様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　３０　年　　５　月　　８　日現在

機関番号：　１７１０４
研究種目：　基盤研究(C)（一般）
研究期間：　2014～2017
課題番号：　２６３３０２７７

研究課題名（和文）ラフ集合非決定情報解析に関連する諸問題の研究

研究課題名（英文）A study of the problems associated with Rough Set Non-deterministic Information Analysis

研究代表者

酒井　浩（SAKAI, Hiroshi）

九州工業大学・大学院工学研究院・教授

研究者番号：６０２０１５１３

交付決定額（研究期間全体）：（直接経費）　　3,600,000円

研究成果の概要（和文）：ラフ集合理論は表データマイニングのための数学的枠組である。特徴的な含意式(ルール)生成に活用され、表データの性質把握や表データに基づく意思決定支援に応用される。研究代表者はラフ集合理論に様相論理(可能世界意味論)を導入し、情報の非決定性まで考慮できるラフ集合非決定情報解析(RNIA)を提案した。場合分けによる指数オーダー問題を解消できたため、RNIAは他に類のない枠組みになっている。コアアルゴリズムNIS-Aprioriにより、関連する問題（少数派意見マイニング、解析ソフトウェアの改善、ビッグデータへの対応、実データへの応用、プライバシー・データ保護、欠損値推定)の解決を図った。

研究成果の概要（英文）：Rough set theory is a mathematical framework for mining table data sets. This theory is utilized for generating the characteristic implications (rules), and is applied to the recognition of the properties and decision support in table data sets. Principal investigator introduced the modal logic (possible worlds semantics) in rough set theory, and proposed Rough set Non-deterministic Information Analysis (RNIA) that can be taken into account to non-deterministic information. Because, the computational complexity problem was solved, RNIA became a quite unique framework. The proposed NIS-Apriori algorithm is employed as the core algorithm, and related problems like minor rule mining, the improvement of the analytic software tool, the improvement toward big data analysis, the analysis of the actual data sets, privacy-preserving data mining, the estimation of missing values, were solved by the NIS-Apriori algorithm.

研究分野：　情報学（ソフトコンピューティング）

キーワード：　ラフ集合　非決定情報　アプリオリアルゴリズム　NIS-アプリオリ　データマイニング　相関ルール　欠損値　SQL

１．研究開始当初の背景

(1) 連続値表データの解析には、通常、平均や分散に基づく統計的手法が用いられる。しかし、離散値表データでは平均や分散の定義が難しい場合も多い（例えば、血液型の平均値、目の色や頭髪の色の平均値など)。従って、統計的手法ではなく「血液型＝A」や「目の色＝青」などのデスクリプタによる含意式を用いて表データを特徴付ける研究が進められていた。

(2) 研究代表者は離散値表データ解析手法であるラフ集合理論、さらにデータマイニング手法であるアプリオリ法を利用し、統計解析では扱いにくいデータの特徴付けを研究していた。特に情報の不完全性も扱う理論的枠組の構築と実際のデータ解析支援ツールの実現を進め、一連の研究をラフ集合非決定情報解析と名付けた。トランザクション形式データのためのアプリオリ法を表データ形式用に拡張し、非決定情報も扱う NIS-アプリオリ法を提案していた。

(3) ツール実現において、ラフ集合理論に基づく同値類の利用を検討し、同値類をリストで表現できる都合の良さから論理型言語 Prolog と C を用いて実現を進めた。しかし、実行環境については不十分であり、解析ツールの整備が必要であった。

２．研究の目的

(1) 提案するラフ集合非決定情報解析の枠組自体の確立、解析ツールの強化、さらに現実問題への応用を図ることを研究の目的とし、具体的 4 課題を設定した。4 課題の解決により、目的の達成を目指した。

[課題 1] アプリオリ法は多数派ルールを生成するには適しているが、出現頻度が低い少数派ルールの生成には向かない。しかし、局所的に強い相関を有するルールは意味あるルールと考えられる。少数派意見マイニングのための手法・アルゴリズムを検討し、本問題の解決を図る。

