解説

ラフ集合のこれまでとこれから[†]

工藤 康生^{*}

1. はじめに

ソフトコンピューティングの大きな特徴の1つとして、情報にまつわる様々な「あいまいさ」の取り扱いが挙げられる。情報のあいまいさを扱う方法として、まず思い浮かぶのはファジィ理論であるが、他の方法の一種としてラフ集合 (rough set) [1] がある。ラフ集合は、ポーランド出身の応用数学者・計算機科学者であった Z. Pawlak 教授 (1926–2006) が 1982 年に提唱した理論である。ファジィ集合が表すあいまいさが、ぼやけた境界を持つ概念のあいまいさ [2] であることに對し、ラフ集合が表すあいまいさは、背景知識の下では正確に定義できない概念のあいまいさである [3]。本稿では、ラフ集合理論の誕生から現在に至る、ラフ集合関連の研究の大まかな流れを概説する。なお、本稿で用いる定義等は文献 [3, 4] を参考にした。

2. ラフ集合のこれまで

2.1 ラフ集合の誕生

考える対象全体の有限集合 U と、背景知識を表す U 上の同値関係 R が与えられているとする。対象に関するある概念（部分集合） $X \subseteq U$ について、背景知識 R による X の下近似 (lower approximation) $\underline{R}(X)$ と上近似 (upper approximation) $\overline{R}(X)$ をそれぞれ以下のように定義する：

$$\underline{R}(X) = \{x \in U \mid [x]_R \subseteq X\}, \quad (1)$$

$$\overline{R}(X) = \{x \in U \mid [x]_R \cap X \neq \emptyset\}. \quad (2)$$

ここで、 $[x]_R$ は R による対象 $x \in U$ の同値類であり、背景知識 R の下で x と区別できない対象の集合である。定義より明らかに、包含関係 $\underline{R}(X) \subseteq X$

および $X \subseteq \overline{R}(X)$ が成り立つ。下近似と上近似の対 $(\underline{R}(X), \overline{R}(X))$ を X のラフ集合と呼び、 $\underline{R}(X) = X = \overline{R}(X)$ であるとき、概念 X は R -定義可能 (R -definable) であるという。そうでないとき、すなわち $\underline{R}(X) \subset X \subset \overline{R}(X)$ であるとき、 X は R -ラフ (R -rough) であるという。 R -ラフな概念 X は、ラフ集合が表すあいまいさを持つ概念である。

2.2 近似から縮約計算・ルール抽出へ

上述のように、本来のラフ集合は概念の集合論的近似に関する理論であるが、転じて、表形式のデータに対する分析手法の一種として認識されるようになった。その理由は、概念の集合論的近似、特に下近似の抽出と、If-then 形式のルールの抽出とが密接に関連しているためである。本節では、データ分析手法としてのラフ集合について概説する。

2.2.1 決定表と識別不能関係

分析対象とする表形式データを決定表 (decision table) と呼ぶ。決定表に現れる項目 (属性) は、特徴を知りたい項目に対応する決定属性 (decision attribute) とその特徴を記述する条件属性 (condition attribute) に分けられる。決定表を $(U, C \cup \{d\}, V, \rho)$ の 4 項組で定義する。ここで、 U は対象の有限集合、 C は条件属性の集合、 $d \notin C$ は決定属性、 V は各属性の値の集合、 $\rho : U \times (C \cup \{d\}) \rightarrow V$ は各対象 $x \in U$ に各属性 $a \in C \cup \{d\}$ での値 $\rho(x, a) \in V$ を定める関数である。なお、決定属性は複数個使用することも可能だが、本稿では決定属性を 1 個に限定する。

