deployment (and very likely the ones emerging in the future): 1) in virtue of the underlying complexity and distributed architectures (for example, consider deep learning networks) are black boxes; 2) going beyond the boundaries of data used to build the ML architectures (built on the principle of inductive reasoning) brings a question of credibility/confidence/relevance of the produced results (which becomes of central importance when deploying the models in critical environments requiring a high level of autonomy, consider autonomous vehicles); 3) learning is commonly realized from scratch, which brings a high level of computing overhead; 4) loss functions emphasize the quality expressed in terms of prediction or classification accuracy; and 5) any possible data attack impacts the construction and the functioning of the model. In light of the arguments put forward and to emphasize this point on reliance on data in the ML design methodology, we can refer to the resulting mapping $\mathcal{D} \rightarrow M_{\mathcal{D}}$ with the subscript emphasizing the origin of the ML model $M_{\mathcal{D}}$.

A. Knowledge

In every real-world problem, there is some knowledge that as of now has not been accommodated as a part of the design environment of ML architectures. Going beyond the data in the development of ML models, sounds attractive and deserves careful attention. One may anticipate far reaching ramifications of this conceptualization. The main direction emerging in this way could be referred to as knowledge-data ML or KD-ML in brief. In essence, denoting the knowledge as K, the resulting construct could be expressed as M_{DK}. There are different ways of representing knowledge and using it in the realization of the model. Quite commonly, one can envision that the design of M_{DK} is guided by a general (extended) loss function whose important component is a regularization term reflecting the knowledge conveyed in the problem.

For the sake of completeness, one could also remark that the standard data-oriented practice of ML is to augment the loss function by admitting a so-called regularization term. Its role is to avoid possible memorization effect by controlling the nature of the parameters of the constructed model. In the context of knowledge involved in the design, this term is referred to as a knowledge regularization component.

B. Taxonomy

When pursuing the agenda of KD-ML, the two fundamental questions that need to be posed and carefully addressed concern the origin and taxonomy of knowledge as well as its suitable representation.

When it comes to the origin and taxonomy of knowledge, there are several key categories encountered [44].

Scientific knowledge This knowledge is articulated through universal laws of physics, chemistry and invariants of mass conservation, moment conservation, etc. For instance, one can refer to Newton's law of motion, Maxwell's law of electromagnetics and conservation laws (conservation of mass, moment, energy, ...).

Interestingly, the considerations developed within this realm have led to the recent developments of physics-informed ML [45], [46], [47], [48].

Commonsense knowledge: Two main categories are distinguished here.

1) *World Knowledge:* This category of knowledge comprises everyday life facts; it is intuitive and validated by human reasoning (subsumes linguistics) and validated through empirical studies; various levels of abstraction (information granules) are present.
2) *Expert Knowledge Again:* This represents common knowledge being held by a particular group of experts. Various levels of abstraction (specificity of information granules) are also considered.

The theme of knowledge representation has been central on the agenda of AI. There have been fundamental discussions on the symbolic and connectionist perspective of knowledge representation with the associated benefits and limitations, as convincingly advocated by Minsky [49]. Interestingly, he made compelling arguments on the role of heuristics being played in the representation schemes .... Our purely numeric connectionist networks are inherently deficient in abilities to reason well; or purely symbolic logical systems are inherently deficient in abilities to represent the all-important heuristic connections between things–the uncertain approximate or analogical links ....

C. Knowledge Representation

There are numerous and highly diversified formal ways aimed at knowledge representation including the following.

1) Algebraic equations, differential equations, simulation results, and spatial invariances (translations and rotations). These fall under the umbrella of ways in which scientific knowledge is being represented. They serve as vehicles to support physics-informed ML.
2) Logic rules and rule-based models, knowledge graphs, relations and relational calculus, semantic networks, and frames with default assignments. These are the schemes proposed and intensively studied in the area of AI.

D. Knowledge Elicitation

Once some way of knowledge representation has been decided upon, it must be elicited. In the area of AI, there have been intensive studies [50], [51] on ways how knowledge can be acquired and its quality evaluated. The key features of acquired knowledge determining its relevance involve completeness and consistency, among others. With the increasing level of complexity and dimensionality of the problem, any human way of knowledge acquisition is faced with a phenomenon of knowledge acquisition bottleneck [52]. As a follow-up, there is no guarantee that the knowledge guidance is ideal and as such its contribution to knowledge-data ML is instrumental but cannot be regarded as the only player in this design environment.

