REVIEW
https://doi.org/10.26599/FIE.2023.9270001

Fuzzy Information and Engineering
ISSN: 1616-8658

Advancing Federated Learning with Granular Computing

Witold Pedrycz ✉

ABSTRACT

Over the recent years, we have been witnessing spectacular achievements of Artificial Intelligence (AI) and Machine Learning (ML), in particular. We have seen highly visible accomplishments encountered in natural language processing and computer vision impacting numerous areas of human endeavours. Being driven inherently by the technologically advanced learning and architectural developments, ML constructs are highly impactful coming with far reaching consequences; just to mention autonomous vehicles, health care imaging, decision-making processes in critical areas, among others. The quality of ML architectures and credibility of generated results are inherently implied by the nature, quality, and amount of available data. The credibility of ML models and confidence quantified their results are also of paramount concern to any critical application. In this study, we advocate that the credibility (confidence) of results produced by ML constructs is inherently expressed in the form of information granules. Several development scenarios are carefully revisited including those involving constructs in statistics (confidence and prediction intervals), probability (Gaussian process models), and granular parameters (fuzzy sets and interval techniques). We augment the commonly encountered and challenging category of applications of ML referred to as federated learning where the aspect of quality of the model and its results calls for a thorough assessment.

KEYWORDS

granular computing; credibility; machine learning; justifiable granularity; federated learning

1 Introductory Notes: At the Junction of Machine Learning and Granular Computing

Information granules are intuitively appealing constructs, which play a pivotal role in human cognitive and decision-making activities[1–3]. We perceive complex phenomena by organizing existing knowledge along with available experimental evidence and structuring them in a form of some meaningful, semantically sound entities, which are central to all ensuing processes of describing the world, reasoning about the environment, and supporting decision-making activities.

The terms information granules and information granularity themselves have emerged in different

Department of Electrical & Computer Engineering, University of Alberta, Edmonton, T6G 2R3, Canada
✉ Address correspondence to Witold Pedrycz, wpedrycz@ualberta.ca

contexts and numerous areas of application. Information granule carries various meanings. One can refer to Artificial Intelligence (AI) in which case information granularity is central to a way of problem solving through problem decomposition, where various subtasks could be formed and solved individually. Information granules and the area of intelligent computing revolving around them being termed granular computing are quite often presented with a direct association with the pioneering studies by Zadeh[3]. He 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. Subsequently, granular computing is about representing, constructing, processing, and communicating information granules. The concept of information granules is omnipresent and this becomes well documented through a series of applications.

Machine Learning (ML) has enjoyed a rapid progress by bringing a wealth of conceptual developments, impressive learning algorithms, and far-reaching applications.

Yet with the broad scope of applications, especially in critical areas including autonomous systems, there are a number of quests that start challenging the core technologies and practices of ML. One can point here at enormous computing overheads, limited interpretability and explainability, arising privacy concerns, and brittleness of ML solutions. If not addressed properly, all of those could form bigger obstacles in the passage of time. To alleviate these issues, there have been a visibly extended agenda of ML as illustrated in Fig. 1 by accommodating new directions of green AI[4–7] and explainable AI (XAI)[8], among others. The intent is to augment the ML area by bringing new technologies that have been around and whose accommodation here could lead to fostering new possibilities. The objective of this study is to identify a role of granular computing in the augmentation of the quality of constructs of ML, especially when it comes to the quantification of credibility of such models and their results. We advocate that their quality can be expressed in terms of information granules. Such proposals cast in the general setting of granular computing have not been studied so far and as such, this study offers a novel perspective. This requirement is of particular relevance in case of federated learning. The original granular augmentation of this learning framework is proposed and investigated in detail.

Further algorithmic aspects of federated learning are discussed with regard to fuzzy rule-based models. In particular, the focus is on rules in the form "if $\boldsymbol{x}$ is A_i then $y = L(\boldsymbol{x}; w_i)$" with A_i being an information granule

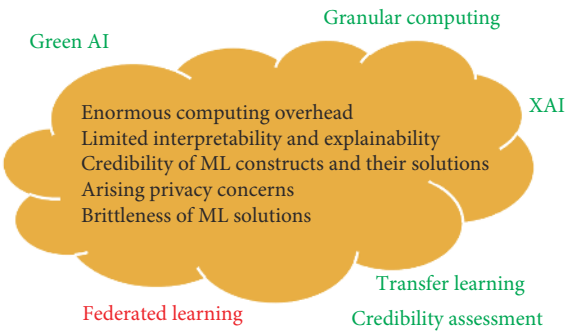

Fig. 1 Augmentations of the agenda of ML.

and L denoting a linear or constant function endowed with parameters w_i.

