EDITOR'S PERSPECTIVE

Beyond i.i.d.: Non-IID Thinking, Informatics, and Learning

Longbing Cao, *University of Technology Sydney, Sydney, NSW, 2007, Australia*

In science, technology, engineering, and their applications, a ubiquitous assumption is independent and identically distributed (i.i.d. or IID). IID simplifies the intricate realities and complexities of the real world for their approximate quantifiability and tractability and asymptotic problem-solving. It, however, has also induced significant limitations and gaps in their knowledge, capability, and capacity. Reality-level problem-solving *has to go beyond i.i.d. and adopt a broad-reaching* non-IID thinking *approach, i.e., exploring the comprehensive* non-IIDness—heterogeneity *and* interaction *(couplings and entanglement)—of an underlying problem and its behaviors, environment, data, and problem-solving system. Accordingly, this article discusses this foundational, cross-disciplinary, cross-domain, and cross-theory and -practice problem-going beyond i.i.d. and proceeding with non-IID. Concepts, challenges, and prospects of non-IIDness, non-IID thinking, informatics, and learning are discussed for developing* reality-level *AI, data science, machine learning, and intelligent systems by identifying, quantifying, and synthesizing heterogeneities, interactions, complexities, and intelligence in complex systems and data.*

Real-world complexities continuously challenge the existing body of knowledge in science and technology. A critical issue is the independent and identically distributed (i.i.d., or IID, for compatibility across different disciplinary habits, here we use IID) assumption, which has laid the foundation of modern science, technology, and engineering. Over the recent half-century, the debate on IID limitations and the demand for new paradigms have increased and expanded from statistics to informatics and computing and other disciplines,[1] including quantum physics, communication, control systems, statistics, informatics, and computing. This article is thus motivated to discuss this important topic—beyond i.i.d., and to advocate non-IID thinking for reality-level science and technology, more specifically non-IID informatics, AI, data science, and machine learning.

PARADIGM SHIFT: FROM IID TO NON-IID

IID and Its Significant Gaps

In probability and statistics, the *IID (or i.i.d.) assumption* refers to the scenario where a collection of random variables are mutually independent and drawn from the same probability distribution.[2] We call such variables to have *IIDness*, i.e., independence and identical distribution, and their data are IID. This IID assumption has served a foundational role in probability operations (e.g., chain rule and *d*-separation), theories (e.g., the law of large numbers and Bayes rule) and statistics (e.g., mean, variance and maximum likelihood estimate), exact and approximate inference, and their applications in modeling uncertainty (e.g., Bayesian networks).[3] IID has further been applied to other disciplines, such as coding, information theory, machine learning, signal processing, and data analytics. It has been treated as a foundation of classic and modern science, technology, and engineering. In data science and data-driven AI, the role and importance of IID has been overwhelming as well, from sampling, feature relation analysis, uncertainty learning, shallow and deep learning theories, and systems, to optimization and generalization theories.

More broadly, the IID assumption has played a paramount role in driving scientific, technological, and engineering developments. It simplifies and transforms complex real-life problems and challenges into representable, quantifiable, and tractable solutions. However, this IID and near-IID simplification has also induced significant limitations, constraints, and gaps. Examples are 1) an incomplete understanding and over-simplification,

1541-1672 © 2022 IEEE
Digital Object Identifier 10.1109/MIS.2022.3194618
Date of current version 12 September 2022.

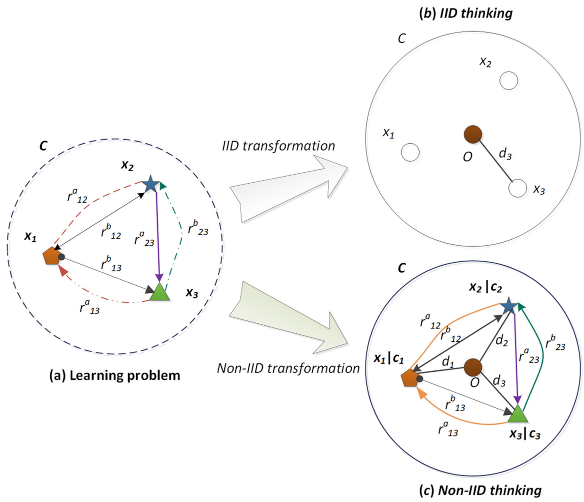

FIGURE 1. IID thinking versus non-IID thinking. For example, from the machine learning perspective, a given learning problem (a) is either (b) IID transformed per the IID assumption (i.e., independent and identically distributed) and then solved by an IID learning system, or (c) non-IID transformed by characterizing its non-IIDness (i.e., heterogeneity and interaction) and then solved by a non-IID system.

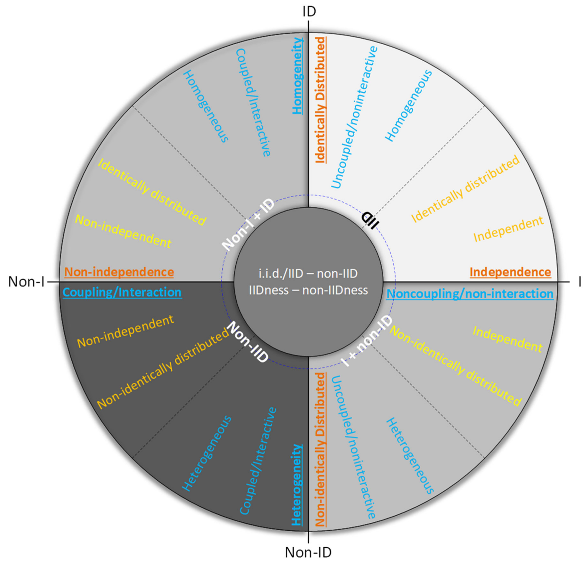

FIGURE 2. IID to non-IID space. Two sets of axes: classic independence/nonindependence-identical distribution/nonidentical distribution versus heterogeneity/homogeneity-coupling/interaction//noncoupling/noninteraction; generating four quadrants: IID, non-I + ID, non-IID, and I + non-ID.

2) a partial characterization, 3) biased solutions, 4) adverse consequences, and 5) significant knowledge gaps of underlying problems, which are briefly discussed in the following.

First, IID only reflects an incomplete understanding of the characteristics and complexities of a system. As shown in Figure 2, IID only focuses on one extreme scenario of the overall problem space, i.e., those are independent and identically distributed or more broadly homogeneous, uncoupled, and noninteractive. They only correspond to specific scenarios of homogeneity and independence, as shown in the first quadrant.