[課題 2] 実データの解析を通しながら、解析ツールの改善とビッグデータ処理への展開を図る。

[課題 3] 情報を意図的に非決定化し、データセキュリティやプライバシー保護分野への応用を進める。「情報の希薄化」の利用可能性を検討する。

[課題 4] ラフ集合と粒状計算による欠損値推定アルゴリズムを研究する。統計分野における欠損値推定と関連付けながら研究を進める。

３．研究の方法

(1) 非決定情報に様相性を導入した枠組みは今までに殆ど無く、研究の方法は従来にない枠組みの提案とそのための解析ツールの実現を図る手順で行った。

(2) 実現した解析ツールの検証には UCI 機械学習レポジトリに公開されている表データを用いた。

４．研究成果

(1) 4 課題についてそれぞれ枠組みの確立や解析ツールの改善を進めることができ、当初の目的を達成できたと考える。以下、順に成果を列挙する。

(2) 研究代表者の提案する枠組みは様相論理における可能世界意味論を継承している。さらに、定義されたルールの集合に対して NIS-アプリオリは健全(得られる含意式はルールである)、かつ完全(任意のルールは必ず得られる)である。このように論理の体系を維持しながら実際のデータを扱う解析ツールまで実現している点に本研究の特徴がある。

(3) [課題 1]についての成果：通常、アプリオリ法では、生起頻度の割合(サポート値)と無矛盾の割合(アキュラシー値)が共に一定以上になる含意式を生成し、生成された含意式を特にルールとよぶ。サポート値の要件を強めると、要件を満たす含意式は減り計算時間も減少する。一方、サポート値の要件を弱めると、要件を満たす含意式は増え計算時間が膨大になることもある。例えば、UCI の Car データセット(レコード数 1728)に対して、サポート値 0.25 とすると、432 レコード以上で成立する含意式を扱い、サポート値 0.001 とすると、2 レコード以上で成立する含意式を扱う。このことからアプリオリ法は多数派を表現する含意式(メジャールールとよぶ)生成に適し、少数派を表現する含意式(マイナールールとよぶ)生成には向かないと考えられる。後者では全数探索とほぼ同等の計算が必要になる。

研究代表者は表データにおける決定属性以外に決定属性値まで指定する手法(ターゲットデスクリプタ付きアプリオリ法)を提案し、解析ツールの実現を行った。決定属性値が 4 つの Car データセットでは同じ操作を 4 回繰り返す。表 1 と表 2 は今までの手法とターゲット付き手法の実行時間の和を比較している。低いサポート値では、ターゲット付きアプリオリ法を属性値の回数だけ繰り返す方が、実行時間面において有効であると思われる。

表 1：Car データセットの実行時間

サポート	アキュラシー	実行時間(sec)	実行時間の和(sec)
0.25	0.6	8.30	13.7
0.10	0.6	38.26	33.83
0.05	0.6	255.23	177.37
0.01	0.6	3343.52	2014.33
0.001	0.6	6004.56	4043.02

表 2：Phishing データセットの実行時間

サポート	アキュラシー	実行時間(sec)	実行時間の和(sec)
0.25	0.6	11.44	15.36
0.10	0.6	350.65	206.67
0.05	0.6	3153.86	995.19
0.01	0.6	11892.18	5761.78
0.001	0.6	15641.92	8825.96

他に balance データセットなどでも同様の傾向になった。決定属性値を指定することで、繰り返しの回数は増えるものの SQL の不要なサーチが減少し、実行時間の和の方が効率的になったものと考える。本研究の成果は現在、投稿中である。

(4) [課題 2]についての成果：Prolog と C 版 NIS-アプリオリの実行ではスタックオーバーフローなどの実行時エラーが頻発したことや、ビッグデータへの対応も考慮し、SQL 版への転換を図った。コマンドプロンプトモード以外にも Xampp と呼ばれるユーザインターフェース付きの実行環境を利用した。SQL のプロシージャ作成、ファイルの管理には Xampp を用い、実行にはコマンドプロンプトを用いた。10 数個の同じ表データセットに対して、独立して構築した 2 つの NIS-アプリオリプログラムは冗長性を除き同じルールを生成した。この結果から、2 つのプログラムが正しく動作していることの確認もできたと考える。本内容は主に論文①、⑨、⑫に対応する。

Prolog と C 版における元データ作成の手間も簡略化された。Prolog では各レコードを data(1,[red,big,high]) や data(2,[[red,blue],[big,medium],high])などと記述する必要があり、csv 形式データの変換を要する。一方、SQL 版では csv 形式ファイルをインポートし、属性の定義を記述すれば十分である。このような操作の簡略化ができたために、UCI 機械学習レポジトリにある欠損値なしデータセット Car Evaluation、Chess、lenses、Balloon、Phishing、Balance 等や欠損値ありデータセット Congressional Voting、Mammographic、Credit Evaluation、Dresses Sales、Hepatitis 等からのマイニングも容易になった。本内容は主に論文①、②に対応する。

[課題 1]の実行例は SQL 版によるもので、オーバーフローの観点から Prolog 版でのサポート値 0.001 の実行は難しいと考える。SQL 版によりビッグデータへの対応のみならず、極めて低いサポート値によるマイニングも可能になった。