2 Granular Computing: Elicitation and Processing Information Granules

For the sake of completeness of the study, in this section, we concisely overview some essential aspects of granular computing by starting with the diversity of formal settings along with the underlying motivation, generalizations of information granules, and proceed with a design of information granules (the principle of justifiable granularity).

2.1 Information granules and their formal frameworks

The framework of granular computing along with a diversity of its formal settings offers a critically needed conceptual and algorithmic environment. A suitable perspective built with the aid of information granules is advantageous in realizing a suitable level of abstraction. It also becomes instrumental when forming sound and pragmatic problem-oriented trade-off among precision of results, their easiness of interpretation, value, and stability (where all of these aspects contribute vividly to the general notion of actionability).

There are numerous well-established formal frameworks of information granules; the commonly encountered include:

Sets (intervals). They realize a concept of abstraction by introducing a notion of dichotomy. Along with the set theory comes a well-developed discipline of interval analysis[9–11].

Fuzzy sets. By admitting a notion of partial membership they deliver a conceptual and algorithmic generalization of sets[12–14]. They facilitate coping with situations where the principle of dichotomy is neither justified nor advantageous.

Shadowed sets. They offer an interesting description of information granules by distinguishing among three categories of elements[15, 16] by discriminating among elements, which fully belong to the concept, which are excluded from it, and those elements whose belongingness is completely unknown.

Rough sets[17, 18] 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 this nature gives rise to so-called lower and upper bounds forming the essence of a rough set.

From the practical perspective, 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.

There is an important direction of generalizations of information granules, namely information granules of higher type. The essence of information granules of higher type comes with a fact that the characterization (description) of information granules is described in terms of information granules rather than numeric entities. Well-known examples are fuzzy sets of type-2[19], granular intervals, or imprecise probabilities. For instance, a type-2 fuzzy set is a fuzzy set whose grades of membership are not single numeric values (membership grades in [0, 1]) but fuzzy sets, intervals or probability density functions truncated to the unit interval. There is a hierarchy of higher type information granules, which are defined in a recursive manner. Therefore, we talk about type-0, type-1, type-2 fuzzy sets, etc. In this hierarchy, type-0

information granules are numeric entities, say, numeric measurements.

2.2 Construction of information granules: A principle of justifiable granularity

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

The principle of justifiable granularity guides a construction of an information granule based on available experimental evidence[1]. In a nutshell, a resulting information granule arises a summarization of data (viz. the available experimental evidence). The underlying rationale behind the principle is to deliver a concise and abstract characterization of the data such that (1) the produced granule is justified in light of the available experimental data, and (2) the granule comes with a well-defined semantics meaning that it can be easily interpreted and becomes distinguishable from the others.

The two intuitively appealing requirements are quantified by the criterion of coverage and the criterion of specificity. Coverage states how much data are positioned behind the constructed information granule. Coverage quantifies an extent to which information granule is supported (legitimized) by available experimental evidence. Specificity, on the other hand, is concerned with the semantics of information granule by stressing the semantics (meaning) of the granule.

The definition of coverage and specificity requires formalization and this depends upon the formal nature of information granule to be formed. As an illustration, consider an interval form of information granule A. In case of intervals built on a basis of one-dimensional numeric data (evidence) x_1, x_2, …, x_N, the coverage measure is associated with a count of the number of data embraced by A, namely

$$cov(A) = card\{x_k | x_k \in A\}/N \tag{1}$$

where *card* $\{\cdot\}$ denotes the cardinality of A, viz. the number (count) of elements x_k belonging to A. In essence, coverage has a visible probabilistic flavor. The specificity of A, $sp(A)$ is regarded as a decreasing function g of the size (length) of information granule. If the granule is composed of a single element, $sp(A)$ attains the highest value and returns 1. If A is included in some other information granule B, then $sp(A) > sp(B)$. In a limit case if A is an entire space $sp(A)$ returns zero. For an interval-valued information granule $A = [a, b]$, a simple implementation of specificity with g being a linearly decreasing function comes as

$$sp(A) = g(length(A)) = 1 - \frac{|b-a|}{range} \tag{2}$$

where *range* stands for an entire space over which intervals are defined. Another alternative is to consider $sp(A) = \exp(-\xi|b-a|)$ with ξ being some nonnegative calibration coefficient.