Second, IID only makes a partial characterization of an underlying problem. The IID understanding captures a partial picture of the overall problem space by simplifying and neutralizing the characteristics and complexities of heterogeneity and interaction to homogeneity and independence. The comprehensive aspects and properties of heterogeneities and interactions of real-life problems structured in Figure 3 are then over-simplified, -normalized, or -abstracted, as shown in Figure 1.

Third, IID leads to biased IID theories, techniques, tools, systems, evaluation, and results. The resultant IID science and technology are biased to their assumptions. Most existing scientific, technological, and engineering developments are biased to IID or near-IID settings. Many evaluation measures and methods apply the IID assumption as well. They thus are only applicable to IID scenarios and settings.

Furthermore, IID results in adverse consequences when their biased understanding and solutions are applied to problem-solving and decision-making. It is important to know IID results are biased. They only capture a partial, incorrect, or misleading picture of the underlying problems and may lead to risky or costly consequences if their results and tools are applied.[4] Hence, IID evaluation may result in biased or incorrect justification and results.

Finally, IID leaves significant knowledge gaps. IID theories, techniques, and tools are short of the required capabilities and capacity to fully address complex problems and their challenges. Figure 4 illustrates various perspectives and opportunities to address the gaps in existing learning systems. In principle, every existing discipline, research area, and the tool have its limitations and constrained applicability, and *beyond IID thinking* is to disclose its gaps and inspire important gap-filling perspectives and opportunities.

Emergence of Non-IID

In both theory and practice, the IIDness and IID assumption do not reflect the reality of many complex scenarios and settings in the real world and its variables, behaviors, environments, data, and solutions.

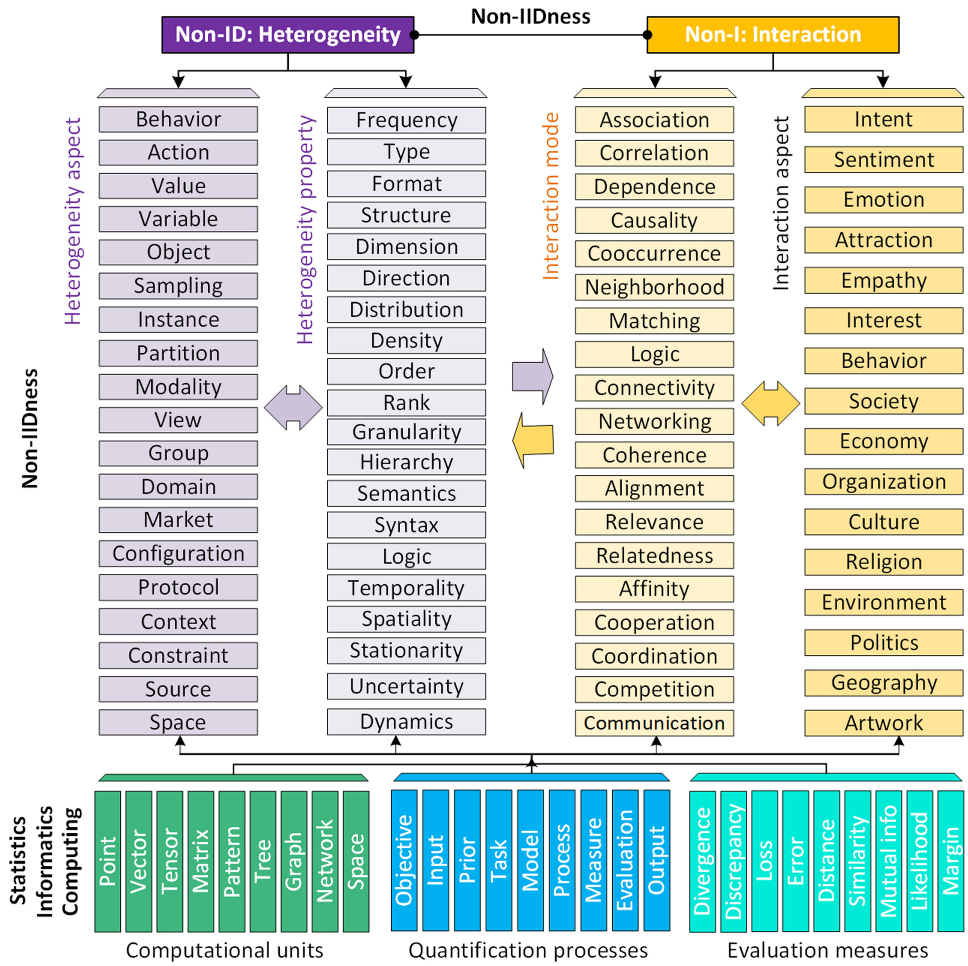

FIGURE 3. Terminology and conceptual map of non-IIDness: non-ID—heterogeneities, and non-I—interactions.

They also do not conform to the characteristics of real-life problems, systems, and data, where variables (or features and attributes in different domains) interact and are not mutually independent (thus nonindependent) and do not follow the same distribution (hence nonidentically distributed). This leads to the *non-IID case*[1] and *beyond IID* scenarios, where variables and their data go beyond the statistical IID assumption but hold *nonindependence* and *nonidentical distribution*. More broadly and generally, *nonindependence* further corresponds to diverse interactions, couplings, and entanglement (interaction for short); *nonidentical distribution* is expanded to comprehensive heterogeneities. *Heterogeneity and interaction forms the general non-IIDness and non-IID settings* in statistics, informatics, and computing.[5–7]

In practice, what consequences may the IID theories and tools cause? In theory, is nonindependent complementary to independent statistically, qualitatively, and quantitatively? What forms the opposite of independence in the real world? What forms the non-IIDnesses of real-life systems and data? Do the state-of-the-art deep neural networks address non-IIDness? How does non-IID affect the theories and practice of AI, data science, machine learning, and more broadly? How can non-IID AI and data science cope with the non-IID nature of the real world? Would non-IID AI, non-IID learning, non-IID analytics, and non-IID processing better support the realitylike and actionable AI, machine learning, data analytics, and signal processing, required to address real-world complexities? Last but not least, theoretically, how can the non-IID intractability be addressed or approximated to an actionable level? Addressing these questions motivates this article to expand our early thinking and argument on non-IIDness and non-IID learning[5,6] to more broad-reaching cases, issues, and perspectives in AI and data science and our early trials and practices in areas, such as