E. Duality of Information Granules

Information granules exhibit an important feature of duality. They can be sought as symbols (S, M, L, ..., etc.) and as such are subject to symbol processing (for instance, governed by grammar rules). On the other hand, they are grounded and their numeric manifestation is subject to number-oriented processing. If S, M, and L are linguistic terms, we may have rules of symbolic processing, such as S + M = VL (very large), etc. but once they are described by *numeric* membership functions, operations on them yield some numeric results.

The key point worth making is that knowledge is inherently represented at the higher than numeric level of abstraction. Hence, the role of information granules becomes inherently visible and their usage in KD-ML turn out to be apparent. Essential is the crucial *symbolic* and *numeric* duality of information granules.

There are two main levels of knowledge integration in knowledge-data ML, namely, knowledge integration completed at the data level and the integration carried out at the level of architectures of ML models.

F. Learning

In the presence of data D, the commonly encountered loss function used in the construction of M (or more specifically M_D) assumes the following form:

$$L = \sum_{D} (M(x_k, w) - t_k)^2 + \mu||w||. \tag{1}$$

where the second component is a regularization term whose contribution to L is guided by the non-negative weight μ. The norm ||.|| is either L_0, L_1, or L_2. By analogy, the inclusion of knowledge in the loss function is completed in an additive form

$$L = \sum_{D} (M(x_k, w) - t_k)^2 + \mu \sum_{D} RK \tag{2}$$

where $R(K)$ denotes the knowledge-oriented regularization. μ is a non-negative hyperparameter that strikes a balance between the impact of data and knowledge. Its value is optimized on a basis of the validation set.

The guidance provided by the knowledge model M_K provides information granule in the design process. Thus, if for some $\boldsymbol{x}$ the model M_D returns $M_D(\boldsymbol{x})$ at the same time M_K returns granule $Y = M_K(\boldsymbol{x})$. The compatibility is expressed as a product of coverage and specificity and its derivative becomes used in the gradient-based optimization of L. In this sense, the use of the smooth non-Boolean membership function (information granule) because advantageous to improve the efficiency of the learning process. The detailed calculations of the derivative of the knowledge regularization term computed with respect to the i^{th} component of $\boldsymbol{w}$, w_i are expressed as

$$\frac{d}{dw_i}(1-cov(M(\boldsymbol{x};\boldsymbol{w}), Y)sp(Y)) = -spY\frac{d}{dw_i}(cov(M(\boldsymbol{x};\boldsymbol{w})) = \\ = -sp(Y)\frac{dcovM(\boldsymbol{x};\boldsymbol{w})}{d(M\boldsymbol{x};\boldsymbol{w})}\frac{d(M\boldsymbol{x};\boldsymbol{w})}{dw_i}. \quad (3)$$

VII. Knowledge and Data Integration in ML

A. Physics-Informed ML

Physics-informed ML has been intensively pursued in recent years. As a convincing testimony, one can refer to the recent collection of studies reported in [53]. There are several compelling arguments behind this direction of ML studies including 1) an ability to describe a physical phenomenon at a global level and 2) endowing the constructed ML model with locally available data driven model. In the design we engage some invariants, including laws of mass and momentum conservation.

The scientific knowledge is described by some relationships present in the system governed by equation g($\boldsymbol{x}$, y, $\boldsymbol{a}$) = 0 with $\boldsymbol{a}$ denoting a vector of parameters. The dynamic systems are described with the use of dynamics-oriented models, namely, differential and partial differential equations. Given the training data $\boldsymbol{D} = (\boldsymbol{x}_k, t_k)$, $k = 1, 2, \ldots N$ and D_* composed of collocation data, the loss function guiding the development of the ML model $M(\boldsymbol{x}; \boldsymbol{w})$ is additive consisting of two components

$$L = \lambda\sum_{D}(M(x_k;\boldsymbol{w}) - t_k)^2 + (1-\lambda)\sum_{D*}|g\boldsymbol{x}_k, M(\boldsymbol{x_k};\boldsymbol{w}), \boldsymbol{a}|. \quad (4)$$

The second term of L enforces M to fit the collocation points. The hyperparameters λ assuming values in [0,1] strikes a sound tradeoff between the usage of guidance stemming from data and knowledge sources. As $g(\boldsymbol{x},y, \boldsymbol{a})$ is commonly acceptable and comprehended by the community, the physics-informed ML promotes interpretability of the resulting model.

