SCIENCE CHINA
Information Sciences

• PERSPECTIVE •

March 2024, Vol. 67, Iss. 3, 136201:1–136201:3
https://doi.org/10.1007/s11432-023-3930-9

Systems science in the new era: intelligent systems and big data

Wenwu YU[1,2*], Duxin CHEN[1], Hongzhe LIU[1], He WANG[1,2], Jinde CAO[1,2], Zengru DI[3], Xiaojun DUAN[4], Bin JIA[5], Xiaodong DING[6] & Yiguang HONG[7]

[1]*School of Mathematics, Southeast University, Nanjing* 210096*, China;*
[2]*Purple Mountain Laboratories, Nanjing* 211111*, China;*
[3]*International Academic Center of Complex Systems, Beijing Normal University, Zhuhai* 519087*, China;*
[4]*College of Science, National University of Defense Technology, Changsha* 410073*, China;*
[5]*School of Systems Science, Beijing Jiaotong University, Beijing* 100044*, China;*
[6]*Business School, University of Shanghai for Science and Technology, Shanghai* 200093*, China;*
[7]*Department of Control Science and Engineering, Tongji University, Shanghai* 201804*, China*

Received 26 July 2023/Accepted 14 September 2023/Published online 25 January 2024

This study introduces the emergence, research challenges, and future development of intelligent systems and big data (ISBD) and discusses the significance of ISBD to systems science. To this end, an overview of systems science is first presented, followed by the conclusion that the development of artificial intelligence (AI) has promoted the derivation and development of intelligent systems (ISs). Then, the fact that ISBD has been a new branch of the systems science discipline is well explained, and the main research challenges involved in ISBD are extensively demonstrated. Finally, the idea of leveraging networked collective intelligence (NCI) to address the above challenges and advance the future development of ISBD is presented.

Overview of systems science. A system is an organic whole with specific functions composed of several components that interact and depend on each other. Systems science is a discipline that studies the general relationships among structures, states, environments, and functions of systems, and the universal laws on the evolution and regulation of systems. One of the basic principles of systems science is that the structure, state, and environment of a system jointly determine its function. Therefore, enhancing the individual functions of the components, optimizing the interaction of the components, and promoting interaction between the system and the external environment are effective ways to improve the system's function. In particular, systems theory generally comprises the following five aspects: method, evolution, cognition, regulation, and practice theories [1]. Method theory guides the selection of specific methods in systems research. Evolution theory mainly explores the emergence and evolution laws between hierarchies and the corresponding functions of systems in time and space. Furthermore, cognition theory focuses on the theories and methods of cognition and learning about the systems themselves and their surroundings, whereas regulation theory mainly studies the general relationships between the structures, states, and functions of the systems and their regulation methods. Additionally, practice theory focuses on the application of systems to specific disciplines. Classical studies on systems theory focus on the relationship between the structures, states, and functions of the systems, and there is relatively less attention paid to the optimization of the system functions based on the cognition and learning of the surrounding environments.

The emergence of ISs. Nowadays, inspired by the generation and access to big data, the development of AI has greatly improved the ability of a system to comprehend and learn about its surrounding environment, as well as its ability to understand and make decisions about the problems involved in the system. Therefore, AI can greatly enrich the intelligent connotation of systems theory and vigorously promote the derivation and development of ISs. Overall, ISs emerge with time.

ISBD—the new branch of systems science. Generally, ISs refer to computer systems that can produce intelligent human behaviors. The intelligence of systems is mainly embedded in the abilities of automatically acquiring and applying knowledge, thinking and reasoning, problem-solving, and automatic learning to be able to self-regulate [2]. Therefore, an IS necessarily contains the following elements: a sensor unit for acquiring information, a unit for retaining the received information as knowledge and storing the knowledge, a unit for acting based on the knowledge, and a functional element for the interconnection of the above three units [3].

A significant difference between ISs and traditional systems is that ISs can interact with their environments and eventually adapt to those environments; i.e., they have the abilities of automatic organization and adaptability. Interactions involve perceiving, learning, reasoning, judging, and acting accordingly. However, training data is the only way for the systems to learn from human and environmental input. Thus, a major and effective interaction mode for ISs is to collect big data in the environment, perform deep learning, causal inference, modeling and identification, and other

* Corresponding author (email: wwyu@seu.edu.cn)

technologies for reasoning, and then make decisions and implement regulations based on optimization, game, control, and other technologies. Therefore, the development of IS theory is closely related to big data. Nowadays, ISBD has developed into a new branch of systems science based on big data processing, optimization, game theory, and control theory.

