

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 4 月 3 日現在

機関番号：17104
 研究種目：基盤研究（C）
 研究期間：2010～2012
 課題番号：22500204
 研究課題名（和文）ラフ集合不完全情報解析の新展開とそのデータマイニングへの応用
 研究課題名（英文）A New Stage of Rough Sets Incomplete Information Analysis and Its Application to Data Mining
 研究代表者
 酒井 浩（SAKAI HIROSHI）
 九州工業大学・工学研究院・教授
 研究者番号：60201513

研究成果の概要（和文）：本研究では、確定情報表の下で提案・応用されているラフ集合理論を、不完全情報表も処理する体系に広げ、情報表におけるデータマイニングの新展開を図った。提案している NIS-アプリアルゴリズムは既存のアプリアルゴリズムを不完全情報表に対応させたものであり、計算量は既存分と殆ど変わらない。本アルゴリズムに関する完全性定理（定義されたルールの集合を過不足なく抽出する）も証明した。現実面では、本アルゴリズムによるソフトウェアを実現し、実行例をホームページに公開している。また、現在、本ソフトウェアのウェブ版公開に向けた取り組みを進めている。

研究成果の概要（英文）：We coped with rough set theory and rough sets-based data mining. We have theoretically extended rough sets-based concepts in a standard table to concepts in a table with non-deterministic information. Especially in rule generation from a table, we have proposed NIS-Apriori algorithm, which can handle tables with non-deterministic information and the computational complexity is almost the same as the original Apriori algorithm. Furthermore, NIS-Apriori is sound and complete for rules defined in a table with non-deterministic information. We have practically implemented software in C and Prolog, and opened the execution logs in the web page. Currently, we are touching a prototype of a web version with NIS-Apriori.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	1,300,000	390,000	1,690,000
2011年度	1,100,000	330,000	1,430,000
2012年度	900,000	270,000	1,170,000
年度	0	0	0
年度	0	0	0
総計	3,300,000	990,000	4,290,000

研究分野：総合領域

科研費の分科・細目：情報学、感性情報学・ソフトコンピューティング

キーワード：ラフ集合、不完全情報、データマイニング、アプリアルゴリズム

1. 研究開始当初の背景

(1) 本研究で中心的役割を果たす NIS-アプリアルゴリズムを提案してはいたが、本アルゴリズムにおける計算式の記述が非常

に複雑であり、説明や証明のための新たな記述方法が必要であった。

(2) 従来のラフ集合理論では、属性値の集

合は離散値有限集合に限定されており、連続値の不完全性を扱うための区間処理の導入が必要であった。

(3) NIS-アプリアリによるデータマイニングツールの実現、実行例の公表、さらにはウェブ版ツールの公開などが望まれた。研究当初は Prolog 言語による記述を開始しており、Prolog を Java に変換する PrologCafe (神戸大学、田村研究室がウェブ上に公開中) の利用や MySQL の利用を検討した。

2. 研究の目的

(1) 現在まで離散値の不完全性 (非決定情報を意味する) を扱う体系「ラフ集合非決定情報解析」の研究を進めてきた。今回、研究対象を連続値の不完全性 (区間情報を意味する) まで広げ、新たに「ラフ集合不完全情報解析」として提案し基礎理論の確立を図る。

(2) 理論面のみならず、提案する枠組みを計算機上に実現し、不完全な情報を含む場合にも特徴あるルール抽出が可能になるデータマイニングツールの実現を図る。

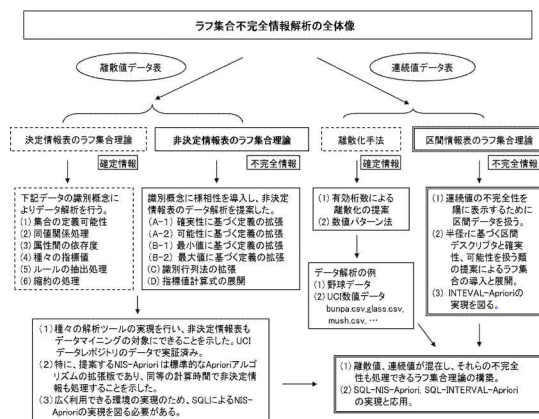


図 1. 研究の全体像

3. 研究の方法

目的で提案した内容を実現するために、従来の手法を比較しながら新たな枠組みを提案した。処理系の実現では、PrologCafe、MySQL などの公開された処理系の利用も検討した。検証のために機械学習レポジトリ (Machine Learning Repository, UC Irvine 校) に登録されたデータを用いた。

