The Contributions of Rough Sets to Artificial Intelligence

Duoqian Miao

Tongji University

Shanghai, P.R. China

May 18, 2024

General Conference Chairs

•**Pawan Lingras** (Saint Mary′s University, Canada)
•**Dominik Ślęzak** (QED Software, Poland)
•**JingTao Yao** (University of Regina, Canada)

Program Committee Chairs

•**Mengjun Hu** (Saint Mary′s University, Canada)
•**Chris Cornelis** (Ghent University, Belgium)
•**Yan Zhang** (California State University San Bernardino, USA)

Honorary Chairs

•**James Peters** (University of Manitoba, Canada)
•**Andrzej Skowron** (Systems Research Institute of Polish Academy of Sciences, Poland)
•**Yiyu Yao** (University of Regina, Canada)

Steering Committee Chairs

•**Witold Pedrycz** (University of Alberta, Canada)
•**Roman Słowiński** (Poznań University of Technology, Poland)
•**Guoyin Wang** (Chongqing University of Posts and Telecommunications, China)

Outline

- Artificial Intelligence
- Rough Set Theory
- Contributions of RS to AI
- Conclusions

Artificial Intelligence

- **ChatGPT:** Chat Generative Pre-trained Transformer, a chat robot program developed by OpenAI. (2022)

- **Sora:** An AI model that can create realistic and imaginative scenes from text instructions. (2024)

Sora

[1] https://openai.com/index/chatgpt
[2] https://openai.com/index/sora

Artificial Intelligence

- **Knowledge Representation:** neural networks
- **Knowledge Acquisition:** machine learning methods
- **Knowledge Application:** expert systems

Artificial Intelligence

- Artificial Intelligence research has been based on “**Knowledge**".

- What is Knowledge?

- Unfortunately: there has been no **formal definition of Knowledge** in AI textbooks.

Outline

- Artificial Intelligence
- Rough Set Theory
- Contributions of RS to AI
- Conclusions

Rough Set Theory

- **Rough sets** (RS) was introduced by Prof.Pawlak in 1982.
- It is an efficient method to handle **imprecise**, **inconsistent** and **incomplete** data based on set theory.
- It has a natural advantage in processing **qualitative data**.

[1] Z. Pawlak, "Rough Sets," in *International Journal of Computer and Information Sciences*, vol. 11, pp. 341-356, 1982.
[2] R. Slowinski, C. Zopounidis, "Application of the rough set approach to evaluation of bankruptcy risk," in *Intelligent Systems in Accounting, Finance and Management*, vol. 4, no. 1, pp.27-41, 1995.

Rough Set Theory

- **Input data** in rough sets

Fact	Condition attributes			Decision attribute	Support
	Weather	Road	Time	Accident	
1	misty	icy	day	yes	6
2	foggy	icy	night	yes	8
3	misty	not icy	night	yes	5
4	sunny	icy	day	no	55
5	foggy	not icy	dusk	yes	11
6	misty	not icy	night	no	15

Rough Set Theory

- **Output results** in rough sets

Decision Rules	cer.
1. $(Weather, misty) \wedge (Road, icy) \rightarrow (Accident, yes)$	1.00
2. $(Weather, foggy) \rightarrow (Accident, yes)$	1.00
3. $(Weather, misty) \wedge (Road, not\ icy) \rightarrow (Accident, yes)$	0.25
4. $(Weather, sunny) \rightarrow (Accident, no)$	1.00
5. $(Weather, misty) \wedge (Road, not\ icy) \rightarrow (Accident, no)$	0.75

Rough Set Theory

- Contor Georgy: **Set** Theory (1883)

- Lotfi Zadeh: **Fuzzy Set** Theory (1965)

- Zdzislaw Pawlak: **Rough Set** Theory (1982)

Rough Set Theory

- **Formal representation of knowledge in RS**

$$\mathbf{K=(U,R)}$$

Where $U = \{x_1, x_2, \cdots, x_n\}$ is the universe,

$R = \{R_1, R_2, \cdots, R_m\}$ is a family of equivalence relations.

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

The First Contribution

- This marks the first significant contribution of RS to AI, bridging the gap in the formal definition of knowledge in AI.

Rough Set Theory

- What is a concept ?