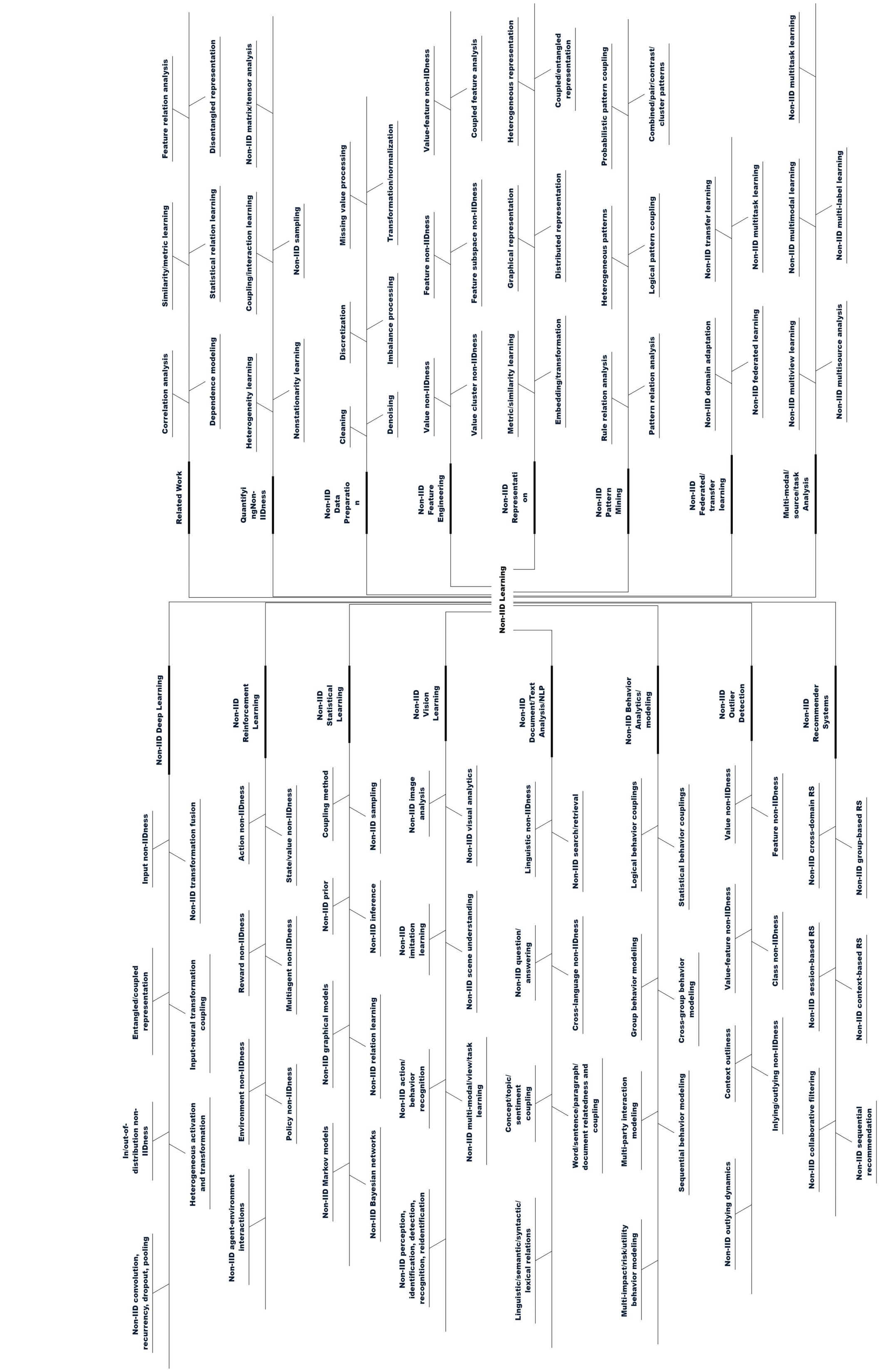

FIGURE 4. Research map of non-IID learning.

similarity/metric learning, unsupervised to supervised learning, and applications, such as outlier detection and recommender systems.[a]

Note, here, we do not focus on scenarios and questions on beyond IID for quantum information theory, where concerns are on issues and topics, including coding theory, higher order analysis, quantum Shannon theory, one-shot information theory, information input-channel entanglement, information spectrum, entropy inequality, nonstandard entropy theory, and asymmetry for quantum thermodynamics and quantum information theory.[7] However, the non-IID thinking discussed here is generally applicable to them and their studies form part of the entire spectrum of non-IID informatics, computing, and learning.

GENERAL NON-IID FRAMEWORK

Here, we first discuss the IID to non-IID thinking transformation and then present a conceptual map of non-IIDness that comprises a spectrum of heterogeneities and interactions for non-IID processing and learning.

IID Thinking versus Non-IID Thinking

First, IID thinking versus non-IID thinking are illustrated in Figures 1 and 2, respectively. *IID thinking* transforms a complex system to be IID, where only IIDnesses—i.e., independence and identical distribution, or noncoupling/noninteraction and homogeneity—in the underlying problem are assumed and characterized in the problem-solving. In contrast, *non-IID thinking* transforms the problem to be non-IID, where non-IIDnesses—i.e., nonindependence and nonidentical distribution, or coupling/interaction and heterogeneity—in the underlying problem are characterized and incorporated into the problem-solving system.

Second, we further explain the IID versus non-IID thinking in Figure 1 using a four-quadrant IID to non-IID space. This IID to non-IID space reflects the full picture of IIDness to partial-to-complete non-IIDness. Figure 2 presents, such an IID to non-IID space. The x-axis reflects two categorizations: 1) *independence* versus *nonindependence* following the classic i.i.d. annotation, and 2) *noncoupling/non-interaction* versus *coupling/interaction* following the broad-reaching cross-disciplinary terminology. Correspondingly, the y-axis also reflects two categorizations: 1) *identical distribution* versus *nonidentical distribution*, and 2) *homogeneity* versus *heterogeneity*. Accordingly, the four quadrants correspond to four scenarios of IID, to partial and complete non-IID settings, respectively. The first quadrant corresponds to IID and IIDness, where variables are mutually independent and follow the same distribution, or in other words, they are homogeneous and uncoupled/noninteractive. The remaining three quadrants correspond to partial-to-complete non-IID and non-IIDness. The second and fourth quadrants are partially non-IID. The second is non-I + ID, where the variables are nonindependent but belong to the same distribution, or homogeneous but coupled/interactive. The fourth corresponds to I + non-ID, where the variables are independent but nonidentically distributed, or heterogeneous but uncoupled/noninteractive. The third quadrant refers to complete non-IID or non-IIDness, where variables are nonindependent and nonidentically distributed, or heterogeneous and coupled/interactive.