4. 研究成果

(1) 平成 22 年度においては、研究対象を連続値情報の不完全性 (区間情報を意味する) まで広げた。離散値の場合には起こり得る可能な場合 (派生する決定情報表) を自然に定義できるが、区間情報の場合には集合の濃度との関係から起こり得る場合が無制限

になることもあり、離散値と比べてその扱いが難しい。そこで、起こり得る場合を規定するために解像度 (Resolution) γ (正の数) を与え、 γ に基づく体系として再構築した。既に提案している NIS-アプリアリアルゴリズムに (解像度 γ による) 離散化処理を追加し、Prolog 言語により実現した。

(2) 平成 23 年度においては、Prolog 版実験システムの高速度化と処理環境の整備が進んだ。論理型言語の特徴をうまく利用することで実験システムを実現しているが、実行速度の面において論理型言語は劣ると考えられる。この問題に対して、論理型言語を手続き型言語に変換し、コンパイルする手法を応用した。結果、いずれのプログラムにおいても実行時間が半分以下に短縮された。実行例はホームページにアップロードしている。

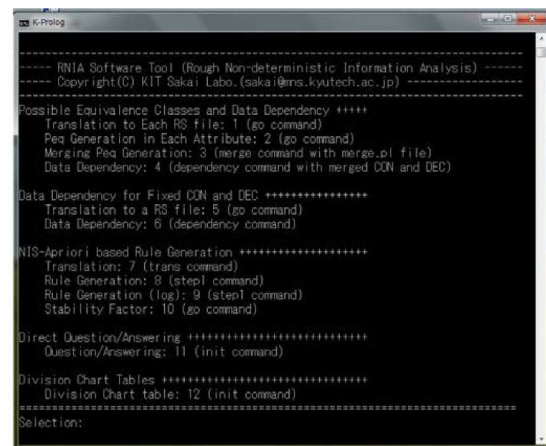


図 2. Prolog 版ソフトウェアメニュー画面

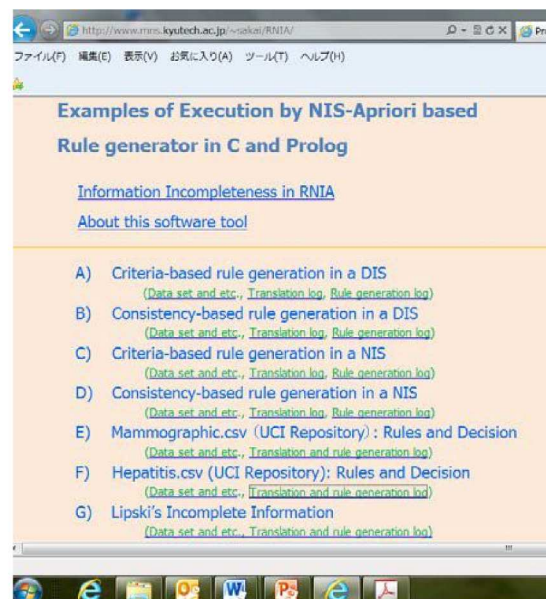


図 3. 実行例の公開ページ

http://www.mns.kyutech.ac.jp/~sakai/RNIA/

(3) 平成23年度、理論的側面においては、新たに非決定情報表における「分割図」と「分割図表」を提案した。これは、統計分野でよく利用される「分割表 (Contingency Table)」をルール生成のために焼き直し、改めて提案した図表である。ラフ集合では、無矛盾性を同値類の包含関係により判定でき、この性質を利用して無矛盾性の計算を行う点に特徴があるが、この計算と同等な結果を分割図により表現できることを示した。そして、この分割図と分割図表を利用し、従来、説明が難解であった定理に新たな証明を与えることができた。

(4) 平成24年度においては、分割図の合併アルゴリズムを得ることができ、これを用いて NIS-アプリアリを書き換えた。さらに、Google の Apps を利用し、Python 言語を用いたウェブ版 NIS-アプリアリをほぼ仕上げている。現在、β版として利用可能な状況になっている。このように、今年度はソフトウェアの実現の面において重要な進展があった。