B. Augmentation of Data Driven Models by Knowledge-Oriented Constraints

The data blinding effect (viz. an exclusive reliance on data) can be reduced by accommodating knowledge -oriented constraints. The constraints can be sought as sources of knowledge describing the input-output relationships existing in the problem.

As an illustrative example, let us consider a single input-single output ML model in the form of a neural network

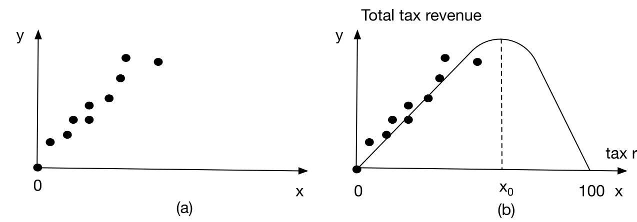

Fig. 5. One-dimensional data (a) and Laffer effect (b) providing knowledge constraints.

$NN(x; \boldsymbol{w})$ built on a basis of some data $D = \{x_1, x_2, \ldots, x_N\}$ as illustrated in Fig. 5. The constructed neural network may suggest a nonlinearly increasing input-output relationship and its usage for predicting higher values of the input may imply higher values of the output. The derived conclusion serves as a simple example of data blinding. This is a direct conclusion of the fact that the problem is being viewed as exclusively through the lenses of data and ignoring the commonsense knowledge. Before proceeding with the design of the neural network, note that are informed that the input variable is tax rate (in range 0–100) and the output is a tax revenue. It is apparent that the continuously increasing relationship could be simply misleading–indeed 100% tax rate does not offer any incentive. The relationship is known as a Laffer effect stating that a total revenue is not maximized with the increase of tax rates; and beyond some point the workers are disincentivized.

Thus, the data can be augmented by some semantically sound relationships. In this case, we express them by bringing three constraints in the form.

1) Including a boundary condition at 100% $NN(x; \boldsymbol{w})|_{x=100} = 0$.
2) Considering a single optimal value x_0 where $NN(x_0; \boldsymbol{w})$ attains maximum, namely, $dNN(x; \boldsymbol{w})/dx|_{x = x0} = 0$. This relationship captures a certain knowledge-implied constraint.
3) Requesting the maximum x_0 is located as far from $\max_{k=1,2,\ldots,N} x_k$; we argue that some maximal and yet optimal (the higher) value of revenue. The following function $\varphi(x)$ is introduced: it is increasing over $[\max_{k=1,2,\ldots,N} x_k, 100]$ and set to 0 otherwise.

The loss function L becomes formed as a linear combination whose structure is the same as in (2) involving the data-driven neural network and the satisfaction of the constraints specified above

$$L = \lambda\sum_{D}(NN(x_k;\boldsymbol{w}) - t_k)^2 + \quad (5) \\ (1-\lambda)\left[NN(100;\boldsymbol{w}) + \left|\frac{dNN(x;\boldsymbol{w})}{dx}\right|_{x=x_0} + \exp(-\phi(x_0))\right].$$

The minimization of the loss function is carried out with respect to λ, $\boldsymbol{w}$, and x_0.

C. Constraints as Information Granules

In general, we may formulate a general format of the design of the ML model M in the presence of knowledge hints articulated as some constraints and required local extreme describing the nonlinear characteristics of the model.

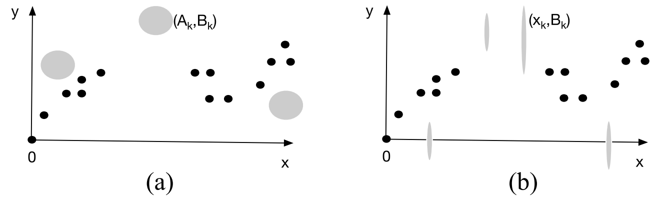

Fig. 6. Examples of data and information granules expressing knowledge-guided constraints: (a) (A_k, B_k) and (b) (x_k, B_k), $k = 1,2, \ldots, N^*$.

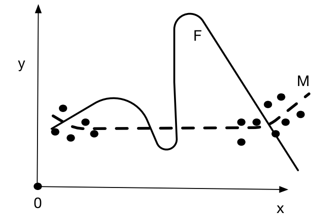

Fig. 7. Physics-oriented model F supporting the design of ML model M.