Conceptual Map of Non-IIDness

Furthermore, we summarize the different aspects of non-IIDness in terms of heterogeneities and interactions (couplings and entanglements). Figure 3 lists both tangible and intangible aspects and properties of heterogeneities and interactions and the relations between them.

First, *heterogeneity* is embodied through tangible and intangible heterogeneity aspects and their properties. *Heterogeneity aspects* refer to entities and aspects that are visible, observable, or easily quantifiable and characterizable in systems and data. Examples of heterogeneity aspects are behavior, action, value (of variables), variable, object (or sample, instance), partition (e.g., of a dataset), modality, view (from cameras or applications), group, domain, market, configuration, protocol, context, constraint, source (of data), and space (e.g., of variables). In a given problem or dataset, they may be homogeneous or heterogeneous. These heterogeneity aspects may hold one-to-multiple heterogeneity properties. *Heterogeneity properties* examples are frequency (e.g., of categorical values), type (e.g., integer and float data), format (e.g., categorical and numerical variables), structure (e.g., shape or form), dimension, direction (including directed and undirected), distribution, density, order, rank, granularity, hierarchy, semantics, syntax, logic, temporality, spatiality, stationarity, uncertainty, and dynamics. For the same property, different heterogeneity aspects may have different meanings or be characterized in different forms. For example, the direction of a value refers to whether it is positive or negative, while the direction of a view may refer to its position.

Second, *interactions*[8] are characterized in terms of interaction aspects and modes. *Interaction aspects* may refer to the mental, cognitive, and psychological

[a][Online]. Available: https://datasciences.org/non-iid-learning/

aspects, such as intent, sentiment, emotion, attraction, empathy, and interest or other aspects, such as behavior, society, economy, organization, culture, religion, environment, politics, geography, and artwork. The *interaction modes* of these aspects may be in the form of association, correlation, dependence, causality, cooccurrence, neighborhood, matching, logic, connectivity, networking, coherence, alignment, relevance, relatedness, affinity, cooperation, coordination, competition, or communication. One interaction aspect may be affiliated with one to multiple interaction modes. For example, in cyberspace, cyber-attacks may be associated with cooperation between team members and competition with counterparties. Interaction modes may be formulated or characterized by different methods or tools. For example, the association is usually measured by support and confidence, while correlation is measured by coefficients. Third, the aspects and properties of heterogeneity may also be applicable to interactions, an interaction aspect or mode may hold some of the properties listed for heterogeneity. For example, interactions may take place between behaviors, and behavioral interactions may hold properties of temporality, spatiality, and uncertainty. On the other hand, the aspects and modes of interaction may be applicable to the aspects and properties of heterogeneity as well. For example, behaviors may take various interaction modes, such as alignment or competition, and in aspects, such as a society or organization.

Finally, both heterogeneity and interaction in a system or dataset must be quantified, characterized, or processed to incorporate non-IIDness into their problem-solver. Their quantification, characterization, and processing are usually converted to computational units, which are then processed by certain quantification processes and evaluated by their corresponding criteria or measures. *Computational units* may take the form of point, vector, array, matrix, tensor, tree, graph, network, pattern, or space. *Quantification processes* may involve the objective, input, prior knowledge, task, model, process, evaluation, measure, and output of heterogeneity or interaction. The *evaluation* of heterogeneity or interaction may be based on the criteria or aspect of divergence, discrepancy, loss, error, distance, similarity, mutual information, likelihood, or margin.

QUANTIFYING HETEROGENEITY AND INTERACTION

The aforementioned heterogeneities and interactions must be quantified so that they can be incorporated into non-IID frameworks, systems, and solutions. Here, we thus discuss their quantification.

Quantifying Heterogeneities

Figure 3 illustrates a collection of aspects and properties of heterogeneities in complex systems, behaviors, and data. How quantifying the heterogeneity aspects and properties in a system or dataset becomes a fundamental task in quantifying the non-IIDness of the system or data. Heterogeneity quantification involves several layers of characterization and formulation:

1) quantifying heterogeneity properties;
2) quantifying heterogeneity aspects in terms of quantified properties;
3) quantifying heterogeneity subjects in terms of their heterogeneity aspects, and properties.

In addition, heterogeneity and interaction are coupled and mutually influence and determine the complexities and intelligence of a system (which will be discussed in the following section). Therefore, quantifying heterogeneity also needs to involve and formulate the relevant interaction, complexity, and intelligence.

First, quantifying heterogeneity properties is to define and characterize each property of a heterogeneity aspect and then represent or formulate the property in terms of their suitable formulation tools. Here, I briefly illustrate the quantification of some heterogeneity properties listed in Figure 3. *Frequency* may be measured by count or converted to the frequency domain by Fourier transform. *Type* needs to be first specified in terms of its heterogeneity aspect and then formulated accordingly, for example, the type of values in integers or floats, versus the type of modality in text, image, or voice. *Format* is similar to or replaceable by type in some cases, for example, the format of numerical, categorical, or textual variables versus the format of constraints in probability conditions or logic clauses. The *structure* of a subject may be specified in terms of geometric, organizational, or graphical shapes or forms. The *direction* may be categorized by polarity, positivity, or cardinal point. *Distribution* may be quantified by probabilistic or mathematical functions. *Density* may be measured by mass concentration. *Order* and *rank* measure the degree of complexity, determinant, or differentiation of a variable or matrix. *Granularity* measures the layer or scale of an object, source, or configuration, etc. *Hierarchy* may be quantified in terms of rank, order, or cluster. The *semantics* and *syntax* of a text or textual aspect may be quantified by linguistic tools, neural, or statistical representation. Some heterogeneity aspects may involve *logic*, e.g., constraints and logic reasoning, and inference, where tools like mathematical logic, computational logic, or formal logic may be applied to quantify them. *Temporality* may apply to the aspects of behavior, action, variable, modality, domain, source, and context, which may be quantified by temporal order, duration, and change, etc. *Spatiality* may be associated with behavior, action, variable, view, domain, source, or space, which may be expressed by spatial factors, such as location, altitude, depth, width, height, or edit distance, etc. *Stationarity* (and nonstationarity) may be affiliated with the aspects of behavior, variable, or context, which may be measured by variance, discrepancy, or residual, etc., and their properties, such as distribution, expressed by statistics, such as the shape or rate change of a distribution. *Uncertainty* may be associated with aspects, such as variable and context or properties, such as structure, distribution, density, or stationarity, etc., which may be expressed in terms of probability, entropy, or mutual information, etc. Finally, *dynamics* may refer to an aspect or property that evolves over time, space, or object, etc., which is quantified either in terms of evolving properties, such as distribution, semantics, shift, drift, change, and nonstationarity or by incorporating it into a dynamic, adaptive, and active modeling process.