Challenges for research on ISBD. Since the autonomous regulation of ISs requires data-based knowledge, AI methods are usually employed to solve ISs problems. Therefore, compared with the solution mode adopted by traditional systems, the mode for ISs has the obvious characteristics that problem solutions mainly depend on data processing to obtain knowledge, and problem solutions often involve exponential computational complexity. Therefore, with the increasing complexity of task scenarios, the scale and difficulty of data processing, as well as the difficulty and complexity of decision-making, will increase sharply. Generally, a system expressing characteristics such as non-linearity, complexity, self-organization, emergence, heterogeneity, and involving big data can be referred to as a complex system.

On the one hand, for various existing complex systems, the training and learning challenges are mainly posed by large-scale complex multisource data. When dealing with complex networked systems, the complexity of the training and learning of data will increase greatly. These challenges mainly include the high training cost arising from the vast amount of data, the instability of training outcomes due to the uneven distribution of multiple data sources, the difficulty in achieving accurate learning outcomes due to the nonlinear spatiotemporal coupling characteristics of complex data, and the prolonged training time required for high-dimensional data representation learning. Traditional AI methods, even the current hot AI large models, constrained by the centralized date training framework and strong assumptions on data distribution, fail to meet the lightweight, interpretable, reliable, and highly generalizable requirements for effectively learning large-scale complex multisource data. Consequently, they are unable to adequately support ISs in handling complex tasks. From the data application perspective, the following critical challenges require urgent attention.

Structural inference of complex systems. Inferring the networked structure of complex systems based on observation data has always been a hot topic and a foundation for understanding complex systems. The bottleneck lies in the precise inference of the Markov equivalence class. Traditional correlation or causality methods cannot achieve the stability of discovering complex system structures, and the inference accuracy, confidence interval, high-order structure extraction, and uniqueness cannot be theoretically proven. However, a high-efficiency structural inference algorithm with a lower computation burden is urgently needed for large-scale systems.

Modeling and identification of complex systems. The complexity of large-scale systems poses significant challenges in system modeling that requires only expertise. Modeling ISs must consider both knowledge-based and data-driven fusion. However, heterogeneous dynamics, networked coupling, switching topology, large scale, and other characteristics make it difficult to infer and reconstruct the system function. Accordingly, traditional system modeling and identification methods are no longer applicable in cases with the absence of prior knowledge. Developing an interpretable system reconstruction and identification method with an adaptive selection of basis functions and a fast optimization process is a future trend.

Prediction and generation of complex systems. Shortcomings include slow convergence, weak interpretability, and low robustness in implementing data generation and prediction tasks of complex systems using AI methods. More challenges lie in out-of-distribution generalization and prediction with observational data distribution. Therefore, new methods for knowledge-driven and data-driven integration need to be developed, especially the fusion of multimodal and distributed data, interpretability, reliability, counterfactual inference, and computational power reservation of AI large models.

On the other hand, for complex task scenarios, optimization, game, and control problems will inevitably be very complicated, such as optimization problems involving multiple objective functions, nonconvex objective functions, mixed-integer decision variables, games with large-scale players, and heterogeneous and diverse controlled subjects. Due to the centralized solution manner, traditional optimization, game, and control algorithms are inefficient, making them unable to obtain a decision and regulation scheme that meets the accuracy requirement in a limited time. Therefore, traditional optimization, game, and control algorithms cannot meet the real-time and accuracy requirements of decision-making and regulation, supporting the intelligent needs of ISs. From the perspective of decision-making and regulation tasks, the following key challenges require addressing urgently.

Nonconvex optimization. When solving nonconvex optimization problems, there are multiple local optimal solutions because of the nonconvexity of the objective and constraint functions. Nowadays, the local optimal solutions are usually achieved using existing methods because gradient-based methods with limited exploration capabilities always get trapped in local minima.

