Biograph: Dr./IRSS Fellow/CAAI Fellow, Duoqian Miao is a Professor and the Vice Director of the Key Laboratory of Embedded System and Service Computing of Ministry of Education, Tongji University, China. His research interests include Artificial Intelligence, Machine Learning, Big Data Analysis, Granular Computing and Rough Sets. Since 1997, he has published over 180 research papers (e.g., IEEE TKDE, IEEE TFS, IEEE TIFS, IEEE TC, IEEE TIP, IEEE TCSVT) in refereed journals and conferences (e.g., AAAI, NeurIPS, ICPR, ICCV, WACV). In addition, he has published 10 academic books and over 20 patents. He serves as AC of Information Sciences, International Journal of Approximate Reasoning, CAAI Transactions on Intelligence Technology, and Chinese Journal of Computer Research and Development, etc. He is the president of International Rough Sets Society (2020- 2022), vice president of Shanghai Computer Federation(2018-2022), vice president of Shanghai Association for Artificial Intelligence, Chair of Granular Computing and Knowledge Discovery Technical Committee of CAAI(2012-2022) , a distinguished member of China Computer Federation(CCF).

Title: The Contributions of Rough Set Theory to Artificial Intelligence

Abstract: After more than 70 years of development, artificial intelligence is moving towards the stage of general artificial intelligence, which has a profound impact on human daily work and life. Rough set theory emerged in the early 1980s and, after 40 years of development, has played an irreplaceable role in handling imprecise, inconsistent, and incomplete data [1]. As we all know, artificial intelligence mainly studies knowledge representation, knowledge acquisition, and knowledge application, which means that the research of artificial intelligence revolves around knowledge. But what is knowledge? Unfortunately, textbooks on artificial intelligence do not provide a formal definition of knowledge [2]. This report will introduce the understanding and formal definition of knowledge in rough set theory, as well as its significant contributions to artificial intelligence, hoping to provide inspiration for future research in artificial intelligence.

The recent emergence of ChatGPT and Sora aroused the attention of large models, which has promoted the advent of general artificial intelligence [3,4]. As we stand on the precipice of a new era in artificial intelligence, the foundational role of knowledge representation in both acquisition and application cannot be overstated. Currently, the field leans towards end-to-end methodologies characterized by simplification and automated computation for knowledge representation. However, this approach is not without its Achilles' heel: a conspicuous lack of interpretability. Knowledge acquisition, the technological linchpin for extracting valuable insights from data, is predominantly facilitated by machine learning models, ranging from classical algorithms like K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) to the more contemporary deep learning paradigms. Furthermore, expert systems have emerged as pivotal instruments for the application of knowledge, permeating various domains within the AI landscape. According to these developments, a formal definition of "knowledge" is not merely desirable but imperative for the continued advancement of artificial intelligence.

Rough sets (RS), introduced by Z. Pawlak in 1982, offer an efficient approach for handling imprecise, inconsistent, and incomplete data [1]. It provides formal knowledge

representation that is understood as the ability to classify objects. In this theory, the knowledge is analogized from knowledge to object partition and further equivalence relation, which can be represented as $K = (U, \mathbb{R})$, where $U = \{x_1, x_2, \cdots, x_n\}$ is the universe and $\mathbb{R} = \{R_1, R_2, \cdots, R_m\}$ is a family of equivalence relations [2]. This is an algebraic representation of knowledge. Subsequently, attribute reduction based on positive region is proposed to obtain an important feature subset. Nevertheless, it is only a decidable definition, and selecting a minimum subset is an NP-hard problem. To address this challenge, we have proposed a heuristic algorithm for attribute reduction based on attribute significance [2]. Moreover, Skowron introduced the discernibility matrix for attribute reduction [5]. Similarly, it is also a decidable definition. Consequently, we have developed a reduction algorithm that is attuned to the attribute frequency within the discernibility matrix [2]. Crucially, we have introduced the information representation of knowledge and defined the lower and upper approximation sets based on it. Some heuristic methods based on information entropy, conditional entropy, and mutual information are proposed for attribute reduction [2,6]. More recently, we integrated zentropy thought to define a zentropy-based uncertainty measure and have adapted it to computing reductions [7]. According to the above analysis, we divided these methods into algebraic representation and information representation, which can be seen in Fig. 1.