Concept intent ; concept extent

- For a concept, how to formally represent it ?

a Concept $\cong$ a Subset

Rough Set Theory

- **What is a imprecise concept ?**

- Philosophical viewpoint

A concept is imprecise, if its boundary region is nonempty.

- **Rough concept**

How to represent it?

[1] Ramsey, Frank P. "Tractatus logico-philosophicus." pp. 465-478, 1923.

[2] L. Zadeh, "Fuzzy sets," in *Information and Control*, vol. 8, no. 3, pp. 338-353, 1965.

[3] J. Yao, A. Vasilakos, W. Pedrycz, "Granular computing: perspectives and challenges," in *IEEE Transactions on Cybernetics*, vol. 43, pp. 1977-1989, 2013.

Rough Set Theory

Rough set model

Let $DIS = (U, C \cup D, V, f)$ be a decision table. For $X \subseteq U, B \subseteq C$, $B(x)$ is the equivalence class of x on B. The **lower and upper approximations** of X on B are defined as follows:

$$B_*(X) = \bigcup_{x \in U} \{B(x) | B(x) \subseteq X\},$$

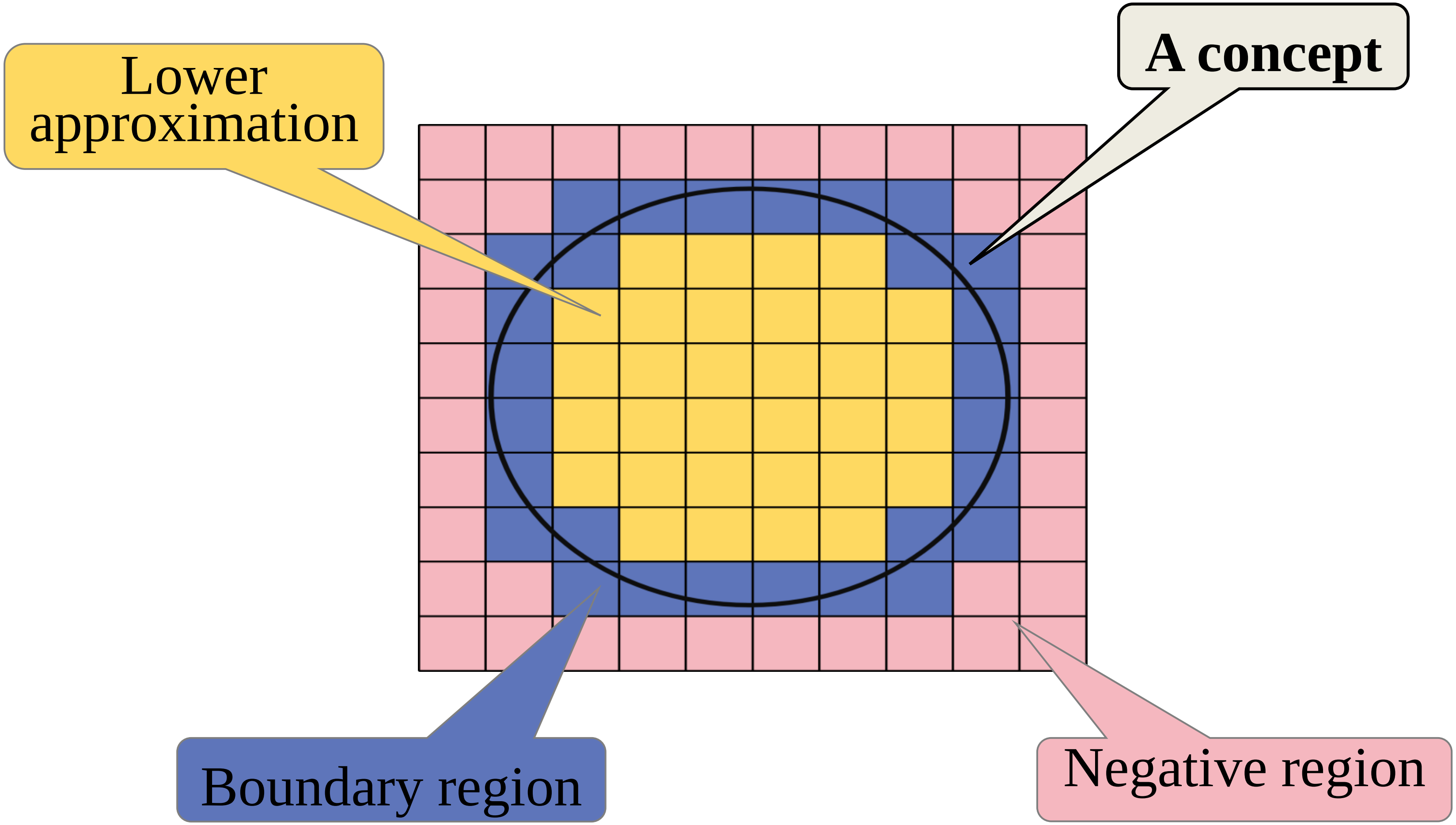
$$B^*(X) = \bigcup_{x \in U} \{B(x) | B(x) \cap X \neq \emptyset\}.$$