1) Boundary condition $M(x;\ \boldsymbol{w})|_{xk=xk\sim} = a_k$ where $(x_k \sim, t_k \sim)$, $k = 1, 2, \ldots, N\sim$ are specified constraints.
2) Local extreme expressed at some predetermined points $(x_k{*}, t_k{*})$, $k = 1, 2, \ldots N$, $dM(xk{*}, \boldsymbol{w})/dx_k{*} = 0$.

The knowledge hints as being expressed at higher level of abstraction could be provided in the form of information granules to capture the available level of precision. The constraints are expressed in the form $D^* = (A_k, B_k)$ where both A_k and B_k are information granules (intervals, fuzzy sets), $k = 1, 2, \ldots, N^*$. The plot of illustrative relationships is shown in Fig. 6.

The loss function is again additive by involving the data-based learning component and the one in which the knowledge hints are to be accommodated. Because of the granularity of the constraints, the key point is to determine the distance between the $Y_k = M(A_k; \boldsymbol{w})$ and B_k. As we are concerned with the two information granules, the calculations depend upon the formalism of information granules under considerations. If Y_k and B_k are intervals, the distance $\|M(A_k;\ \boldsymbol{w})\text{-}B_k\|$ computed by considering distances between the bounds of the intervals. If fuzzy sets are considered, their α-cuts are determined, namely, $Y_{k\alpha}$ and $B_{k\alpha}$ and the distances between intervals are integrated over [0,1]. The loss function reads as follows:

$$L = \lambda \sum_{D} (M(\boldsymbol{x}_k;\boldsymbol{w}) - t_k)^2 + (1-\lambda) \sum_{k=1}^{N*} ||M(A_k;\boldsymbol{w}) - B_k||. \quad (6)$$

The form of knowledge hints can assume a form of pairs (x_k, B_k) where the input is numeric. As before, the loss function is additive and consists of two components

$$L = \lambda \sum_{D} (M(\boldsymbol{x}_k;\boldsymbol{w}) - t_k)^2 + (1-\lambda) \sum_{k=1}^{N*} (1 - cov(M(\boldsymbol{x}_k;\boldsymbol{w}),\ B_k) sp(B_k)). \quad (7)$$

The second part realizes the matching between B_k and $M(x_k,\ \boldsymbol{w})$.

D. Physics-Oriented ML Models With Granular Expansion of the Knowledge-Driven Model

The real-world system we are interested to model is governed by the mapping f^*: $R \rightarrow [0,1]$. f^* is a complex system with eventually varying or stochastically implied parameters. The physics-driven model F of f^* is usually formed because of intensive research and equipped with some parameters estimated over data. Such models are deemed to capture the essentials of the system, and quite relevant to a broad range of the input variables. However, they are general and while applicable to a vast array of situations (regions of the input domain), might not be accurate for the actual situation for which some input-output data are available. The limited accuracy is observed globally. There is no guarantee that such data support the design of the ML model, however we may anticipate that its accuracy is high for the local regions implied by the location of the data. Fig. 7 illustrates this situation.

Thus, the data-oriented model M is designed with the use of D and its parameters are estimated leading to $M(\mathrm{x};\ \boldsymbol{w})$. While M interpolates outside the regions occupied by D, its results are not validated.

On the other hand, F is of global nature, however, not necessarily fits well the data as M does in a local way. To quantify this effect, we complete granular embedding by admitting agranular nature of its parameters $\boldsymbol{a}$, $\boldsymbol{A} = G_\varepsilon(\boldsymbol{a})$ which gives rise to $F(x,\ A_\varepsilon)$ with ε standing for the level of information granularity; this hyperparameter is optimized (maximized) on a basis of D.

The loss function consists of two parts

$$L = \lambda \sum_{D} (M(\boldsymbol{x}_k;\boldsymbol{w}) - t_k)^2 + (1-\lambda) \quad (8)$$
$$\sum_{k=1}^{N*} (1 - cov(M(\boldsymbol{x}_k;\boldsymbol{w}),\ F(\boldsymbol{x}_k, A_\varepsilon)) sp(F(\boldsymbol{x}_k, A_\varepsilon)).$$

The optimization is realized with respect to $\boldsymbol{w}$, ε, and λ assuming values in [0,1] where the minimization of L is accomplished in the presence of some validation set. One may look at this design scenario as resembling transfer learning and knowledge distillation [54], however there is an apparent difference. In this scheme, we admit that the performance of teacher is manifested in the form of information granules and because of the form of the second term of (8), its impact on the design of M varies from data to data, while this type of individual contribution of data is not available in the generic mechanism of transfer learning. Furthermore, in knowledge distillation discussed in the setting of classification problems, the hyperparameter of temperature softening the original logics is optimized and its role resemblances the one of the levels of information granularity.