Second, the heterogeneity aspects can be formulated in terms of their quantified properties by following certain aspect-specific mathematical, statistical or logical tools, information theory, or representation methods, etc. In general, we first identify the properties of each heterogeneity aspect, quantify each property by its suitable formulation, and then quantify the aspect in terms of all quantified properties. In practice, their quantification or representation method may depend on the task of quantifying heterogeneity and how heterogeneities are involved in the problem-solving. For example, given a variable, we can quantify its value type, distribution, and uncertainty, etc., respectively. These properties will then be fed into a variable representor (a model that characterizes the variable) to catch their interactions and joint effect on the variable. When an aspect has a hierarchical structure, i.e., hierarchical aspects, an iterative quantification aggregation would be applied, starting with the atomic properties of the lowest layer, and aggregating to the highest layer of the aspect. For example, given a data source consisting of multiple modalities of data, we specify the properties (variables or attributes) of each object in each modality; the properties of different modalities affiliated with the same object are joined or jointly modeled to form the multimodal representation of the object.

Quantifying Interactions

The diverse aspects and modes of interactions illustrated in Figure 3 and Cao[8] can be universally seen in both simple and complex natural, social, economic, cultural, environmental, geopolitical, artificial, and virtual organizations and systems. Without interactions, a system would become lifeless and highly simple. Interactions serve as one of the most critical system properties, mechanisms, and behaviors of complex systems.[9,10] Interactions cause, drive, or influence system complexities, dynamics, emergence, and consequence, thus system intelligence.[9] In the increasingly complex, uncertain, and volatile world, it is essential to deeply understand, quantify, model, analyze, predict, and intervene in interactions, which would thus make it possible to better understand and manage their complexities and intelligence.

Given a complex system and its dataset, to deeply understand, quantify and manage their interactions, many questions have to be addressed as follows:

1) What do their interactions look like?
2) How do the interactions relate to their system complexities and intelligence?
3) What does their interaction taxonomy look like in general?
4) How are their interactions usually quantified in their field?
5) Where are the gaps in quantifying, understanding, and managing interactions?
6) What prospects are there in deeply understanding, quantifying, and managing interactions?

In complex systems, interactions take place in *interaction media,* including material, information, and energy spaces. Interactions between two to multiple entities are presented in the forms of connections, communications, relations, or couplings. Interactions can be quantified in terms of interaction aspects and their *interaction properties*, such as interaction party, type, format, heterogeneity, dimensionality, order, hierarchy, granularity, dynamics, and context, etc. Furthermore, an *interaction taxonomy* can be defined in terms of diversified and contrastive aspects, properties, and modes of interactions: subjective versus objective, explicit versus implicit, virtual versus physical, hidden versus observable, local versus global, homogeneous versus heterogeneous, hierarchical versus flat, static versus dynamic, and synchronous versus asynchronous. For example, in decentralized AI (DeAI) and edge intelligence,[11] interactions take place between peers at the device and edge levels. Interaction networks emerge in DeAI systems, where ends and edges are connected to each other for various purposes and in different forms, such as P2P interaction networks, chained or ring interaction networks, or asynchronous interaction networks.

Typical quantitative methods formulating interactions comprise tools for *mathematical relations,* such as matching, cooccurrence, correlation, and association; *statistical relations,* such as dependency and chain rules; *logic relations,* such as by temporal logic or fuzzy logic; *semantic relations,* such as linguistic relations; and *information theory,* such as mutual information and cross-entropy, etc. In high-dimensional, high-order, and high-rank interaction spaces, nonlinear, hidden, and indirect interactions are typically quantified in terms of geometric, distributional, process-based, and manifold spaces, relations, and tools. They are typically applicable to quantified or quantifiable interactions. In complex systems, interactions and couplings may be intangible, implicit, indirect, and transitional, such as social interactions, economic interactions, cultural interactions, organizational interactions, emotional interactions, thought exchange, and inspiration. These interactions may take decentralized forms and take place between decentralized agents.

Further *deep interaction modeling and learning*[12] thus emerges as a critical agenda to characterize, represent, model, analyze, discover, and learn interaction intelligence. The aims and tasks of interaction modeling and learning are to define, identify and formulate complex interaction aspects and properties; characterize, represent, model, analyze, and learn interaction networks, modes, strengths, and dynamics; quantify and evaluate interaction effects, and intelligence; discover, coordinate, consolidate, and federate interaction intelligence between end, edge and control agents; and exchange, transfer, and transit interaction intelligence within a multiagent system or between the system and its environment for interactive problem-solving. As DeAI and edge intelligence play increasingly important roles, another interesting perspective is to characterize, analyze, discover, utilize, and manage emergent interaction intelligence of a large scale of decentralized ends or edges, including important individual, cluster, and collective interaction intelligence. Typical theories and approaches for interaction modeling and learning include interaction networks, graph modeling, network modeling, relation learning, data-driven discovery, and deep learning approaches.[13]

To model interactions and couplings, *coupling learning*[8,12,14] has emerged as an effective approach, which learns complex couplings and interactions in a system or dataset. Significant progress and prospects have been demonstrated in many references and areas,[12,14] including coupled representation learning, pattern relation analysis, coupled similarity/metric learning, multimodal coupling learning, cross-domain coupling learning, heterogeneous hierarchical coupling learning for unsupervised to supervised tasks, cross-market coupling learning, coupled behavior analysis, and learning couplings in documents, multiple time series, recommendation, outliers, multisource data, images, and user-item interactions, etc. New mathematical tools, such as coupled matrix factorization, statistical learning models, such as hierarchical nonparametric Bayesian networks, machine learning advances, such as coupled multikernel learning, coupled similarity, and metrics, and integrating attribute couplings with deep neural learning are developed to learn the diversified couplings between attributes, objects, object clusters, and data sources, etc.

RESEARCH MAP OF NON-IID LEARNING

Research Questions

In machine learning, non-IID learning may involve parts or the whole of a learning system, from its input, feature engineering, input–output mapping (i.e., modeling), and evaluation to output. This perspective generally applies to any data-driven, model-driven, domain-driven, and hybrid learning system. Incorporating non-IID thinking into shallow and deep machine learning thus forms the area of non-IID learning.[5,6,15–18] *Non-IID learning* aims to 1) identify, represent, analyze, discover, and manage the non-IIDnesses of a system and its behaviors, environment, and data; and 2) develop the corresponding non-IID learning theories and systems.