Multi-objective optimization. For multi-objective optimization problems, a general approach is to transform them into single-objective problems using scalarization or constraint-based methods. However, the individual search property in these approaches makes it challenging to effectively balance conflicts among multiple objectives. Consequently, it is difficult to efficiently obtain all Pareto optimal points. Furthermore, as the dimensionality of the objectives increases, the conflicts among them become more pronounced; thus, only local optima are usually achieved using the existing methods.

Mixed-integer programming. To address mixed-integer programming problems, branch and bound or cutting-plane methods are usually employed. However, the ergodicity of these methods causes an extensive search domain. As the scale of the problem increases, the required computational resources increase exponentially. Consequently, due to the constraints imposed by computation time and resources, the suboptimal solution is usually achieved using the existing centralized methods.

Intelligent game. The search for equilibria of the games should use complete information for all players. However, when large-scale players are involved, the central unit introduced for the traditional equilibrium searching method needs to undertake extensive information storage and computational tasks, which greatly increases the time cost of searching equilibrium.

Intelligent control. In complex task scenarios, as the scale of controlled subjects increases, it is difficult to maintain efficient and robust coordination of multiagent systems due to the communication limitations, conflicting cooperative objectives, and uncertainties with traditional centralized control algorithms.

Networked collective intelligence for solving these challenges. With the development of society and the progress of time, the collaborative work of large-scale ISs for complex tasks will be inevitable, thereby increasing the difficulty of data training and learning and the difficulty of efficient and accurate solutions to the problems. Collective intelligence (CI) has been well studied and applied in different research fields [4]. In the traditional sense, CI mainly focuses on heuristic algorithms, such as ant colony and particle swarm algorithms, which have been clearly listed as one of the eight basic theories of AI in the new era in the "New Generation Artificial Intelligence Development Plan" issued by The State Council in 2017. Different from traditional CI, networked collective intelligence (NCI) in the new era has a richer definition. Based on the concept of network, it is studied as a comprehensive theoretical framework, including several basic theories pointed out in the "New Generation Artificial Intelligence Development Plan," such as CI, autonomous collaborative control and optimal decision-making, and advanced machine learning. The development of NCI provides new insight for solving the above bottleneck problems on data learning, decision-making, and regulation. NCI can effectively leverage the role of the group components, allowing the learning task of large-scale complex data, decision-making, and regulation tasks to be separated. This separation enables the collaborative resolution of multiple simple subtasks, ultimately resulting in an efficient and accurate solution for the entire complex learning, decision-making, and regulation process.

Nowadays, research on ISBD in the sense of systems science is closely related to AI, complex systems, and other interdisciplinary technologies. The theories and technologies of big data and decision-making, as well as the regulation of ISs in the sense of NCI, will greatly enrich the connotation of systems theory. Furthermore, the big data, decision-making, and regulation theories and technologies of ISs in the sense of AI should focus on the development of the following parts: complex ISs-based structural inference, prediction and generative model, systems modeling and identification, large-scale pretraining model, and reinforcement learning theories and methods; NCI-based cooperative (convex or nonconvex) optimization, large-scale multi-objective optimization, large-scale mixed-integer programming, multiagent game theories and methods, and cooperative control.

In summary, AI and NCI theories are of great significance for systems science to guide the development of systems intelligence in the new era. The development of ISBD theory opens numerous avenues for future interdisciplinary investigations on systems science from the theoretical and application perspectives. It also promotes the gradual development of systems science into a base-solid and intersection-strong interdiscipline integrating mathematics, AI, electrical engineering, information science, network security, medicine, transportation, computer, automation, mechanical engineering, and many other disciplines. Overall, ISBD, as a new branch of systems science, will achieve great development in the new era.

Acknowledgements This work was supported by National Key R&D Program of China (Grant No. 2022ZD0120001) and National Natural Science Foundation of China (Grant No. 62233004).

References

1 Guo L. What is systematology (in Chinese). J Syst Sci Complex, 2016, 36: 291–301
2 Shin Y C, Xu C. Intelligent Systems: Modeling, Optimization, and Control. Boca Raton: CRC Press, 2017
3 Kaspar C, Ravoo B J, van der Wiel W G, et al. The rise of intelligent matter. Nature, 2021, 594: 345–355
4 Warnat-Herresthal S, Schultze H, Shastry K L, et al. Swarm learning for decentralized and confidential clinical machine learning. Nature, 2021, 594: 265–270