決定表における対象の有限集合 U でも、2.1 節と同様に、背景知識の下で概念の近似を行うことができる。背景知識を表す同値関係として、以下で定義される識別不能関係 (indiscernibility relation) を用いる。属性の任意の部分集合 $A \subseteq C \cup \{d\}$ に対して、対象の集合 U 上の識別不能関係 R_A を

$$R_A = \{(x, y) \in U \times U \mid \rho(x, a) = \rho(y, a), \forall a \in A\} \quad (3)$$

[†] A Review of Research Trends and Future Issues of Rough Set Theory
Yasuo KUDO

^{*} 室蘭工業大学大学院工学研究科
Graduate School of Engineering, Muroran Institute of Technology

と定義する。識別不能関係が同値関係となることは容易に確認できる。決定属性による識別不能関係から得られた同値類 $D_1, \dots, D_n$ を決定クラス (decision class) と呼び、近似の対象となる概念として用いる。

2.2.2 決定ルール

条件属性の部分集合 $A \subseteq C$ を用いて識別不能関係 R_A を考える。任意の決定クラス D_j ($1 \leq j \leq n$) に対して、2.1 節で述べた下近似との包含関係 $\underline{R}_A(D_j) \subseteq D_j$ が成り立つため、下近似 $\underline{R}_A(D_j)$ が空集合でなければ、その要素はすべて決定クラス D_j に含まれる。対象 $x \in \underline{R}_A(D_j)$ および $A = \{a_1, \dots, a_k\} \subseteq C$ 、決定属性 d について、包含関係 $\underline{R}_A(D_j) \subseteq D_j$ は、「 a_1 の値が $\rho(x, a_1)$ かつ $\dots a_k$ の値が $\rho(x, a_k)$ ならば、 d の値が $\rho(x, d)$ である」という規則性を表す。この規則性を以下の表現で表し、決定ルール (decision rule) と呼ぶ:

$$\bigwedge_{a_i \in A} [a_i = \rho(x, a_i)] \rightarrow [d = \rho(x, d)]. \quad (4)$$

含意記号 $\rightarrow$ の左側を決定ルールの前提 (antecedent)、右側を結論 (consequent) と呼ぶ。決定クラスの下近似に含まれる対象から作成した決定ルールは、前提に対して結論が一意に定まる無矛盾な決定ルールである。なお、決定クラスの下近似に含まれない対象からも、決定ルールを作成することは可能である。このような決定ルールは、前提に対して結論が一意に定まらないため、矛盾した決定ルールとなる。

2.2.3 相対縮約

決定表内の規則性の記述として、決定ルールの前提部分は短い方が、ルールとしての可読性が高まる。一方、決定ルールによる対象の分類能力は、すべての条件属性を用いた場合と同等に維持されなければならない。このような前提部分を与える条件属性の部分集合を求める計算を、縮約計算あるいは属性縮約 (attribute reduction) と呼び、その結果得られる条件属性の部分集合を相対縮約 (relative reduct) と呼ぶ。

決定表からすべての相対縮約を抽出する手法の 1 つとして、識別行列 (discernibility matrix) を用いた手法 [5] が知られている。一方、与えられた決定表からすべての相対縮約を抽出する問題は NP-困難であることも証明されている [5]。よって、規模の大きいデータからすべての相対縮約を抽出することは現実的に不可能であるため、QuickReduct [6] を始め、少数個の相対縮約の候補を発見的に抽出するアルゴリズムが多数提案されている (例えば [7–11])。一方、属性の個数が多いデータから多数の相対縮約を抽出する手法も提案されている [12]。

2.3 様々な一般化・他手法との融合

ラフ集合で用いられる諸概念を緩和・一般化することや、他の手法と融合することは自然な発想である。本節ではその一部を概説する。

包含関係の緩和 代表例として、可変精度ラフ集合 (variable precision rough set) [13] が挙げられる。予め設定した誤差許容度 β ($0 \leq \beta < 0.5$) の下で、集合 X が集合 Y からはみ出す度合いが β 以下であるとき、 X は Y に包含されると見なす。これにより、多少の誤りを許容した上下近似などが構成できる。