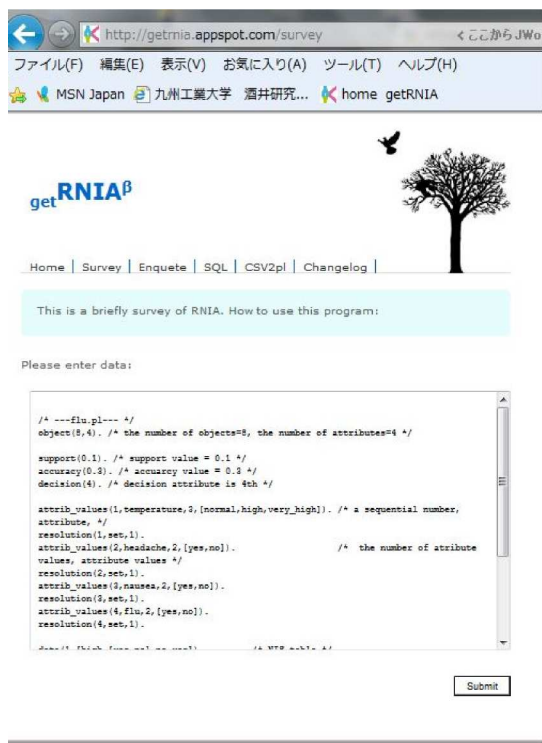


図4. ウェブ上で動く NIS-アプリアリ

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計20件)

- ① Hiroshi Sakai, Mao Wu, Michinori Nakata, Dominik Slezak, Rough sets-based ma-

chine learning over non-deterministic data: A brief survey, Communications in Computer and Information Science, Springer, 査読有, Vol.322, 2012, pp.3-12,

(DOI:10.1007/978-3-642-35326-0_1)

- ② Mao Wu, Naoto Yamaguchi, Michinori Nakata, Hiroshi Sakai, Learning a table from a table with non-deterministic information: A perspective, Communications in Computer and Information Science, Springer, 査読有, Vol.322, 2012, pp.25-34,

(DOI:10.1007/978-3-642-35326-0_3)

- ③ Hiroshi Sakai, Mao Wu, Michinori Nakata, Association rule-based decision making in table data, International Journal of Reasoning-based Intelligent Systems, Inderscience, 査読有, Vol. 4, No. 3, 2012, pp.162-170,

(DOI:10.1504/IJRIS.2012.050374)

- ④ Hiroshi Sakai, Mao Wu, Naoto Yamaguchi, Michinori Nakata, Division charts and their merging algorithm in rough non-deterministic information analysis, Proc. IEEE. Granular Computing, 査読有, 2012, pp.1-6,

(DOI:10.1109/GrC.2012.6468685)

- ⑤ Hiroshi Sakai, Michinori Nakata, Dominik Slezak, Management of information incompleteness in rough non-deterministic information analysis, Communications in Computer and Information Science, Springer, 査読有, Vol.297, 2012, pp.280-289,

(DOI:10.1007/978-3-642-31709-5_29)

- ⑥ Hiroshi Sakai, Hitomi Okuma, Mao Wu, Michinori Nakata, A descriptor-based division chart table in rough non-deterministic information analysis, Smart Innovation, Systems and Technologies, Springer, 査読有, Vol.15, 2012, pp.25-34,

(DOI:10.1007/978-3-642-29977-3_3)

- ⑦ Hiroshi Sakai, Michinori Nakata, Toward association rules based decision making in Lipski's Incomplete Information Databases, Proc. IEEE. Granular Computing, 査読有, 2011, pp.570-575,

(DOI:10.1109/GRC.2011.6122659)

- ⑧ Hiroshi Sakai, Kohei Hayashi, Michinori Nakata, Dominik Slezak, A mathematical extension of rough set based issues toward uncertain information analysis, New Mathematics and Natural Computation, World Scientific, 査読有, Vol. 7, No. 3, 2011, pp.543