Let us introduce the following product of the criteria.

$$V = cov(A)sp(A) \tag{3}$$

It is apparent that the coverage and specificity are in conflict; the increase in coverage associates with the drop in the specificity. The design of information granule is accomplished by maximizing the above product of coverage and specificity. Thus, the desired solution (optimal values of a and b) is the ones where the value of V attains its maximum.

If the data are weighted by the corresponding weights w_1, w_2, ..., w_N, then the coverage is modified to be a weighted sum in the form

$$cov(A) = \sum_{x_i \in [a,b]} w_i / \sum_{i=1}^{N} w_i \tag{4}$$

3 Credibility of ML Constructs

There are two main challenges when it comes to the construction and an efficient deployment of ML architectures. They have to be carefully addressed:

(1) Development of ML models by optimizing some loss function

There are a variety of learning schemes aimed at the minimization of the loss function.

Typically, structural and parametric optimization tasks are envisioned. Structural optimization in which a number of hyperparameters are optimized focuses in the realm of population-based optimization or a prudent search strategy over a relatively limited search space. The parametric optimization involves some gradient-based optimization.

(2) Quantification of credibility of the model and its results

This phase, although crucial to any applications addressing the need to express how much confidence could be associated with the constructed ML model, is less visible in comparison to the first one. Yet, the credibility of the ML model and its ensuing parameters is highly relevant implying the usefulness of the developed model and the credibility in the results. The issue of awareness of the quality of the model becomes more central and will play even more visible role given the scope of existing and future applications, especially those concerning critical and autonomous systems.

A numeric result of prediction or classification does not carry any associated credibility measure. We advocate that the credibility can be associated with the numeric results by making its granular description, viz. by forming an information granule formed around the original numeric finding. The information granule delivers a well quantifiable result of credibility of the outcome and makes the user or associated system (e.g., an autonomous system) aware about the quality of the result implying possible activity to be taken, in particular to take some action or rather to collect more experimental evidence.

It is worth noting that this line of thought invoking information granule has been studied in the past under some particular assumptions. For instance, in linear regression analysis, the results are provided in terms of interval information granules guided by some probabilistic evidence and leading to confidence or prediction intervals. In case of nonlinear models, one has to consider more specialized approaches such as a delta method, mean-value estimation (MVE), and bootstrapping.

Another alternative is to resort to Bayesian models and Gaussian processes[20], in particular. In these cases, the results of the model are probabilistic information granules.

From the architectural perspective, we can think of a granular embedding the original numeric ML model as illustrated in Fig. 2. The embedding mechanism is endowed with a level of information granularity ε which can be thought as a design asset. From the algorithmic perspective, the embedding is realized by

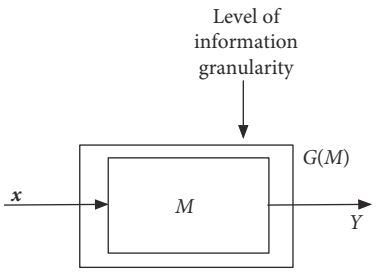

Fig. 2 A granular embedding of ML model; included is a level of information granularity treated as a certain design asset.

optimizing a certain performance index characterizing the quality of granular results when being confronted with the data.

Let us start with a numeric model M expressed as $y = M(\boldsymbol{x}; \boldsymbol{w})$ that has been designed in a supervised mode on the basis of pairs of input-output data $(\boldsymbol{x}_k, target_k)$, k=1, 2, ..., N. Here $\boldsymbol{x}$ stands for the vector of input variables, $\boldsymbol{w}$, $\dim(\boldsymbol{w}) = m$. $\boldsymbol{w}$ denotes a vector of estimated parameters of the model, $target_k$ is the output data for the corresponding $\boldsymbol{x}_k$.

The parameters $\boldsymbol{w}$ are elevated to a numeric counterpart in the following fashion.

$$\boldsymbol{w} \xrightarrow{G,\varepsilon} W \tag{5}$$

i.e.,

$$W = G(\boldsymbol{w}, \varepsilon) \tag{6}$$

where the level of information granularity ε gives rise to granular information granules W. Namely, if we admit information granules in the form of intervals, we have the following expressions.

$$w_i \xrightarrow{\varepsilon} [\min(w_i(1+\varepsilon), w_i(1-\varepsilon)), \max(w_i(1+\varepsilon), w_i(1-\varepsilon))], \varepsilon \in [0,1] \tag{7}$$

$$w_i \xrightarrow{\varepsilon} [\min(w_i(1+\varepsilon), w_i/(1+\varepsilon)), \max(w_i(1+\varepsilon), w_i/(1-\varepsilon))], \varepsilon \geqslant 0 \tag{8}$$

The level ε is optimized by evaluating the resulting information granule Y=$G(M(\boldsymbol{x}; \boldsymbol{w}))$ in terms of the coverage of data and its specificity. Consider a dataset (either training, validation, or testing).