Various research questions may be targeted in developing non-IID learning, such as:

- Where and what non-IIDness presents in a system or its data?
- What is on the checklist to verify whether and to what extent non-IIDness exists in a system or dataset?
- How to measure and evaluate whether and to what extent a dataset is non-IID?
- How to verify whether and to what extent non-IIDness needs to be addressed in learning?
- How can broad-reaching non-IIDnesses be addressed statistically either in the current statistical principles or by extending them?
- What are the issues and gaps in existing statistical learning in addressing non-IIDness? How to address them?
- What are the issues and gaps in existing information theories in addressing non-IIDness? How to address them?

- How and to what extent do the existing shallow machine learning systems address non-IIDness?
- What are the issues and gaps in the existing shallow machine learning systems in addressing non-IIDness? How to address them?
- Whether and to what extent the existing deep neural networks address non-IIDness?
- What are the issues and gaps in the existing deep neural networks in addressing non-IIDness? How to address them?
- How do we quantify non-IIDness computationally?
- How does non-IIDness incur the intractability during learning? How do we address it?
- Partial to complete non-IID learning may be NP-hard or NP-complete, so how do we approximate it?

More foundational non-IID thinking is essential in breaking through the constraints of existing learning principles, paradigms, and systems. Non-IID thinking encourages the independent thinking of the intrinsic, fundamental, and original non-IIDnesses in a system or its data from perspectives of heterogeneity, interaction (coupling, entanglement), complexity and intelligence. Independent non-IID thinking is to go beyond the constraints of existing learning paradigms and principles.

Research Map

Accordingly, Figure 4 presents a research map of non-IID learning to expand the existing machine learning capacity. It illustrates the areas from shallow to deep learning by incorporating non-IIDness into their frameworks or transforming their existing learning systems to non-IID. They include quantifying non-IIDness, non-IID data preparation, non-IID feature engineering, non-IID representation learning, non-IID pattern mining, non-IID federated learning, non-IID transfer learning, non-IID multitask/label learning, non-IID multimodal/source learning, non-IID reinforcement learning, non-IID deep learning, non-IID statistical learning, non-IID vision learning, non-IID document/text analysis and natural language processing, non-IID behavior analytics, non-IID outlier detection, and non-IID recommender systems.

Quantifying non-IIDness may involve many topics, such as heterogeneity learning, coupling and interaction learning, non-IID sampling, non-IID matrix and tensor analysis, and nonstationarity learning. They characterize and quantify various aspects of heterogeneities, interactions, and their properties and challenges.

Non-IID data preparation explores non-IIDness in data characteristics and complexities[10] and takes them into consideration during data cleaning, discretization, missing value processing, denoizing, imbalance processing, and data transformation and normalization, etc. These are fundamental to retain the intrinsic and intricate non-IIDnesses of raw data in their prepared transformations for further genuine non-IID processing, analysis, and learning, etc.

Non-IID feature engineering is to develop new methods and tools for feature selection, feature relation analysis, feature mining, and feature construction by identifying and characterizing value non-IIDness, feature non-IIDness, value-feature non-IIDness, value cluster non-IIDness, and feature subspace non-IIDness, and for coupled feature analysis, selection, and mining, etc.

Non-IID representation learning is to characterize and learn representations of data or entities by 1) identifying and characterizing their non-IIDnesses, and 2) developing non-IID representation methods. Examples are non-IID metric and similarity learning, non-IID graphical representation, heterogeneous representation learning, coupled and entangled representation learning, distributed representation learning, and non-IID embedding and transformation by considering couplings and heterogeneities in observations and between inputs and representations.

Non-IID pattern mining considers the non-IIDnesses between items, itemsets, item attributes or properties, and patterns. It then discovers interesting non-IID patterns, such as by pattern (or rule) relation analysis, logic pattern coupling analysis, or discovering heterogeneous patterns, probabilistic pattern couplings, combined patterns, paired patterns, cluster patterns, or contrast patterns.

Non-IID federated learning considers the non-IIDnesses within and between local sources, tasks, and models, and within and between local and global sources, tasks, and models. It then comprises topics and opportunities, such as heterogeneous federated learning from heterogeneous local sources, tasks, and models and with heterogeneity between local and global sources, tasks, and models, and coupled federated learning by coupling local sources, tasks, and models and coupling local and global sources, tasks, and models.

Non-IID transfer learning relates to non-IID multitask learning, non-IID multi-label learning, non-IID cross-domain learning, and non-IID domain adaptation, etc. They consider the non-IIDness between domains, between labels, between tasks, and between models. They can thus form research areas and opportunities of heterogeneous transfer learning, and coupled transfer learning, etc.

Non-IID multimodal/source learning considers the non-IIDnesses within and between modalities, and sources. When multiple models are involved in each modality or source, the non-IIDness between models may also be considered. These thus form areas and opportunities of heterogeneous multimodal learning, coupled multimodal learning, heterogeneous multisource learning, and coupled multisource learning, etc.

Non-IID reinforcement learning involves non-IIDness between agents in terms of their actions, action states, values, rewards, and policies and between agents, and their environments. The non-IIDness of the environment may also be considered in the agent–environment interactions. This may thus form new perspectives of heterogeneous reinforcement learning, coupled reinforcement learning, and reinforcement learning with the non-IID environment, etc.

Non-IID deep learning is to engage non-IIDness in input, neural architectures, transformations and mechanisms, output, and specific settings, etc. This thinking can then be specified in terms of learning input non-IIDness, entangled, and coupled representation learning, learning in and out-of distribution non-IIDness between training and test data, developing non-IID convolution, recurrency, attention, dropout and pooling mechanisms, supporting heterogeneous activation and transformation, learning input-neural feature couplings, and enabling non-IID transformation fusion, etc.

Non-IID statistical learning goes beyond classic probability assumptions and principles by incorporating probability couplings, non-IID priors, non-IID graphical models, non-IID sampling, non-IID inference, and non-IID relation learning, etc.

Non-IID vision learning may inspire the areas and perspectives of non-IID image analysis, non-IID action recognition, non-IID behavior detection, non-IID imitation learning, non-IID scene understanding, and non-IID visual analytics. Other topics include non-IID perception, identification, detection, recognition, and reidentification, and non-IID multi-model, multilabel, and multitask learning.

Non-IID document/text analysis and *non-IID natural language processing* inspect the heterogeneities and interactions between words, concepts, sentences, paragraphs, documents, topics, and sentiment, between questions and answers, between queries and answers for search and retrieval, and between semantics and syntax in texts or documents.