VIII. Rule-Based Models in the Realization of Data and Knowledge Environment

Knowledge and data facet of ML supplement each other to eliminate potential gaps. Realistically, the data do not “cover” the entire space (as we can never account for all possible situations and collect pertinent data). Likewise, knowledge might not be complete–a problem of knowledge bottleneck

error \ change of error	NB	NS	Z	PS	PB
NB	P	P	P	P	Z
NS	P	P	P	Z	N
Z	P	P	Z	N	N
PS	P	Z	N	N	N
PB	Z	N	N	N	N

Fig. 8. Example set of control rules (fuzzy controller).

was flagged as one of the issues in knowledge-based systems; some facts and relationships (because of their evident relationships) might be overlooked quite often in the process of knowledge elicitation.

A. Qualitative (Symbolic) Models

Qualitative models [55] and symbolic processing [56] have been under intensive studies in the realm of AI, especially in the eighties, when there was a great deal of studies on knowledge representation. The key idea behind these models is to describe relationships among variables (inputs and outputs) at the qualitative (symbolic) level where the variables are quantified through a collection of symbolic (qualitative) landmarks. Qualitative reasoning about continuous domains requires quantization of the domain to a discrete and small number of symbols. The values being chosen reflect regions of qualitatively uniform component behavior, whereas interesting transitions occur at such boundary points.

Apparently, the qualitative models are more abstract than number-oriented constructs. They do not focus on numeric detailed aspects but rather tend to reveal qualitative dependencies and in this sense orient toward forming some interpretable relationships.

B. Rule-Based Architectures

Rule-based models assume a visible position in the realization of symbolic models. This interpretability aspect has been emphasized in the current trend of studies on XAI. The rules are conditional statements whose condition and conclusion parts are articulated with symbols. For instance, the rule reads as "if x is S and z is M then y is L" where the landmarks small (S), medium (M), zero (Z), and alike are positioned in the input and output space. A limited repertoire of landmarks could consist various symbols (qualitative values), such as $-$ (negative), Z (zero), and $+$ (positive) as discussed in qualitative models. The rules are elicited from domain expert. At the implementation end, a critical step is to ground the symbols, namely, endow them with sound semantics in the context of a problem at hand. The symbol grounding involves intervals and their variants, such as rough intervals or fuzzy sets.

Fuzzy controllers are positioned in the plethora of symbolic models (symbolic controllers), see Fig. 8. The constructs well known in the fuzzy control community are based on control rules that capture the essence of control strategies. The inputs are error and change of error while the output is control variable. In this particular case, the landmarks are implemented as fuzzy sets. The symbol grounding is completed by admitting some specific membership functions along with the parameters of the corresponding fuzzy sets. The table captures 25 rules which at the symbolic level are easily comprehended, for instance, "if error is negative big (NB) and change of error is Z (Zero) then control is P (Positive)." On the other hand, as each symbol is grounded through the corresponding membership function, the mechanisms of approximate reasoning applied here give rise to numeric values of the output (control).

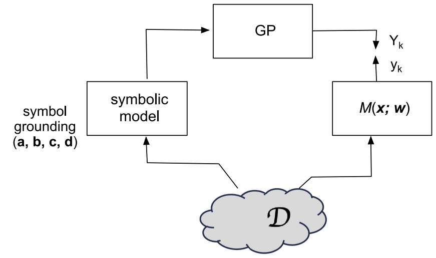

Fig. 9. Knowledge-data design environment with qualitative model.

Bringing data and knowledge conveyed by symbolic models requires a mindful process of model design in this heterogeneous environment. It is worth emphasizing that in light of the symbolic nature of the qualitative models (and the results produced there), a suitable synergistic mechanism is required as well as a way of building the model and its validation.

An overall way of building ML models in the presence of symbolic model and data is displayed in Fig. 9.

On the one hand, as usual, the construction of the ML model is guided by available numeric data $D = (x_k, t_k)$, $k = 1,2, \ldots, N$. On the other, there is some useful guidance delivered by the symbolic model. In virtue of the qualitative nature of this model, they are conveyed in the form of some information granule which is used directly in the optimization of the augmented loss function.