Then, the **boundary region** of X is $BN_B(X) = B^*(X) - B_*(X)$, and ***X* is rough** when **the boundary region is nonempty**.

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

Rough Set Theory

Rough set model



The Second Contribution

- **Rough Set Theory provides a method to calculate the boundary of a concept, realizing the computability of the philosopher's decidable definition.**

Rough Set Theory

Attribute reduction

Let $DIS = (U, C \cup D, V, f)$ be a decision table. For $X \subseteq U$, the postive region of D is defined as follows:

$$POS_B(D) = \cup_{X \in U/D} B_*(X).$$

For $b \in B$, the attribute b is ***indispensable*** in B, if

$$POS_{B-\{b\}}(D) \neq POS_B(D).$$

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

Rough Set Theory

Attribute reduction

Let $DIS = (U, C \cup D, V, f)$ be a decision table. For $B \subseteq C$, the B is a reduct, if

(1) $POS_{\mathrm{B}}(D) = POS_{\mathrm{B}}(D)$;

(2) For $b \in B$, each attribute is indispensable in B.

Note that:

- *This is from an algebra viewpoint, it is a decidable definition.*
- *The search of minimum reduction is NP-hard problem.*

Rough Set Theory

Reduction results --- 1

Fact	Condition attributes		Decision attribute	Support
	Weather	Road	Accident	
1	misty	icy	yes	6
2	foggy	icy	yes	8
3	misty	not icy	yes	5
4	sunny	icy	no	55
5	foggy	not icy	yes	11
6	misty	not icy	no	15

Rough Set Theory

Reduction results --- 2

Fact	Condition attributes		Decision attribute	Support
	Weather	Time	Accident	
1	misty	day	yes	6
2	foggy	night	yes	8
3	misty	night	yes	5
4	sunny	day	no	55
5	foggy	dusk	yes	11
6	misty	night	no	15

Rough Set Theory

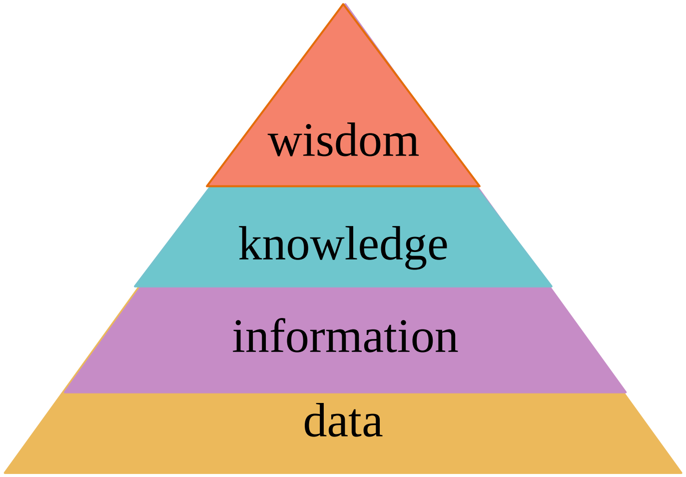
Reduction results --- decision rules

Decision Rules	cer.
1. $(Weather, misty) \wedge (Road, icy) \to (Accident, yes)$	1.00
2. $(Weather, foggy) \to (Accident, yes)$	1.00
3. $(Weather, misty) \wedge (Road, not\ icy) \to (Accident, yes)$	0.25
4. $(Weather, sunny) \to (Accident, no)$	1.00
5. $(Weather, misty) \wedge (Road, not\ icy) \to (Accident, no)$	0.75

The third Contribution

RS as a machine learning model

It is an efficient tool for handling **qualitative data**, which provides **multi-granularity representation**, **learning**, and **reasoning** model in artificial intelligence.