- 570、
(DOI:10.1142/S1793005711002062)
- ⑨ Hiroshi Sakai、Michinori Nakata、Dominik Slezak、A NIS-apriori based rule generator in Prolog and its functionality for table data、Proc. RSKT、Springer LNCS、査読有、Vol. 6954、2011、pp. 226-231、
(DOI:10.1007/978-3-642-24425-4_31)
- ⑩ Hiroshi Sakai、Michinori Nakata、Dominik Slezak、A prototype system for rule generation in Lipski's incomplete information databases、Proc. RSFDGrC、Springer LNAI、査読有、Vol. 6743、2011、pp. 175-182、
(DOI:10.1007/978-3-642-21881-1_29)
- ⑪ Aboul Ella Hassanién、Hiroshi Sakai、Dominik Slezak、Michir K. Chakraborty、William Zhu、Special issue: Rough and fuzzy methods for data mining、International Journal of Hybrid Intelligent Systems、IOS Press、査読有、Vol. 8、No. 1、2011、pp. 1-2、
(DOI:10.3233/HIS-2011-0125)
- ⑫ Hiroshi Sakai、Hitomi Okuma、Michinori Nakata、Dominik Slezak、Stable rule extraction and decision making in rough non-deterministic information analysis、International Journal of Hybrid Intelligent Systems、IOS Press、査読有、Vol. 8、No. 1、2011、pp. 41-57、
(DOI:10.3233/HIS-2011-0130)
- ⑬ J. Peters、A. Skowron、H. Sakai、M. Chakraborty、D. Slezak、A. E. Hassanién、W. Zhu、Preface of Special Issue、Transaction on Rough Sets、Springer-Verlag、査読有、Vol. 14、No. 1-2、2011、pp. 1-2、
(DOI:10.1007/978-3-642-21563-6)
- ⑭ Hiroshi Sakai、Michinori Nakata、Dominik Slezak、Toward rough sets based rule generation from tables with uncertain numerical values、Advances in Intelligent and Soft Computing、Springer、査読有、Vol. 68、2010、pp. 395-406、
(DOI:10.1007/978-3-642-11960-6_37)
- ⑮ Hiroshi Sakai、Michinori Nakata、Dominik Slezak、Rule generation in Lipski's incomplete information databases、Proc. RSCTC、Springer LNCS、査読有、Vol. 6086、2010、pp. 376-385、
(DOI:10.1007/978-3-642-13529-3_40)
- ⑯ Hiroshi Sakai、Michinori Nakata、Dominik Slezak、The lower and the upper systems of rules in tables with missing values、Springer、CCIS、査読有、Vol. 118、2010、pp. 132-141、
(DOI:10.1007/978-3-642-17622-7_14)
- ⑰ Michinori Nakata、Hiroshi Sakai、Two rough approximations for information tables containing missing values、査読有、Proc. IEEE Granular Computing、2010、pp. 363-368、
(DOI:10.1109/GrC.2010.139)
- [学会発表] (計 8 件)
- ① Hiroshi Sakai、Mao Wu、Naoto Yamaguchi、Michinori Nakata、Rough non-deterministic information analysis and its software tool、招待講演、FIM 学会、2012 年 12 月 16 日、パンジャブ大学 (インド)
- ② 中田典規、酒井浩、可能性分布で表された情報を扱うラフ集合、ファジイシステムシンポジウム、2012 年 9 月 14 日、名古屋工業大学 (名古屋)
- ③ 卯伍、山口直人、中田典規、酒井浩、分割図と NIS アプリオリアルゴリズムに関する考察、ファジイシステムシンポジウム、2012 年 9 月 13 日、名古屋工業大学 (名古屋)
- ④ 酒井浩、大隈ひとみ、中田典規、非決定情報システムにおける粒状集合 \inf と \sup による依存度計算の考察、ファジイシステムシンポジウム、2011 年 9 月 13 日、福井大学 (福井)
- ⑤ 中田典規、酒井浩、不完全な情報テーブルにおける二つのラフ近似、ファジイシステムシンポジウム、2011 年 9 月 13 日、福井大学 (福井)
- ⑥ Hiroshi Sakai、Dominik Slezak、Rough set theory and rough non-deterministic information analysis: A brief survey、実解析学シンポジウム、Vol. 4、2010 年 11 月 14 日、九州工業大学 (北九州)
- ⑦ 中田典規、酒井浩、欠損値を含む情報テーブルにおける二つのラフ近似、ファジイシステムシンポジウム、2010 年 9 月 14 日、広島大学 (広島)
- ⑧ 大隈ひとみ、中田典規、酒井浩、非決定情報システムにおける可能ルールの信頼度数について、ファジイシステムシンポジウム、2010 年 9 月 14 日、広島大学 (広島)
- [図書] (計 2 件)
- ① Hiroshi Sakai、Hitomi Okuma、Mao Wu、Michinori Nakata、Rough Non-deterministic Information Analysis for Uncertain Information、The Handbook on Reasoning-Based Intelligent Systems、Chapter 4、World Scientific、査読有、2013、pp. 81-118、
(DOI:10.1142/9789814329484_0004)

- ② Hiroshi Sakai, Hitomi Okuma, Michinori Nakata, Rough Non-deterministic Information Analysis: Foundations and Its Perspective in Machine Learning, Smart Innovation, Systems and Technologies, Springer, Chapter 9、査読有、2013、pp. 215-247、
(DOI:10.1007/978-3-642-28699-5_9)

[その他]

ホームページ等

<http://www.mns.kyutech.ac.jp/~sakai/RNIA/>

<http://getrnia.appspot.com/>

6. 研究組織

(1) 研究代表者

酒井 浩 (SAKAI HIROSHI)

九州工業大学・工学研究院・教授

研究者番号：60201513