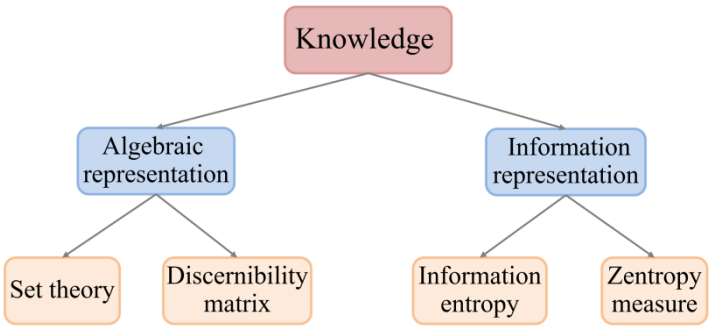

Fig.1 The relationship of knowledge representation in rough sets

Concept is the fundamental unit for the recognition and acquisition of knowledge [8]. In machine intelligence, the processing of precise concepts often surpasses human capabilities. Currently, how to handle imprecise concepts plays a vital role in artificial intelligence [9,10]. Philosophical scholars contend that a concept is imprecise if its boundary region is nonempty [11]. However, it is only a semantic limitation that cannot be recognized by computers, bring challenges of imprecise concepts in computational recognition. Rough set theory provides a mathematical framework for calculating the boundary region of imprecise concepts. It approximates imprecise concepts by two precision sets (lower and upper approximate sets) and describes their uncertainty via boundary region [2,12,13]. Subsequently, the positive, negative, and boundary regions are induced by rough sets, and each region has different actions. Based on this, three-way decision (3WD) is introduced by Y. Yao, which is an exploration of rough set theory and has successfully applied to various scenarios [14].

Rough set theory is an efficient machine learning model for processing qualitative data [15]. It provides a multi-granulation representation, learning, and reasoning methods for knowledge in artificial intelligence. In essence, rough set theory has made significant contributions to AI through three pivotal aspects: the formal representation of knowledge, the

computation of boundary regions for imprecise concepts, and the multi-granulation knowledge representation and reasoning methods. As a holistic, data-driven model, rough set theory shows robust potential prospects and developmental avenues in artificial intelligence.

References

[1] Z. Pawlak, Rough Sets, *International Journal of Computer and Information Sciences*, vol. 11, pp. 341-356, 1982.

[2] D. Miao, Rough set theory and its applications in machine learning, Doctoral dissertation, *Institute of Automation, Chinese academy of sciences*, 1997.

[3] J. Lechien, G. Briganti, L. Vaira, Accuracy of ChatGPT-3.5 and -4 in providing scientific references in otolaryngology-head and neck surgery, *European Archives of Oto-Rhino-Laryngology*, vol. 281, pp. 2159 – 2165, 2024.

[4] L. Fan, C. Guo, Y. Tian, et al., Sora for foundation robots with parallel intelligence: three world models, three robotic systems, *Frontiers of Information Technology & Electronic Engineering*, doi: https://doi.org/10.1631/FITEE.2400144, 2024.

[5] A. Skowron, R. Cecylia, The discernibility matrices and functions in information systems, *Intelligent decision support: handbook of applications and advances of the rough sets theory. Dordrecht: Springer Netherlands*, pp. 331-362, 1992.

[6] K. Yuan, W. Xu, D. Miao, A local rough set method for feature selection by variable precision composite measure, *Applied Soft Computing*, vol. 155, pp. 111450, 2024.

[7] K. Yuan, D. Miao, Y. Yao, et al., Feature selection using zentropy-based uncertainty measure, *IEEE Transactions on Fuzzy Systems*, vol. 32, no. 4, pp. 2246-2260, 2024.

[8] D. Guo, W. Xu, W. Ding, et al., Concept-cognitive learning survey: mining and fusing knowledge from data, *Information Fusion*, vol. 109, pp. 102426, 2024.

[9] D. Slezak, R. Glick, P. Betliski, et al., A new approximate query engine based on intelligent capture and fast transformations of granulated data summaries, *Journal of Intelligent Information Systems*, vol. 50, pp. 385C414, 2018.

[10] Y. Yao, Interval-set algebra for qualitative knowledge representation, *the 5th International Conference on Computing and Information*, pp. 370-375, 1993.

[11] F. Ramsey, Review of Ludwig Wittgenstein's "Tractatus Logico-Philosophicus", *Mind*, vol. 32, no. 128, pp. 465-478, 1923.

[12] P. Lingras, M. Chen, D. Miao, Rough cluster quality index based on decision theory, *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, pp. 1014-1026, 2009.

[13] J. Yao, A. Vasilakos, W. Pedrycz, Granular computing: perspectives and challenges, *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 1977-1989, 2013.

[14] Y. Yao, Three-way decisions with probabilistic rough sets, *Information sciences*, vol. 180, no. 3, pp.341-353, 2010.

[15] D. Miao, Y. Zhao, Y. Yao, et al., Relative reducts in consistent and inconsistent decision tables of the Pawlak rough set model, *Information Sciences*, vol. 179, no. 24, pp. 4140-4150, 2009.