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

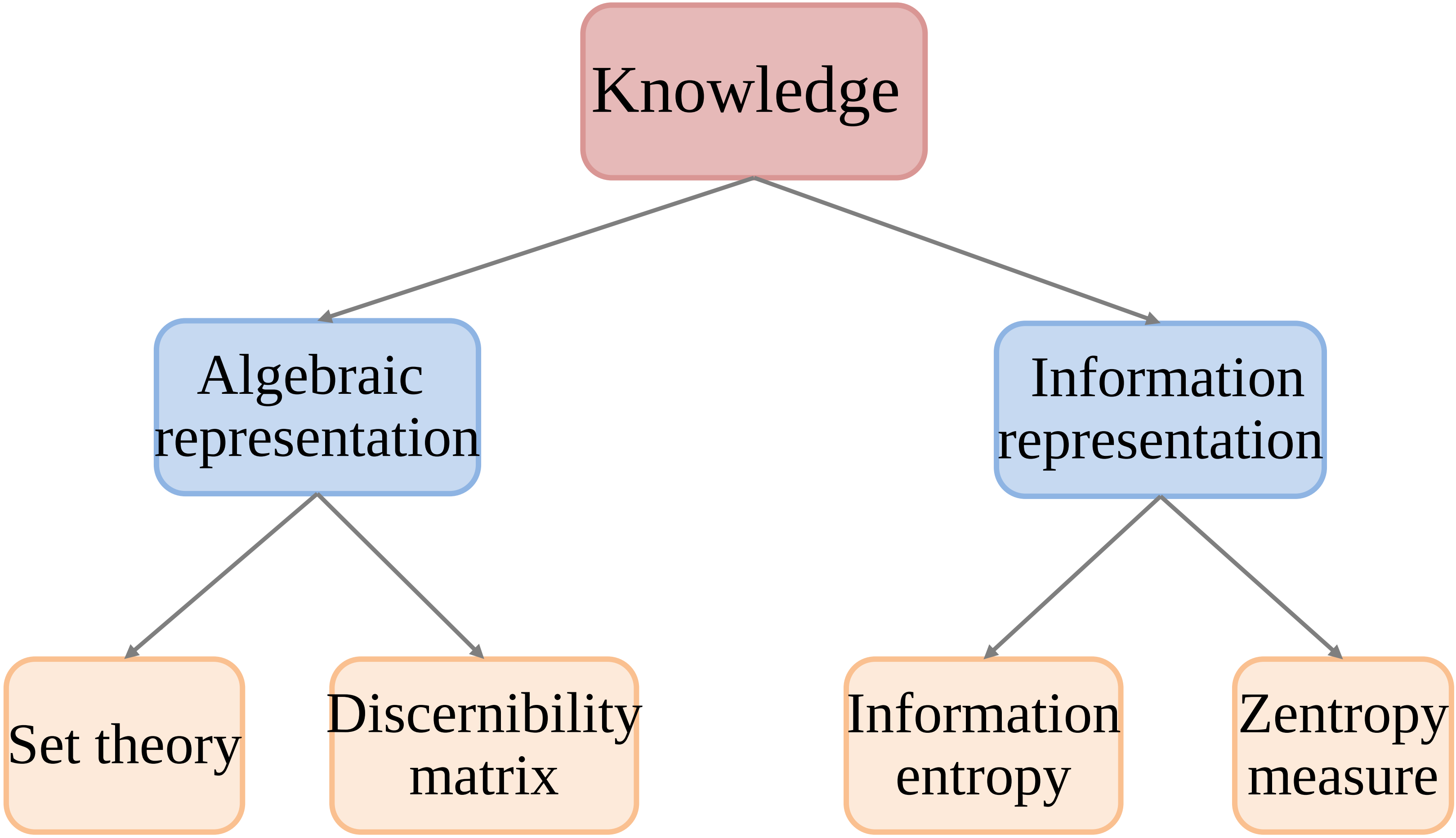
[2] D. Slezak, R. Glick, P. Betliski, et al., A new approximate query engine based on intelligent capture and fast transformations of granulated data summaries, in *Journal of Intelligent Information Systems*, vol. 50, pp. 385-414, 2018.