同値関係の一般化 同値関係を一般的な 2 項関係に置き換える研究は非常に多数行われている。一例として文献 [14] が挙げられる。

分割の一般化 同値類の集合は全体集合の分割となる。これを緩和する研究も非常に多数行われており、例として近傍ラフ集合 (neighborhood rough set) [15] や被覆に基づくラフ集合 (covering-based rough set) [16] などが挙げられる。また、この方向での一般化は粒状計算 (granular computing) とも親和性が高い。

名義尺度の緩和 Pawlak のラフ集合では、各属性は名義尺度に限定されている。これを緩和した代表例として、各属性で序数性を考慮した支配関係に基づくラフ集合 (dominance-based rough set) [17] が挙げられる。

値の欠損の許容 Pawlak のラフ集合における決定表では、値の欠損が許容されていない。これを一般化し、値の欠損を許容するラフ集合が多数提案されている。一例として文献 [18] が挙げられる。

ラフ集合とファジィ ラフ集合に対するファジィ化も多数提案されている。代表例として、ラフファジィ集合 (rough fuzzy set) やファジィラフ集合 (fuzzy rough set) が提案されている [19]。

ラフ集合と論理 ラフ集合と様相論理の可能世界意味論との対応関係は早くから指摘されており、本誌でも村井による解説 [20] が掲載されている。先に挙げた文献 [14] はこの一例でもある。

また、ラフ集合と DS 証拠理論 (Dempster-Shafer theory of evidence) との融合やラフ集合と形式概念分析 (formal concept analysis) との融合に関する研究も多数行われている。

3. おわりに：ラフ集合のこれから

本稿では、ラフ集合の誕生からこれまでの研究の大きな流れを概説した。特に、ラフ集合による対象の集合論的近似が、表形式データの分析と密接に関連し

ていることを述べた。ラフ集合の根幹は対象の集合論的近似であることを知って頂けると幸いである。

ラフ集合という特定の分野の中でも、研究の流行がある。近年のラフ集合に関する国際会議（International Joint Conference on Rough Sets : IJCRS）での発表論文を見ると、Yao et al. [21] が提唱した決定理論的ラフ集合（decision-theoretic rough set）およびこれに基づく three-way decisions [22] に関する論文が非常に多い。一方、ラフ集合を実問題解決に応用する研究は、近年あまり目立たない印象がある。大規模なデータに対するラフ集合の適用は、今後の重要な課題の 1 つである。