(5) 今まで、与えられた非決定情報表におけるマイニングを主体的に研究したが、別の観点から応用を試みた。図 1 は通常の表データ

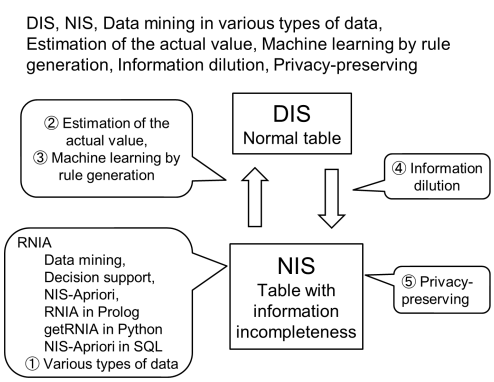

図 1：DIS と NIS に関する論点

(DIS)と非決定情報も含む表データ(NIS)の関係において生じる論点を列挙している。[課題 3]では DIS を意図的に NIS に置き換えることでデータセキュリティの強化やプライバシー保護への応用を図れると考える。[課題 4]は NIS から DIS を推定する論点を扱う。本内容は主に論文④に対応する。

(6) [課題 3]についての成果：情報を意図的に非決定化しプライバシー保護マイニングへの応用を検討した。プライバシー保護における k-匿名性と類似した匿名性を非決定情報によって実現できると考える。回答しにくいアンケート調査において「A または B」などの非決定回答ができれば、回答者にも都合がいい。このようにして得られる回答は非決定情報を含み、従来のツールでの解析には向かないが、NIS-アプリオリ法ではこのようなデータセットのマイニングも可能である。図 2 において、決定情報表 DIS では表ΨとΨのルール集合 Rule(Ψ)が決まる。NISΦではΨ1 からΨn のいずれかかが真の表であり、Rule(Ψ)の部分集合(確実ルールの集合)と上位集合(可能性ルールの集合)により、Rule(Ψ)を把握する。本内容は主に論文⑧に対応する。

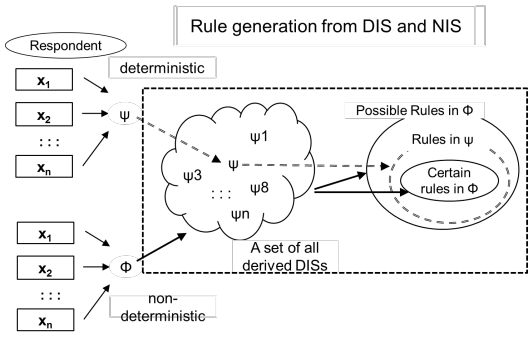

図 2：プライバシー保護アンケート

(7) [課題 4]についての成果：NIS から DIS の妥当な推定を行う手順を提案し、SQL を用いてツールを実現した。統計における最尤推定では観測されたサンプルによる尤度関数が最大になるようにパラメータ値を設定する。これはサンプルの影響を最も大きくする手法と考えられる。この方針を参考にし、研

究代表者は確実ルールの生成後、確実ルールを可能な限り多く出現させるように非決定情報から決定情報を決める手順を提案した。本手法は属性間の依存関係を逐次的に確認し、動的にNISからDISを推定する手順である。欠損値の推定は主に確率分布を用いる場合が多いが、本手法では新たな追加情報は必要ない。また、NIS-アプリオリ法が無ければ本手法の提案もできない。本内容は主に論文④、⑤、⑮に対応する。

(8) 上記、4 課題への回答を付けることにより、提案するラフ集合非決定情報解析の枠組自体の確立、解析ツールの強化、さらに現実問題への応用を進めることができたと考える。

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１９件）

① H.Sakai、K.Y.Shen、M.Nakata、On two Apriori-based rule generators: Apriori in Prolog and Apriori in SQL、Journal of Advanced Computational Intelligence and Intelligent Informatics、Fuji Press、査読有(Accepted)

② H.Sakai、K.Y.Shen、G.H.Tzeng、M.Nakata、Rule based decision support in table data sets with uncertainty and its execution environment、Japonica Mathematica、ISMS、査読有（Accepted）

③ H.Sakai、C.Liu、Association rule-based modal analysis for various data sets with uncertainty、Advances in Intelligent Systems and Computing、Springer、査読有、730 巻、2018、277-286
DOI:10.1007/978-3-319-75792-6_21

④ H.Sakai、M.Nakata、Y.Yao、Pawlak's many valued information system, non-deterministic information system, and a proposal of new topics on information incompleteness toward the actual application、Studies in Computational Intelligence、Springer、査読有、708 巻、2017、187-204
DOI:10.1007/978-3-319-54966-8_9