Outline

- Artificial Intelligence
- Rough Set Theory
- Contributions of RS to AI
- Conclusions

On Knowledge Representation

- Knowledge representation



Algebraic Representation

- **Discernibility matrix**: was proposed by Prof. **Skowron** in 1992.

Let $DIS = (U, C \cup D, V, f)$ be a decision table. The discernibility matrix $M = [m_{ij}]_{n \times n}$ is defined by the following equation.

$$m_{ij} = \begin{cases} \{c \in C\}, c(x_i) \neq c(x_j)\ and\ D(x_i) \neq D(x_j) \\ \emptyset\ , \qquad otherwise \end{cases}$$

[1] A. Skowron, R. Cecylia, “The discernibility matrices and functions in information systems,” in *Intelligent decision support: handbook of applications and advances of the rough sets theory. Dordrecht: Springer Netherlands*, pp. 331-362, 1992.
[2] J. Wang, D. Miao, Y. Zhou, ”A review on rough set theory and its applications,” in *Pattern Recognition and Artificial Intelligence*, vol. 9, no. 4, pp. 337-344, 1996.

Algebraic Representation

- **Attribute Reduction Based on the Discernibility matrix**

- Let M is the iscernibility matrix of $DIS = (U, C \cup D, V, f)$, the $B \subseteq C$ is a reduct , if

  (1) $\forall \emptyset \neq m_{ij} \in M, B \cap m_{ij} \neq \emptyset$;

  (2) $\forall b \in B, B - \{b\}$ does not satisfy (1).

Note that:

This is merely a decidable definition for attribute reduction.

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

[2] J. Wang, D. Miao, Y. Zhou, ”A review on rough set theory and its applications,” in *Pattern Recognition and Artificial Intelligence*, vol. 9, no. 4, pp. 337-344, 1996.

Algebraic Representation

Algorithm 1 Attribute reduction based on discernibility matrix.

Input: A decision table $DIS = (U, C \cup D, V, f)$.

Output: A reduct subset of DIS.

Step 1. Compute the core $C_0 = CORE_D(C)$ and $R = C_0$;

Step 2. Obtain $Q = \{m_{ij}: m_{ij} \cap R \neq \emptyset, i \neq j, i, j = 1,2 \cdots, n\}$,

$M = M - Q, B = A - R$;

Step 3. $\forall a_k \in B$, compute its frequency in M and set $p(a_q) = max_k\{p(a_q)\}$;

Step 4. $R \leftarrow R \cup \{a_q\}$;

Step 5. Repreat above process until $M = \emptyset$.

Step 6. Return R.

[1] J. Wang, R. Wang, D. Miao, Meng Guo, et al., "Data enriching based on rough set theory," in *Chinese J. Computers*, vol. 21, no. 5, pp. 393-400, 1998 .

Information Representation

- Let $DIS = (U, C \cup D, V, f)$ be a decision table. For $B, P \subseteq C$,

$$U/B = \{X_1, X_2, \cdots, X_s\}, \quad U/P = \{Y_1, Y_2, \cdots, Y_q\}.$$

The **probability distribution** of B and P is defined as follows.

$$[X; B] = \begin{bmatrix} X_1 & X_2 & \cdots & X_s \\ P(X_1) & P(X_s) & \cdots & P(X_s) \end{bmatrix},$$

$$[Y; P] = \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_q \\ P(Y_1) & P(Y_2) & \cdots & P(Y_q) \end{bmatrix},$$

where $P(X_i) = \frac{|X_i|}{|U|}, i = 1,2, \cdots, s,\ P(Y_j) = \frac{|Y_j|}{|U|}, j = 1,2, \cdots, q.$

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

Information Representation

- Let $DIS = (U, C \cup D, V, f)$ be a decision table. For $B, P \subseteq C$. The **information entropy** of B, **conditional entropy** of B related to P, and **mutual information** of B and P are defined as follows.

$$H(B) = -\sum_{i=1}^{s} P(X_i) log P(X_i),$$

$$H(B|P) = -\sum_{i=1}^{s} P(X_i) \sum_{j=1}^{q} P(Y_j|X_i)\, log P(Y_j|X_i)\,,$$

$$I(B;P) = H(P) - H(P|B).$$

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

Information Representation

- Let $DIS = (U, C \cup D, V, f)$ be a decision table. For $B \subseteq C$, the **lower and upper approximation** can be defined from information viewpoint.

$$B_*(X) = \cup_{x \in U}\{B(x) | H(X|B(x)) = 1\}$$

$$B^*(X) = \cup_{x \in U}\{B(x) | H(X|B(x)) > 0\}$$

Then, the boundary region of X is

$$BN_B(X) = B^*(X) - B_*(X)$$

X is rough when the boundary region is nonempty.