Proceeding with the details, we start by symbol grounding at the side of the symbolic model. Consider that the symbols are mapped on some ordinal scale viz. the symbols are associated with the numeric values. For instance, the symbols NL (negative *Large*), NM (negative *Medium*), ..., defined over variable of speed are organized with some order imposed NL $\prec$ NM $\prec \ldots$ and with each of them we associate the numeric entries $a_1, a_2, a_3, \ldots,$ organized in increasing order $a_1 < a_2 < \ldots$ Without loss of generality, we confine these entries to the unit interval so $\boldsymbol{a} = [a_1, a_2, \ldots, a_n]$. They are sought because of grounding the symbols. Furthermore, assume that the values of a_is are distributed over the range of values the corresponding variable in D takes on.

We complete the same embedding mechanism for other input variables, namely, $\{b_i\}$, $\{c_l\}$, and the output variable $\{d_j\}$. These numeric values are structured into a vector form giving rise to $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{d}$. Next, they are concatenated into a single vector $\boldsymbol{A} = [\boldsymbol{a}\ |\boldsymbol{b}|\ \boldsymbol{c}|\ \boldsymbol{d}]$ whose entries are subject to optimization as discussed in a while. In this sense $\boldsymbol{A}$ can be sought as the knowledge conveyed by the symbolic model $D_K = \boldsymbol{A}$. Using $\boldsymbol{A}$, we build a GP regression. Once done, for any x_k coming from D, the knowledge-based model produces a probabilistic Gaussian information granule with the corresponding mean value and the standard deviation, viz. we obtain GP($m(x_k)$,$\sigma(x_k)$ |$\boldsymbol{A}$). This granule is transformed into an equivalent interval or fuzzy set information granule following the optimized equivalence relationship discussed in Section IV-A.

Finally, for each x_k, the corresponding output of the knowledge-based model becomes Y_k. In the sequel, the loss function consists of the two terms

$$L = \lambda \sum_{D} ||M(\boldsymbol{x}_k, \boldsymbol{w}) - t_k|| + (1-\lambda) \sum_{D} (1 - cov(M(\boldsymbol{x}_k, \boldsymbol{w}), Y_k) sp(Y_k)). \tag{9}$$

where $\lambda \in [0, 1]$ is a hyperparameter with values in [0, 1] balancing the contribution of guidance coming from data D and D_K. As in the previous situations, the value of λ is optimized on a basis of the validation set. The minimization is completed with respect to the model coming from some class of models M and its parameters described by $\boldsymbol{w}$, $M(\boldsymbol{x}; \boldsymbol{w})$.

This minimization is also realized for the assumed values of the grounding parameters $\boldsymbol{A}$. They themselves could be optimized, namely, $\boldsymbol{A}_{opt} = \arg\min_A L$. In other words, there are two nested optimization loops-optimization with respect to $\boldsymbol{A}$

$$- \text{given } D_K \text{ (that is the GP model), optimize } M \in M \text{ and } \boldsymbol{w}. \tag{10}$$

While the optimization of (9) could be gradient-based, the nature of the optimization of D_K, some population-based optimization (PSO, ant colonies, differential evolution) could be considered.

IX. Conclusion

The study has raised and elaborated on the promising facets of synergy of ML and granular computing. By identifying the ongoing challenges of privacy, computational sustainability, confidence (credibility) and interpretability, we have stressed the importance of the key features of information granules aimed at the reformulation and conceptualization of the problems and building efficient solutions. It has been demonstrated that the new and promising and far-reaching direction of knowledge-data ML can be achieved by structuring a design framework and formulating coherent loss functions.

There is a broad and productive agenda of future pursuits located at the junction of ML and granular computing. An ultimate objective is to enhance the developments of ML by endowing its constructs by mechanisms of evaluation their performance by admitting inherent information granularity (of higher-type/higher order). This may raise awareness about the limit performance of the models and trigger further actions by engaging ways of its enhancements (e.g., through active learning). One can gain flexibility in establishing tradeoffs and identifying limitations among design criteria of accuracy, stability, interpretability by positioning the development at a suitable problem-implied level of abstraction. Further investigations worth pursuing could be focused on the industrial aspects of knowledge-oriented systems, including control systems [57].

References

[1] W. Pedrycz, *Granular Computing*. Boca Raton, FL, USA: CRC Press, 2013.

[2] W. Pedrycz, *An Introduction to Computing with Fuzzy Sets: Analysis, Design, and Applications*. Cham, Switzerland: Springer, 2020.

[3] W. Pedrycz, "Granular computing for data analytics: A manifesto of human-centric computing," *IEEE/CAA J. Automatica Sinica*, vol. 5, no. 6, pp. 1025–1034, Nov. 2018.