The value of the level of information granularity ε is determined through the optimization of the granular results confronted with the numeric data. In the optimization, a product of coverage cov and specifity sp (which are essential descriptors of information granules) is calculated. The optimization yields a certain compromise between these two conflicting criteria of coverage and specificity. With the increase of the values of ε, the coverage increases but this also results in lower values of specificity.

The pertinent formulas are given as

$$\overline{cov} = \frac{1}{N}\sum_{k=1}^{N} cov(target_k, Y_k) \tag{9}$$

$$\overline{sp} = \frac{1}{N}\sum_{k=1}^{N} sp(Y_k) \tag{10}$$

where the above measures are defined in Eqs. (1) and (2) and averaged over the corresponding data. We

aim to maximize both the measures as in the case of the principle of justifiable granularity. In other words, we have ε_{opt} being a solution to the same problem as already discussed in Eq. (3).

$$\varepsilon_{\text{opt}} = \arg\max_{\varepsilon}(cov(\varepsilon)sp(\varepsilon)) \tag{11}$$

The higher the product of coverage and specificity is, the better the generated granular results are.

The level of information granularity could be more refined by admitting that each parameter of the model, w_1, w_2, ..., w_m can be transformed to its granular counterpart by associating ε_1, ε_2,..., ε_m with the corresponding parameters

$$w_i \overset{G,\varepsilon_i}{\rightarrow} W_i \tag{12}$$

$\boldsymbol{\varepsilon}=[\varepsilon_1, \varepsilon_2, ..., \varepsilon_m]$. This yields the following optimization problem

$$\varepsilon_{\text{opt}} = \arg\max_{\varepsilon}(cov(\varepsilon)sp(\varepsilon)) \tag{13}$$

The calculus of intervals with the algebraic operations follows the well-known formulas[9].

In case of monotonic functions, we have $f([a, b])=[f(a), f(b)]$ for increasing functions and $f([a, b])=[f(b), f(a)]$ for decreasing functions. In general case the extension principle is applied[13, 14].

In rule-based models the parameters of the local functions in the conclusion are made granular. In the sequel the output is computed as follows.

4 Federated Learning

Federated learning[21–24] has emerged as a real-world challenging suite of problems in which we engage ML learning in the environments where data are available only locally and cannot be shared. Our objective is to build a single holistic model when not having a direct access to data. It is a visible quest that goes beyond the settings of commonly encountered learning environment where a single dataset is available for which learning is completed.

In the environment of federated learning, there are p datasets (data islands) $\mathscr{D}_1, \mathscr{D}_2, \ldots, \mathscr{D}_p$ that are isolated and the data cannot be interacted with beyond their local environment. An organization of learning is structured in Fig. 3. There are a number of clients associated with the data islands and a single server. The role of the server is to construct a holistic model. The design process is arranged as a series of interactions among the clients and the server during which the parameters of the model are formed in an iterative

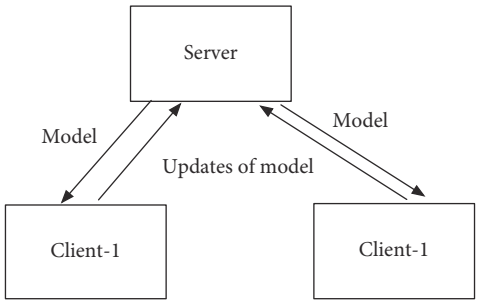

Fig. 3 An overall scheme of federated learning.

manner.

The environment of federated learning is characterized by data format and learning schemes.

4.1 Data format

There are two key modes in which data are encountered, namely horizontal and vertical ones. In the horizontal mode, the data islands $\mathscr{D}_1, \mathscr{D}_2, \ldots, \mathscr{D}_p$ are composed of different data items $\mathscr{D}_1, \mathscr{D}_2, \ldots, \mathscr{D}_p$ while they are defined in the same feature space, $\mathscr{F}_1, \mathscr{F}_2, \ldots, \mathscr{F}_p$.