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

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

Information Representation

☐ The Attribute Reduction Defined by Information representation

Let $DIS = (U, C \cup D, V, f)$ be a decision table. For $B \subseteq C$, the B is a reduct subset induced by information entropy, if

(1) $H(B) = H(C)$;

(2) For $b \in B$, $H(B - \{b\}) \neq H(C)$.

Subsequently, I developed a **heuristic algorithm** for attribute reduction based on Information representation.

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

Information Representation

Algorithm 2 A heuristic algorithm based on mutual information.

Input: A decision table $DIS = (U, C \cup D, V, f)$.

Output: A reduct subset of DIS.

Step 1. Compute the mutual information $I(C; D)$ of C and D.

Step 2. Compute the core $C_0 = CORE_D(C)$ of C related to D;

Step 3. Let $B = C_0$, for $\forall p \in C - B$ repreats the following:

① Compute the conditional mutual information $I(p; D|B)$;

② Select the attribute satisfying $p_0 = argmax_{p \in C-B} I(p; D|B)$;

③ $B \leftarrow B \cup \{p_0\}$;

④ Stop this step when $I(B; D) = I(C; D)$, turn to ①

Step 4. Return B.

[1] D. Miao, G. Hu, "A heuristic algorithm for reduction of knowledge," in *Journal of Computer research & development*, vol. 36, no. 6, pp. 681-684, 1999.

Information Representation-New

- **Zentropy** was proposed by Liu et al. in 2022.
- It emphasizes that system entropy is **the whole reflection of different scales**, which have been used for physical property prediction.

- We first applied this thought to RS and constructed a zentropy-based measure for uncertain knowledge processing.

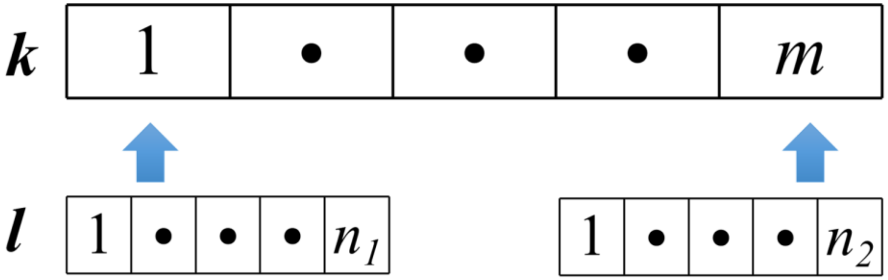
[1] Z. Liu, Y. Wang, S. Shang, "Zentropy theory for positive and negative thermal expansion," in *Journal of Phase Equilibria and Diffusion*, vol. 43, pp. 598-605, 2022.
[2] K. Yuan, D. Miao, Y. Yao, et al., "Feature selection using zentropy-based uncertainty measure," in *IEEE Transactions on Fuzzy Systems*, vol. 32, no. 4, pp. 2246-2260, 2024.

Information Representation-New

- The definition of **Zentropy**

$$Z_B(\mathrm{D}) = -\boxed{\sum_{k=1}^{m} p_k log p_k} + \boxed{\sum_{k=1}^{m} p_k Z_k},$$

current scale entropy lower scale entropy



- There are m parts at the current scale;
- Each part can be decomposed into the finer parts at a lower scale.

[1] D. Miao, "Rough set theory and and its applications in machine learning," *in Doctoral dissertation, Institute of Automation, Chinese academy of sciences*, 1997.
[2] K. Yuan, D. Miao, Y. Yao, et al., "Feature selection using zentropy-based uncertainty measure," in *IEEE Transactions on Fuzzy Systems*, vol. 32, no. 4, pp. 2246-2260, 2024.