Non-IID behavior analytics is to quantify and analyze heterogeneities and interactions within the behaviors of a subject, between the behaviors of different actors, and between the behaviors in different scenarios. This thus forms the prospects of learning coupled behavior sequences, logical behavior couplings, statistical behavior couplings, group behavior couplings, coupled group behaviors, multiparty interactions, and multi-impact, -risk, or -utility behaviors.

Non-IID outlier detection is to identify, characterize, and formulate the non-IIDness of values, features, labels, or contexts; between inlying and outlying features, objects, or labels; and during outlying dynamics and to incorporate them into outlierliness scoring, and outlier detectors, etc. Exemplary topics and perspectives include learning value couplings and feature couplings for outlierliness scoring and outlier detection.

Non-IID recommender systems address user non-IIDness, item non-IIDness, user-item interactions, rating non-IIDness, context non-IIDness, and the non-IIDness during sessions, across domains, between groups, or over time, etc. Examples are coupled matrix factorization for coupled collaborative filtering, and CoupledCF by learning input-neural couplings in deep networks.

The abovementioned only illustrates some research perspectives of applying non-IID thinking to several existing learning systems. In fact, many other research prospects and opportunities exist when non-IIDness and non-IID thinking are considered both in the machine learning field and beyond. Other disciplines may also substantially benefit from incorporating non-IIDness and non-IID thinking into their existing IID or near-IID methodologies and frameworks,[10] for example, information theory, quantum information processing, frequentist, and Bayesian statistics, algebra and numerical computation, control theories, complexity science, management science, and decision science.

COMPLEXITY AND INTELLIGENCE

In reality, heterogeneity, and interaction are associated with the complexities of a system or dataset. Their quantification in non-IID systems thus contributes to advanced intelligence. This applies to general AI and intelligent systems, DeAI, and centralized AI (CeAI) systems.[11] Heterogeneity, interaction, complexity, and intelligence also form fundamental characteristics, properties, mechanisms, challenges, or prospects of any complex systems, problems, and applications,[9,19] in particular, in the digital era. Examples are various interactions, complexities, and intelligences in CeAI and DeAI and various sources, types, and properties of X-complexities, X-intelligence,[10] and X-interaction[8] in complex systems, behaviors, and data. It is thus important to think deeply and broadly about the non-IIDnesses of a system or its data from these perspectives.

In addition, heterogeneity, interaction, complexity, and intelligence of centralized and DeAI systems and settings not only share similarities but also own their respective significant differences. It is essential to identify, define, specify and address their similarities and discrepancies in CeAI and DeAI. In decentralized settings, such as edge intelligence and smart blockchain systems,[11] it is even more challenging to quantify their heterogeneities, interactions, complexities, and intelligence for DeAI. They are fundamental for creating tangible techniques and enablers to understand, develop and manage AI in decentralized organizations, environments, tasks, and settings.

Quantifying Complexity

There are nonuniform definitions and characterizations of the term *complexity*,[19] from various perspectives, on distinctive entities, and for specific purposes. However, several fundamental elements have been commonly agreed on or discussed in terms of complexity as follows:

1) *system* which is composed of multiple parts (or objects, entities, parties, etc.);
2) *environment* of the system (exterior to the system);
3) *interactions* between parts and between the system and its environment;
4) *feedback* (or history, memory, and response) from the interactions;
5) *properties* (or attributes) of the system and its parts, environment, interactions, feedback, or results; and
6) *results* (or state, consequence, and emergence) of the system. In addition, a system is associated with its behaviors and data, which may be complex as well.

Accordingly, systems may be categorized into simple systems or complex systems, open systems or closed systems, static systems or dynamic systems, small systems or large systems, reactive systems or proactive systems, responsive systems or adaptive systems, and organized systems or disorganized systems (or order versus chaos), etc. Two fundamental factors differentiating the two categories in each pair are heterogeneity and interaction and their properties and roles in the system. It is thus fair to say *heterogeneity and interaction determine the complexities of a system*.

Furthermore, the characterization of the complexities of a system is taken from the respective entities, aspects, properties, and their relations in a system, forming the X-complexities of a system. Consequently, we may discuss *X-complexities* in terms of, but not limited to, object complexity, domain complexity, environment complexity, behavior complexity, data complexity, evolution complexity, and consequence complexity,[10] and in particular, their heterogeneity and interaction complexities. Similarly, heterogeneity and interaction form fundamental perspectives to specify and quantify these aspects of complexities.

Specifically, *heterogeneity and interaction complexities* influence or determine the complexities of a system. The complexities of heterogeneities and interactions are further attributed to, driven, or determined by the aspects, properties (such as type, heterogeneity, dimensionality, hierarchy, granularity, and dynamics, etc.) of heterogeneity and interaction. Quantifying heterogeneities and interactions is thus to characterize their aspects, properties, and roles in a system in the context of the system, domain, environment, and objective, etc.

In quantification, each of the X-complexities is indicated, represented, or characterized by their respective aspects, properties, or characteristics. For example, *data complexities* may be characterized by data characteristics in terms of the largeness of scale, non-IIDness, high dimensionality, high order, unclear hidden structures, distributions, and relations in a dataset.[10]

Quantifying Intelligence

Over the 70 years of AI evolution, different types, levels, and generations of intelligence have been explored in the AI communities, forming X-intelligences.[10] Such *X-intelligences* evolve from object intelligence to symbolic intelligence, cognitive intelligence, connectionist intelligence, learning intelligence, behavioral intelligence, natural intelligence, networking intelligence, data intelligence, social intelligence, algorithmic intelligence, emotional intelligence, ethical intelligence to humanlike or human-level intelligence, system intelligence, and metasynthetic intelligence.[20] A complex system may hold, take advantage of, or fulfill some of these X-intelligence, which are often synthesized in system-specific manners. For example, the intelligence of a drone relies on its capabilities and capacity of offline and online perceivability, analysis, learning, communication, reflection, and intervention. An automated adaptive drone goes beyond its predefined and reactive programming but makes autonomous detection, prediction, recommendation, intervention, or reflection on what it perceives, memorizes and learns from history, context, feedback, and performance, and how it optimizes, etc.