In the vertical mode, the data islands are composed of the same data items, however they are defined in different feature spaces. In this sense, we can regard these data islands processed in the federated mode give rise to a multi-view perspective at data.

4.2 Learning schemes

There are two main ways in which the process of federated learning is completed. The essence is about a way how knowledge about the behavior of the model provided by the clients is communicated to the server and subsequently the server updates the model to be next communicated to the clients.

Federated average. The parameters of the models are produced by averaging the parameters provided by the individual clients. The server starts with some initial model ($\boldsymbol{w}$) and sends $\boldsymbol{w}$ to the clients. The clients develop their own models starting from the provided initial parameters based on $\boldsymbol{D}_{ii}$ and sends the optimized parameters $\boldsymbol{w}_{ii}$ to the server. The server aggregates (averages) them in the following form:

$$\boldsymbol{w} = 1/p \sum_{ii=1}^{p} \gamma_{ii} \boldsymbol{w}_{ii} \tag{14}$$

where γ_{ii} is some additional weight reflecting the contribution of the individual clients on the optimization of the model. The results are sent to the clients and the iterative process continues.

Gradient-based. In the gradient-based mode of learning, the updates of the model constructed by the server are based on the gradient of the loss function Q computed by each client once the model from the server has been provided.

$$\boldsymbol{w}(iter+1) = \boldsymbol{w}(iter) - \alpha \sum_{ii=1}^{p} \gamma_{ii} \nabla_{\boldsymbol{w}} Q \tag{15}$$

where α stands for some learning rate.

To proceed with more focused discussion, we consider a design process of rule-based models and look at the construction process. The rules come in the form where the conclusions are constant functions

$$\text{if } \boldsymbol{x} \text{ is } A_i, \text{then } y = w_i \tag{16}$$

The result produced by the model is expressed as $\sum_{i=1}^{c} A_i(\boldsymbol{x}) w_i$. Let us recall that there are two design phases of the design: (1) construction of condition parts of the rules, and (2) development of conclusions of the rules. They are well established and there are a number of optimization methods.

The condition parts are information granules in the form of fuzzy sets. They are constructed with the use

of fuzzy clustering-Fuzzy C-Means (FCM)[14]. The result comes in the form of c prototypes $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, ..., $\boldsymbol{v}_c$ and fuzzy sets A_i are computed on a basis of the prototypes. Their membership functions are computed as

$$A_i(\boldsymbol{x}) = \frac{1}{\sum_{j=1}^{c} \frac{||\boldsymbol{x}-\boldsymbol{v}_i||^2}{||\boldsymbol{x}-\boldsymbol{v}_j||^2}} \tag{17}$$

The conclusion parts are linear functions of inputs so there is an analytical formula producing a global minimum of Q. Alternatively, the optimization comes in the iterative form.

$$w_i(iter+1) = w_i(iter) - \alpha \frac{\mathrm{d}Q}{\mathrm{d}w_i} \tag{18}$$

where α is a learning rate.

These two design steps have to be carefully revisited in federated learning of these models and accommodate in this environment. We proceed with the data present in the horizontal mode and discuss the gradient-based learning.

Horizontal mode of data. The commonly used FCM algorithm has to be revisited so that is can produce the common structure across the data islands. Referring to the minimized objective function, the collaborative clustering involves an optimization of the prototypes[25]. Denote by ii is the index of the ii-th data island and the corresponding partition matrices and prototypes. The objective function (loss function) at the ii-th client reads as

$$Q[ii] = \sum_{i=1}^{c} \sum_{k=1}^{N[ii]} u_{ik}^2[ii] ||x_k[ii] - v_i||^2 \tag{19}$$

The gradient-based optimization concerns the optimization of the prototypes in the following form:

$$v_i(iter+1) = v_i(iter) - \alpha \sum_{ii=1}^{p} {}_{v_i} Q[ii] \tag{20}$$

For the updated prototypes, the partition matrix is given as

$$u_{ik}[ii] = \frac{1}{\sum_{j=1}^{c} \left(\frac{||x_k[ii]-v_i[ii]||}{||x_k[ii]-v_j[ii]||} \right)^{1/(m-1)}} \tag{21}$$

The optimization of the parameters of the conclusion parts follows the iterative scheme

$$V[ii] = \sum_{k=1}^{N[ii]} (target_k[ii] - y_k[ii])^2 \tag{22}$$

and

$$\frac{\partial V[ii]}{\partial w_i} = -2 \sum_{k=1}^{N[ii]} (target_k[ii] - y_k[ii]) A_i(x_k[ii]) \tag{23}$$

An overall scheme is portrayed in Fig. 4.