Comparison of Attribute Reduction Algorithms

Examples of attribute reduction

Table 1 The reduct resutls of different methods

Representation	Methods	Reducts
Algebraic representation	Positive region	{Weather, Road} and {Weather, time}
	Discernibility matrix	{Weather, Road} and {Weather, time}
Information representation	Mutual information	{Weather, Road} and {Weather, time}
	Zentropy measure	{Weather, Road} and {Weather, time}

Comparison of Attribute Reduction Algorithms

Examples of attribute reduction

Making some changes in the Input :

Fact	Condition attributes			Decision attribute	Support
	Weather	Road	Time	Accident	
1	misty	icy	day	yes	6
2	foggy	icy	dusk	yes	8
3	misty	not icy	night	yes	5
4	sunny	icy	day	no	55
5	foggy	not icy	dusk	yes	11
6	misty	not icy	night	no	15

Comparison of Attribute Reduction Algorithms

Examples of attribute reduction

Table 2 The reduct resutls when Input changes

Knowledge	Methods	Reduct subset
Algebraic representation	Positive region	{Weather, Road} and {Weather, time}
	Discernibility matrix	{Weather, Road} and {Weather, time}
Information representation	Mutual information	{Weather, Road} and {Weather, time}
	Zentropy measure	**{Weather, time}**

Comparison of Attribute Reduction Algorithms

Compared with other methods, the zentropy measure describes **the information presented at different granular levels**, So, It can accurately character information changes with attribute variation and obtain the reducts with Stronger discriminability.

Contributions of RS to AI

- Formal definition of knowledge
- Boundary region calculation Method of imprecise concept
- A machine learning model for qualitative data (imprecise, inconsistent, and incomplete)

Outline

- Artificial Intelligence
- Rough Set Theory
- Contributions of RS to AI
- Conclusions

Conclusions

- Rough sets provides **multi-granularity representation, learning, and reasoning** methods for AI.
- **Three-way decision(Proposed by Dr. Y.Yao)** is an important extension of RS and has been successfully applied to various scenarios.
- Rough sets as a **data-driven Machine Learning Model** showes strong application value and development potential.

[1] Y. Yao, "Three-way decisions with probabilistic rough sets," in *Information sciences*, vol. 180, no. 3, pp.341-353, 2010.
[2] D. Miao, K. Yuan, "The contributions of rough sets to artificial intelligence," submitted to the special issue of *40 Years Rough Sets on Applied Soft Computing Journal*.

Paper list in My Lab. ([http://iip.tongji.edu .cn](http://iip.tongji.edu.cn))

[1] Y. Zhang, Q. Zhang, Z. Gong, Y. Shi, Y. Liu, **D. Miao***, "MLIP: Efficient multi-perspective language-image pretraining with exhaustive data utilization", ***ICML*** **2024**.

[2] Z. Wu, Q. Zhang, **D. Miao*,** "HyDiscGAN: A hybrid distributed cGAN for audio-visual privacy preservation in multimodal sentiment analysis," ***IJCAI*** **2024**.

[3] A. Lao, Q. Zhang, C. Shi, L. Cao, K. Yi, L. Hu, **D. Miao**, "Frequency spectrum is more effective for multimodal representation and fusion: A multimodal spectrum rumor detector," ***AAAI*** **2024**.

[4] Y. Zhang, Y. Liu, **D. Miao***, "MG-ViT: A multi-granularity method for compact and efficient vision transformers," ***NeurIPS*** **2023**.

[5] Y. Li, Y. Liu, H. Zhang, C. Zhao, Z. Wei, **D. Miao*,** "Occlusion-aware transformer with second-order attention for person re-identification," ***IEEE Transactions on Image Processing***, doi: 10.1109/TIP.2024.3393360, 2024.

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

[7] D. Chen, **D. Miao***, X. Zhao, "Hyneter: Hybrid network transformer for multiple computer vision tasks," ***IEEE Transactions on Industrial Informatics***, doi: 10.1109/TII.2024.3367043, 2024.

[8] L. Ying, **D. Miao***, Z. Zhang, "3WM-AugNet: A feature augmentation network for remote sensing ship detection based on three-way decisions and multigranularity," ***IEEE Transactions on Geoscience and Remote Sensing***, vol. 63, pp. 1-19, 2023.

Thank you !

The Contributions of Rough Sets to Artificial Intelligence

Duoqian Miao

Tongji University

Shanghai, P.R. China

May 18, 2024