[4] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets Syst.*, vol. 90, pp. 111–127, Sep. 1997.

[5] A. Gacek and W. Pedrycz, "Clustering granular data and their characterization with information granules of higher type," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 4, pp. 850–860, Aug. 2015.

[6] L. Hu and K. C. C. Chan, "Fuzzy clustering in a complex network based on content relevance and link structures," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 2, pp. 456–470, Apr. 2016.

[7] X. Hu, Y. Shen, W. Pedrycz, X. Wang, A. Gacek, and B. Liu, "Identification of fuzzy rule-based models with collaborative fuzzy clustering," *IEEE Trans. Cybern.*, vol. 52, no. 7, pp. 6406–6419, Jul. 2022.

[8] X. Hu, W. Pedrycz, and X. Wang, "Granular fuzzy rule-based models: A study in a comprehensive evaluation and construction of fuzzy models," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 5, pp. 1342–1355, Oct. 2017.

[9] J. Leng, Q. Chen, N. Mao, and P. Jiang, "Combining granular computing technique with deep learning for service planning under social manufacturing contexts," *Knowl. Based Syst.*, vol. 143, pp. 295–306, Mar. 2018.

[10] W. Pedrycz, R. Al-Hmouz, A. Morfeq, and A. Balamash, "The design of free structure granular mappings: The use of the principle of justifiable granularity," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 2105–2113, Dec. 2013.

[11] W. Pedrycz and W. Homenda, "Building the fundamentals of granular computing: A principle of justifiable granularity," *Appl. Soft Comput.*, vol. 13, no. 10, pp. 4209–4218, 2013.

[12] W. Pedrycz, R. Al-Hmouz, A. S. Balamash, and A. Morfeq, "Hierarchical granular clustering: An emergence of information granules of higher type and higher order," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 6, pp. 2270–2283, Dec. 2015.

[13] W. Pedrycz, R. Al-Hmouz, A. Balamash, and A. Morfeq, "Granular autoencoders: Concepts and design," *Soft Comput.*, vol. 23, no. 20, pp. 9869–9880, 2020.

[14] W. Pedrycz, "Evaluating quality of models via prediction information granules," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 12, pp. 5551–5556, Dec. 2022.

[15] G. Alefeld and J. Herzberger, *Introduction to Interval Computations*. New York, NY, USA: Academic, 1983.

[16] R. Moore, *Interval Analysis*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1966.

[17] R. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*. Philadelphia, PA, USA: SIAM, 2009.

[18] H. Nguyen and E. Walker, *A First Course in Fuzzy Logic*. Boca Raton, FL, USA: Chapman Hall/CRC Press, 1999.

[19] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning-I, II, III," *Inf. Sci.*, vol. 8, pp. 199–251, 1975.

[20] W. Pedrycz, "Shadowed sets: Representing and processing fuzzy sets," *IEEE Trans. Syst., Man, and Cybern., Part B (Cybern.)*, vol. 28, no. 1, pp. 103–109, Feb. 1998.

[21] Y. Yao, S. Wang, and X. Deng, "Constructing shadowed sets and three-way approximations of fuzzy sets," *Inf. Sci.*, vols. 412–413, pp. 132–153, Oct. 2017.

[22] Y. Yao, "Three-way decision and granular computing," *Int. J. Approx. Reason.*, vol. 103, pp. 107–123, Sep. 2018.

[23] V. Torra, "Hesitant fuzzy sets," *Int. J. Intell. Syst.*, vol. 25, no. 6, pp. 529–539, 2010.

[24] Z. Xu, *Hesitant Fuzzy Sets Theory*. Cham, Switzerland: Springer, 2014.

[25] Z. Pawlak, "Rough sets," *Int. J. Inf. Comput. Sci.*, vol. 11, vol. 15, pp. 341–356, 1982.

[26] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning About Data*. Dordrecht, The Netherlands: Kluwer Acad. Publ., 1991.

[27] K.T. Atanassov, *On Intuitionistic Fuzzy Sets Theory*. Berlin, Germany: Springer, 2012.

[28] K.T. Atanassov, *Interval-Valued Intuitionistic Fuzzy Sets*. Cham, Switzerland: Springer, 2020.

[29] R. R. Yager, "Pythagorean membership grades in multicriteria decision making," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 4, pp. 958–965, Aug. 2014.