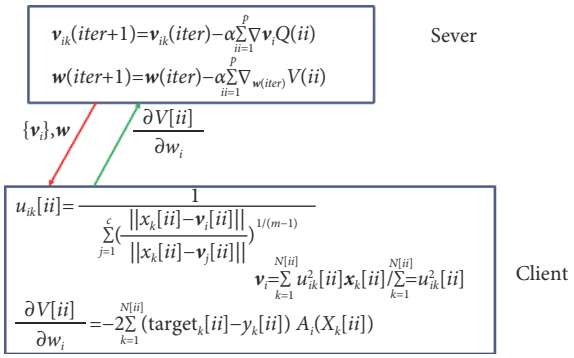

Fig. 4 Federated learning: horizonal mode of data; *m*—fuzzification coefficient.

Vertical mode of data. In this mode, there are different features for each data island, however there are the same data records. This means that in the clustering the server optimizes a partition matrix that is common across the data islands however prototypes are not a subject of communication among the server and the clients. The objective function at the individual clients is expressed in the following form:

$$Q[ii]=\sum_{i=1}^{c}\sum_{k=1}^{N}u_{ik}^2\|x_k[ii]-v_i[ii]\|_{F[ii]}^2 \tag{24}$$

The corresponding gradients communicated to the server are expressed as

$$u_{ik}(iter+1)=u_{ik}(iter)-\alpha\sum_{ii=1}^{p}\frac{\partial Q[ii]}{\partial u_{ik}(iter)} \tag{25}$$

The prototypes and partition matrix are updated in the form

$$v_i[ii]=\sum_{k=1}^{N}u_{ik}^2x_k[ii]/\sum_{k=1}^{N}u_{ik}^2 \tag{26}$$

$$u_{ik}=\frac{1}{\sum_{j=1}^{c}\left(\frac{\|x_k[ii]-v_i[ii]\|_{F[ii]}}{\|x_k[ii]-v_j[ii]\|_{F[ii]}}\right)^{1/(m-1)}} \tag{27}$$

The detailed optimization scheme is presented in Fig. 5.

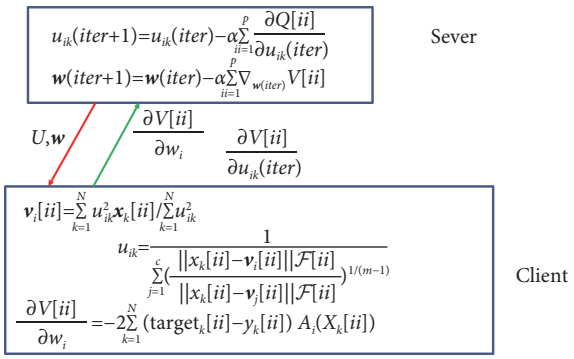

Fig. 5 Federated learning: vertical mode of data; *m*—fuzzification coefficient.

5 Federated Learning with Information Granules: Development of Credibility of Models

Data islands usually have different statistical characteristics; as a matter of fact, the *i.i.d.* assumption is not satisfied. This implies that the model constructed by the server has to be endowed by some credibility measure.

In light of the non-*i.i.d* nature of data (one cannot claim in federated learning that the data islands satisfy the assumption of independent and identically distributed random variables) supplied by individual clients, it is unlikely the model constructed by the server fits all of them. The careful quantification of the resulting model is essential in many ways. In particular, it helps monitor contributions provided by the individual clients (e.g., by weighting the gradient of the loss function) and finally provides a certain figure of merit to the constructed model. The quality is assessed by elevating the numeric model to its counterpart, where the level of information granularity is viewed as a measure of the performance of the model, which directly manifests in the granular character of the generated results.

The model M designed in the federated learning is confronted with local data D_{ii} located at ii-th client. This results in its granular counterpart $G(M)|_{D_{ii}}$. Such a granular model is designed as outlined in Section 6. In the same way, we proceed with all remaining data islands obtaining the granular characterizations of the model $G(M)|_{D_1}$, $G(M)|_{D_2}$, …, $G(M)|_{D_p}$ where p stands for the number of data islands.