Specifically, quantifying the heterogeneities and interactions of a system contribute to system intelligence and quantifying intelligence. Heterogeneities and interactions are major factors causing, driving, or influencing the formation, form, strength, evolution, and emergence of X-intelligences. Problem-solving intelligent systems have to quantify heterogeneities and interactions and utilize heterogeneous and interactive problem-solving. Different X-intelligences are attributed to their responsive and/or proactive influential parties, types, hierarchy, heterogeneity, strength, dynamics, and feedback of heterogeneities and interactions. Understanding, characterizing, and managing X-intelligences must involve and quantify heterogeneity and interaction intelligence. For example, in developing algorithmic intelligence, a typical challenge is to identify and quantify heterogeneities and interactions in aspects, such as objects and distributions and their properties in the data, between aspects, and between properties of heterogeneity and interaction, as shown in Figure 3. Furthermore, *heterogeneity and interaction intelligence* present in different forms in CeAI and DeAI systems. For example, in CeAI, vertical interaction intelligence is focused, to concentrating on hierarchical cooperation, coordination, communication, and problem-solving. By contrast, in DeAI, horizontal interaction intelligence is centered, featuring edge-to-edge and end-to-end engagement, sharing, communication, and task undertaking.

Furthermore, the intelligence of a complex system and its problem-solving rely on the metasynthesis of various X-intelligences,[10] including heterogeneity intelligence and interaction intelligence. X-intelligence metasynthesis may take a qualitative-to-quantitative metasynthetic engineering process.[10] In achieving this, both target system complexities and intelligence and its problem-solving systems and intelligence are involved, interactive, and combined. For example, a programmed drone is embedded with chips, sensors, reactors, and algorithms to detect, analyze, react, and optimize its behaviors, data, and decisions per its prior knowledge, predefined programs, rules, strategies, and policies. These form its system intelligence. A smart autonomous drone may further actively and proactively develop its online rules, strategies, and policies over time, scenarios, and environmental change. This requires active and online problem-solving intelligence, including learning from data, behavior, and feedback.

CONCLUDING REMARKS

Non-IID thinking reflects the nature and complexities of the real world, which goes beyond i.i.d., a widely accepted assumption in classic and contemporary science, technology, and practice. Non-IID thinking advocates a genuine, deep, and wide understanding and characterization of the intricate and intrinsic reality and their complexities in the real world and of any complex systems. Investigating non-IIDness is not just fundamental for pursuing beyond-i.i.d. AI, data science, and machine learning but a general principle for developing *reality-level science and technology*. Specifically, significant knowledge gaps and opportunities exist in deeply exploring the diversified heterogeneities and interactions in a complex system and its behaviors, environment, and data. Non-IID thinking and informatics form an essential paradigm to manage such gaps and opportunities, where non-IID learning represents a new frontier of data-driven discovery.

ACKNOWLEDGMENTS

This work was supported by the Discovery under Grant DP190101079 and by Future Fellow under Grant FT190100734.

REFERENCES

1. G. Shorack, "Convergence of reduced empirical and quantile processes with application to functions of order statistics in the non-i.i.d. case," *Ann. Statist.*, vol. 1, no. 1, pp. 146–152, 1973.
2. R. Golden, *Statistical Machine Learning: A Unified Framework*. London, U.K.: Chapman & Hall/CRC, 2020.
3. X. Wang and S. Kadıoğlu, "Modeling uncertainty to improve personalized recommendations via Bayesian deep learning," *Int. J. Data Sci. Anal.*, pp. 1–11, 2021.
4. C. Wang, Z. She, and L. Cao, "Coupled attribute analysis on numerical data," in *Proc. IJCAI Int. Joint Conf. Artif. Intell.*, 2013.
5. L. Cao, "Non-IIDness learning in behavioral and social data," *Comput. J.*, vol. 57 no. 9, pp. 1358–1370, 2014.
6. L. Cao, "Non-IID recommender systems: A review and framework of recommendation paradigm shifting," *Engineering*, vol. 2, pp. 212–224, 2016.
7. "Beyond IID information theory workshop series." Accessed: Jul. 24, 2022. [Online]. Available: https://sites.google.com/view/beyondiid8/Home?authuser=0
8. L. Cao, "Coupling learning of complex interactions," *J. Inf. Process. Manage.*, vol. 51, no. 2, pp. 167–186, 2015.
9. X. Qian, J. Yu, and R. Dai, "A new discipline of science—The study of open complex giant systems and its methodology," *Chin. J. Syst. Eng. Electron.*, vol. 4, no. 2, pp. 2–12, 1993.
10. L. Cao, *Data Science Thinking: The Next Scientific, Technological and Economic Revolution*. Berlin, Germany: Springer, 2018.
11. L. Cao, "Decentralized AI: Edge intelligence and smart blockchain, metaverse, Web3 and DeSci," *IEEE Intell. Syst.*, vol. 37, no. 3, 2022.
12. "Coupling and interaction learning." Accessed: Jul. 24, 2022. [Online]. Available: https://datasciences.org/coupling-learning/

13. P. Kumar and R. Dohare, “An interaction-based method for detecting overlapping community structure in real-world networks,” *Int. J. Data Sci. Anal.*, vol. 14, no. 1, pp. 27–44, 2022.
14. C. Wang, F. Giannotti, and L. Cao, “Learning complex couplings and interactions,” *IEEE Intell. Syst.*, vol. 36, no. 1, pp. 3–5, Jan./Feb. 2021.
15. “Non-IID learning.” Accessed: Jul. 24, 2022. [Online]. Available: https://datasciences.org/non-iid-learning/
16. KDD2017 Tutorial, “Non-IID learning in Big Data.” Accessed: Jul. 24, 2022. [Online]. Available: https://www.youtube.com/watch?v=3RwyGoiYcLg and https://www.youtube.com/watch?v=So0GNQiF9vE
17. IJCAI2019 Tutorial, “Non-IID learning of complex data and behaviors.” Accessed: Jul. 24, 2022. [Online]. Available: https://datasciences.org/publication/Non-IID%20Learning-CAO-Full.pdf
18. KDD2022 Tutorial, “Shallow and deep Non-IID learning on complex data,” 2022. Accessed: Jul. 24, 2022. [Online]. Available: https://kdd.org/kdd2022/lectureTutorial.html
19. M. Mitchell, *Complexity: A Guided Tour*. New York, NY, USA: Oxford Univ. Press, 2011.
20. L. Cao, “A new age of AI: Features and futures,” *IEEE Intell. Syst.*, vol. 37, no. 1, pp. 25–37, Jan./Feb. 2022.

LONGBING CAO is a professor and an ARC future fellow (professorial level) with the University of Technology Sydney, Sydney, NSW, 2007, Australia. His research interests include AI, data science, machine learning, behavior informatics, and their enterprise applications. He is a senior member of IEEE. Contact him at longbing.cao@uts.edu.au.