参考文献

- [1] Z. Pawlak: “Rough Sets,” *Int. J. of Computer and Information Science*, Vol.11, pp. 341-356, 1982.
- [2] 日本ファジィ学会: “ファジィ測度,” 講座 ファジィ 3, 日刊工業新聞社, 1993.
- [3] Z. Pawlak: “Rough Sets,” *Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publisher, 1991.
- [4] 乾口 雅弘: “ラフ集合と決定表の解析,” ラフ集合と感性 データからの知識獲得と推論, 海文堂出版, pp. 163-181, 2004.
- [5] A. Skowron and C. M. Rauszer: “The discernibility matrix and functions in information systems,” *Intelligent Decision Support: Handbook of Application and Advance of the Rough Set Theory*, R. Słowiński ed., Kluwer Academic Publishers, pp. 331-362, 1992.
- [6] A. Chouchoulas and Q. Shen: “Rough Set-Aided Keyword Reduction for Text Categorization,” *Applied Artificial Intelligence*, Vol.15, No.9, pp. 843-873, 2001.
- [7] K. Hu, L. Diao, Y. Lu, and C. Shi: “A Heuristic Optimal Reduct Algorithm,” *Proc. of IDEAL2000*, Springer, pp. 139-144, 2000.
- [8] F. Hu, G. Wang, and L. Feng: “Fast Knowledge Reduction Algorithms Based on Quick Sort,” *Rough Sets and Knowledge Technology*, Springer, pp. 72-79, 2008.
- [9] Y. Kudo and T. Murai: “Heuristic Algorithm for Attribute Reduction Based on Classification Ability by Condition Attributes,” *J. Adv. Comput. Intell. Intell. Inform.*, Vol.15, No.1, pp. 102-109, 2011.
- [10] S. Tan, H. Xu, and J. Gu: “Efficient Algorithms for Attributes Reduction Problem,” *Int. J. of Innovative Computing, Information and Control*, Vol.1, No.4, pp. 767-777, 2005.
- [11] Z. Xu, C. Zhang, S. Zhang, W. Song, and B. Yang: “Efficient Attribute Reduction Based on Discernibility Matrix,” *Rough Sets and Knowledge Technology*, Springer, pp. 13-21, 2007.
- [12] Y. Kudo and T. Murai: “A Parallel Computation Method for Heuristic Attribute Reduction Using Reduced Decision Tables,” *J. Adv. Comput. Intell. Intell. Inform.*, Vol.17, No.3, pp. 371-376, 2013.
- [13] W. Ziarko: “Variable Precision Rough Set Model,” *J. of Computer and System Science*, Vol.46, pp.39-59, 1993.
- [14] Y. Y. Yao and T. Y. Lin: “Generalization of Rough Sets using Modal Logics,” *Intelligent Automation & Soft Computing*, Vol.2, Issue 2, pp. 103-119, 1996.
- [15] T. Y. Lin: “Granulation and Nearest Neighborhoods: Rough Set Approach,” *Granular Computing: An Engineering Paradigm*, Springer, pp. 125-142, 2001.
- [16] W. Zhu: “Topological approaches to covering rough sets,” *Information Sciences*, Vol.177, No.6, pp. 1499-1508, 2007.
- [17] S. Greco, B. Matarazzo, and R. Słowiński: “Rough sets theory for multi-criteria decision analysis,” *European J. of Operational Research*, Vol.129, No.1, pp. 1-47, 2001.
- [18] M. Kryszkiewica: “Rough set approach to incomplete information systems,” *Information Sciences*, Vol.112, pp. 39-49, 1998.
- [19] D. Dubois and H. Prade: “Rough Fuzzy Sets and Fuzzy Rough Sets,” *Int. J. of General Systems*, Vol.17, Issue 2-3, pp. 191-209, 1990.
- [20] 村井哲也: “ラフ集合と様相論理,” 知能と情報, Vol.13, No.6, pp. 571-580, 2001.
- [21] Y. Y. Yao, S. K. M. Wong, and P. Lingras: “A decision-theoretic rough set model,” *Methodologies for Intelligent Systems*, 5, *Proc. of the 5th Int. Symposium on Methodologies for Intelligent Systems*. Knoxville, Tennessee, pp. 17-25, 1990.
- [22] Y. Y. Yao: “Rough Sets and Three-Way Decisions,” *Proc. of RSKT 2015*, Springer, pp. 62-73, 2015.

(2018 年 6 月 18 日 受付)

【問い合わせ先】

〒050-8585 北海道室蘭市水元町 27-1

室蘭工業大学

工藤 康生

E-mail: kudo@csse.muroran-it.ac.jp

著者紹介

くどう やすお
工藤 康生 [正会員]

1995 年北海道教育大学教育学部函館校総合科学課程卒業。1997 年北海道大学大学院工学研究科システム情報工学専攻博士前期課程修了。2000 年同博士後期課程修了、博士(工学)。同年室蘭工業大学サテライト・ベンチャー・ビジネス・ラボラトリー博士研究員。2003 年同大学情報工学科助手。以来、助教、准教授を経て、2016 年同大学大学院工学研究科しくみ情報系領域教授、現在に至る。ラフ集合の理論と応用や情報推薦システムなどの研究に従事。IEEE、日本知能情報ファジィ学会、日本感性工学会などの会員。