[30] T. Senapati and R. R. Yager, "Fermatean fuzzy sets," *J. Ambient Intell. Humaniz. Comput.*, vol. 11, no. 2, pp. 663–674, 2020.

[31] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010.

[32] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, NY, USA: Plenum, 1981.

[33] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 645–678, May 2005.

[34] O. Kosheleva and V. Kreinovich, "Euclidean distance between intervals is the only representation-invariant one," *Reliable Comput.*, vol. 28, no. 3, pp. 4–9, 2021.

[35] M. Lifshits, *Lectures on Gaussian Processes*. Cham, Switzerland: Springer, 2012.

[36] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[37] R. He, S. Liu, S. He, and K. Tang, "Multidomain active learning: Literature review and comparative study," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 7, no. 3, pp. 791–804, Jun. 2023.

[38] M. Liu, W. L. Buntine, and G. Haffari, "Learning how to actively learn: A deep imitation learning approach," in *Proc. 56th Ann. Meeting Assoc. Comput. Linguist.*, 2018, pp. 1874–1883.

[39] B. Settles, "Active learning literature survey," Comput. Sci., Univ. Wisconsin–Madison, Madison, WI, USA, Rep. 1648, 2010.

[40] D. Yoo and I. S. Kweon, "Learning loss for active learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 93–102.

[41] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[42] A. B. Arrieta et al., "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020.

[43] W. Pedrycz and S. M. Chen, Eds., *Interpretable Artificial Intelligence: A Perspective of Granular Computing*. Cham, Switzerland: Springer, 2021.

[44] L. von Rueden et al., "Informed machine learning—A taxonomy and survey of integrating prior knowledge into learning systems," *IEEE Trans. Knowl. Eng.*, vol. 35, no. 1, pp. 614–633, Jan. 2023.

[45] D. Weikun, K. T. P. Nguyen, K. Medjaher, G. Christian, and J. Morio, "Physics-informed machine learning in prognostics and health management: State-of-the-art and challenges," *Appl. Math. Model.*, vol. 124, pp. 325–352, Dec. 2023.

[46] C. Mao and Y. Jin, "Uncertainty quantification study of the physics-informed machine learning models for critical heat flux prediction," *Prog. Nuclear Energy*, vol. 170, May 2024, Art. no. 105097.

[47] S. Karimpouli and P. Tahmasebi, "Physics informed machine learning: Seismic wave equation," *Geosci. Front.*, vol. 11, pp. 1993–2001, Nov. 2020.

[48] W. Yan et al., "A novel physics-regularized interpretable machine learning model for grain growth," *Mater. Design*, vol. 222, Oct. 2022, Art. no. 111032.

[49] M. L. Minsky, "Logical versus analogical or symbolic versus connectionist or neat versus scruffy," *AI Mag.*, vol. 12, no. 2, pp. 34–51, 1991.

[50] B. R. Gaines and M. L. G. Shaw, "Eliciting knowledge and transferring it effectively to a knowledge-based system," *IEEE Trans. Knowl. Data Eng.*, vol. 5, no. 1, pp. 4–14, Feb. 1993.

[51] R. Rush and W. A. Wallace, "Elicitation of knowledge from multiple experts using network inference," *IEEE Trans. Knowl. Data Eng.*, vol. 9, no. 5, pp. 688–696, Sep./Oct. 1997.

[52] E. A. Feigenbaum and P. McCorduck, *The Fifth Generation*. Reading, MA, USA: Addison-Wesley Publ. Company, 1983.

[53] F. Piccialli, M. Raissi, F. A. C. Viana, G. Fortino, H. Lu, and A. Hussain, "Guest editorial: Special issue on physics-informed machine learning," *IEEE Trans. Artif. Intell.*, vol. 5, no. 3, pp. 964–966, Mar. 2024.

[54] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015, *arXiv:1503.02531*.

[55] J. De Kleer and J. S. Brown, "A qualitative physics based on confluences," *Artif. Intell.*, vol. 24, pp. 7–83, Dec. 1984.

[56] S.-F. Su, W. Pedrycz, T.-P. Hong, and F. D. A. T. De Carvalho, "Guest editorial special issue on Granular/symbolic data processing," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 342–342, Feb. 2026.

[57] M. R. G. Raman and A. P. Mathur, "A hybrid physics-based data-driven framework for anomaly detection in industrial control systems," *IEEE Trans. Syst., Man Cybern. Syst.*, vol. 52, no. 9, pp. 6003–6014, Sep. 2022.