Note that $G(M)|_{D_{ii}}$ is characterized by the level of information granularity ε_{ii}. This level is the result of the maximization of the product of coverage and specificity, namely $\varepsilon_{ii} = \arg\max(cov \times sp)$. The corresponding levels of information granularity ε_1, ε_2, …, ε_p are aggregated subsequently giving rise to some value of ε using which a granular model $G(M)$ is formed and delivers granular results. Various aggregation schemes *agg* could be sought, namely logic-based operators, a family of generalized averaging operators[26], among other alternatives. For positive z_1, z_2, …, z_n, $z = [z_1, z_2, \ldots, z_n]$ and non-negative weights $\boldsymbol{g} = [g_1, g_2, ..., g_n]$, the weighted generalized mean (power mean) is expressed as

$$agg(z, g) = \left(\sum_{i=1}^{n} g_i z_n^r / \sum_{i=1}^{n} g_i \right)^{1/r} \tag{28}$$

For $g = 1$ and $r = 1$, we obtain a well-known average, $r = 0$ returns a geometric mean whereas r tending to infinity or minus infinity returns the maximum or minimum, respectively.

The performance of the model is assessed at the level of the individual clients who return the values ε_1, ε_2, …, ε_p. At the level of the server, these levels are aggregated using some aggregation operator *agg* leading to ε^*. In the next step, the quality of the granular model built with the aid of ε^* is assessed in the presence of the corresponding data islands $\boldsymbol{D}_i$ and the following sum is formed

$$V = V_1(\varepsilon^*) + V_2(\varepsilon^*) + \cdots + V_p(\varepsilon^*) \tag{29}$$

The choice of *agg* coming from some family of operators A and its parameters are optimized through the optimization of Eq.(29).

$$agg_{\text{opt}} = \arg\max_{A, g^*} V^{\sim} \tag{30}$$

The levels of information granularity could also be involved in the successive steps of learning by impacting contributions coming from the clients (see Fig. 6).

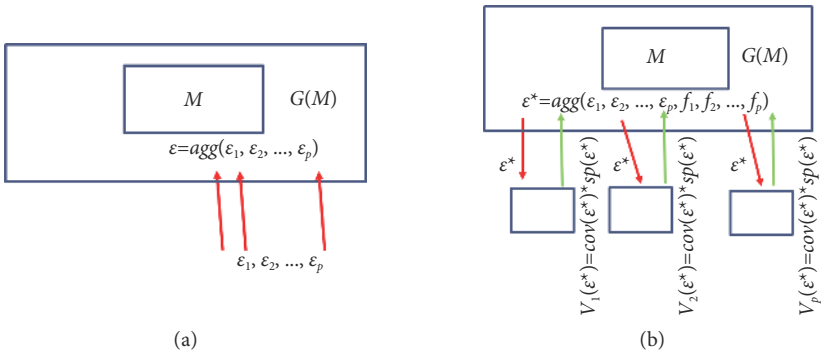

Fig. 6 (a) Designing granular model of federated learning with the use of levels of information granularity and (b) optimization of the aggregation operator.

6 Conclusion

The study has offered a new perspective at the area of federated learning by stressing a need to efficiently express the credibility of ML models constructed in this environment. We have presented a role of information granules in the formation of a granular envelope of originally formed models and showed how granular parameters of the models are optimized by flowing thew principle of justifiable granularity. The key characteristics of the granular results are delivered through the coverage and specificity measures-two fundamental features studied in granular computing when designing information granules. Detailed considerations were focused on the design of rule-based models in which case the scenarios of horizontal and vertical modes of data are investigated leading to substantial differences in the realization of clustering algorithms. It has been shown that another aspect of optimization has to be explored in a way in which levels of information granularity are aggregated.

While the analysis and design presented here concern federated learning involving interval information granules and rule-based models, further studies would focus on other alternatives embracing various formal settings of information granules and ML models.

Publication History

Received: 31 January 2023; Revised: 2 February 2023; Accepted: 24 February 2023

References

[1] W. Pedrycz, *Granular Computing: Analysis and Design of Intelligent Systems*. Boca Raton, FL, USA: CRC Press, 2013.
[2] W. Pedrycz, Granular computing for data analytics: A manifesto of human-centric computing, *IEEE/CAA J. Autom. Sin.*, vol. 5, no. 6, pp. 1025–1034, 2018.
[3] L. A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets Syst.*, vol. 90, no. 2, pp. 111–127, 1997.
[4] R. C. Castanyer, S. Martínez-Fernández, and X. Franch, Which design decisions in AI-enabled mobile applications contribute to Greener AI?, arXiv preprint arXiv: 2109.15284, 2021.
[5] W. Pedrycz, Towards green machine learning: Challenges, opportunities, and developments, *J. Smart Environ. Green Comput.*, vol. 2, pp. 163–174, 2022.
[6] R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, Green AI, arXiv preprint arXiv: 1907.10597, 2019.

[7] T. Tornede, A. Tornede, J. Hanselle, M. Wever, F. Mohr, and E. Hüllermeier, Towards green automated machine learning: Status quo and future directions, arXiv preprint arXiv: 2111.05850, 2021.
[8] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, et al., Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, *Inf. Fusion*, vol. 58, pp. 82–115, 2020.
[9] G. Alefeld and J. Herzberger, *Introduction to Interval Computations*. New York, NY, USA: Academic Press, 1983.
[10] R. E. Moore, *Interval Analysis*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1966.
[11] R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*. Philadelphia, PA, USA: SIAM, 2009.
[12] D. Dubois and H. Prade, An introduction to fuzzy systems, *Clin. Chim. Acta*, vol. 270, no. 1, pp. 3–29, 1998.
[13] H. Nguyen and E. Walker, *A First Course in Fuzzy Logic*. Boca Raton, FL, USA: CRC Press, 2000.
[14] W. Pedrycz, *An Introduction to Computing with Fuzzy Sets: Analysis, Design, and Applications*, Cham, Switzerland: Springer, 2021.
[15] W. Pedrycz, Shadowed sets: Representing and processing fuzzy sets, *IEEE Trans. Syst., Man, Cybern., Part B* (*Cybern.*), vol. 28, no. 1, pp. 103–109, 1998.
[16] Y. Yao, S. Wang, and X. Deng, Constructing shadowed sets and three-way approximations of fuzzy sets, *Inf. Sci.*, vol. 412−413, pp. 132–153, 2017.
[17] Z. Pawlak, Rough sets, *Int. J. Comput. Inf. Sci.* , vol. 11, no. 5, pp. 341–356, 1982.
[18] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Dordrecht, The Netherlands: Springer, 1991.
[19] J. M. Mendel, R. I. John, and F. Liu, Interval Type-2 fuzzy logic systems made simple, *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 6, pp. 808–821, 2006.
[20] C. E. Rasmussen and G. K. I. Williams, *Gaussian Processes for Machine Learning*, Cambridge, MA, USA: MIT Press, 2006.
[21] S. Abdulrahman, H. Tout, H. Ould-Slimane, A. Mourad, C. Talhi, and M. Guizani, A survey on federated learning: The journey from centralized to distributed on-site learning and beyond, *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5476–5497, 2021.
[22] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, et al. , *Advances and open problems in federated learning*, *Found. Trends® Mach. Learn.*, vol. 14, nos. 1–2, pp. 1–210, 2021.
[23] Q. Yang, Y. Liu, Y. Cheng, Y. Kang, T. Chen, and H. Yu, *Federated Learning*, San Rafael, CA, USA: Morgan & Claypool Publishers, 2019.
[24] Q. Yang, L. Fan, and H. Yu, *Federated Learning: Privacy and Incentive*, Cham, Switzerland: Springer, 2020.
[25] 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, 2022.
[26] P. S. Bullen, *Handbook of Means and Their Inequalities*, Dordrecht, The Netherlands: Springer, 2003.

Witold Pedrycz is a professor in the Department of Electrical & Computer Engineering, University of Alberta, Canada. He is also with the Systems Research Institute of the Polish Academy of Sciences, Poland. He is a foreign member of the Polish Academy of Sciences and a fellow of the Royal Society of Canada. He is a recipient of several awards including a Norbert Wiener Award from the IEEE Systems, Man, and Cybernetics Society, an IEEE Canada Computer Engineering Medal, a Cajastur Prize for Soft Computing from the European Centre for Soft Computing, a Killam Prize, a Fuzzy Pioneer Award from the IEEE Computational Intelligence Society, and the 2019 Meritorious Service Award from the IEEE Systems Man and Cybernetics Society. His main research directions involve computational intelligence, granular computing, and machine learning. He serves as an editor-in-chief of *Information Sciences* and *WIREs Data Mining and Knowledge Discovery*, and co-editor-in-chief of *Granular Computing* and *Journal of Data*, *Information and Management*.