⑤ H.Sakai、M.Nakata、J.Watada、A proposal of machine learning by rule generation from tables with nondeterministic information and its prototype system、LNCS、Springer、査読有、10313 巻、2017、535-551 DOI:10.1007/978-3-319-60837-2_43

⑥ K.Y.Shen、H.Sakai、G.H.Tzeng、Stable rules evaluation for a rough-set-based bipolar model: A preliminary study for credit loan evaluation、LNCS、Springer、査読有、10313 巻、2017、317-328
DOI:10.1007/978-3-319-60837-2_43

⑦ M.Nakata、H.Sakai、K.Hara、Rough sets in incomplete information systems with order relations under Lipski's approach、LNCS、Springer、査読有、10313 巻、2017、487-506 DIO:10.1007/978-3-319-60837-2_40

⑧ H.Sakai、C.Liu、M.Nakata、S.Tsumoto、A proposal of a privacy-preserving questionnaire by non-deterministic information and its analysis、Proc.IEEE Bigdata、査読有、2016、1956-1965
DOI:10.1109/BigData.2016.7840817

⑨ H.Sakai、C.Liu、X.Zhu、M.Nakata、On NIS-Apriori based data mining in SQL、LNAI、Springer、査読有、9920 巻、2016、514-524
DOI:10.1007/978-3-319-47160-0_47

⑩ H.Sakai、C.Liu、M.Nakata、Information dilution: Granule-based information hiding in table data -A case of lenses data set in UCI machine learning repository-、Proc. Third International Conference on Computing Measurement Control and Sensor Network、査読有、2016、52-55
DOI:10.1109/CMCSN.2016.28

⑪ M.Nakata、H.Sakai、Describing rough approximations by indiscernibility relations in information tables with incomplete information、CCIS、Springer、査読有、661 巻、2016、355-366
DOI:10.1007/978-3-319-40581-0_29

⑫ C.Liu、H.Sakai、X.Zhu、M.Nakata、On Apriori-based rule generation in SQL -A case of the deterministic information system-、Proc. SCIS-ISIS 2016、査読有、178-182、2016
DOI:10.1109/SCIS-ISIS.2016.0047

⑬ H.Sakai、M.Wu、N.Yamaguchi、M.Nakata、Granules for association rules and decision support in the getRNIA system、Intelligent Decision Technologies、IOS press、査読有、9(4)巻、2015、309-320
DOI:10.3233/IDT-140226

⑭ H.Sakai、C.Liu、M.Nakata、Families of the granules for association rules and their properties、LNCS、Springer、査読有、9436 巻、2015、175-187
DOI:10.1007/978-3-319-25754-9_16

⑮ H.Sakai、C.Liu、A consideration on learning by rule generation from tables with missing values、Proc.Advanced Applied Informatics (IIAI-AAI2015)、査読有、2015、183-188 DOI:10.1109/IIAI-AAI.2015.175

⑯ M.Wu、H.Sakai、On parallelization of the NIS-Apriori algorithm for data mining、Procedia Computer Science、Elsevier、査読有、60 巻、2015、623-631
DOI:10.1016/j.procs.2015.08.198

⑰ H.Sakai、M.Wu、N.Yamaguchi、C.Liu、M.Nakata、Reconsideration of rules in tables with non-deterministic data、Proc.IEEE GrC2014、査読有、2014、235-240
DOI:10.1109/GRC.2014.6982841

⑱ H.Sakai、M.Wu、The completeness of NIS-Apriori algorithm and a software tool getRNIA、Proc.Advanced Applied Informatics (IIAI-AAI2014)、査読有、2014、115-121 DOI:10.1109/IIAI-AAI.2014.33
⑲ N.Yamaguchi、M.Wu、M.Nakata、H.Sakai、Application of rough set-based information analysis to questionnaire data、Journal of Advanced Computational Intelligence and Intelligent Informatics、Fuji Press、査読有、18巻、2014、953-961 DOI:10.20965/jaciii.2014.p0953

〔学会発表〕（計１６件）
① H.Sakai、NIS-Apriori based data mining with uncertainty: A Survey of the SQL software tool、Twenty-Sixth International Conference Forum of Interdisciplinary Mathematics、2017
② K.Y.Shen、Conceptual framework for a bipolar weighting model with non-deterministic attributes for group decision-making: Multiple approaches、Twenty-Sixth International Conference Forum of Interdisciplinary Mathematics、2017
③ M.Nakata、Rough approximations by indiscernibility relations in incomplete information tables with continuous domains、Twenty-Sixth International Conference Forum of Interdisciplinary Mathematics、2017
④ 中村赳、MySQL による NIS-アプリオリ処理環境の実現、第 33 回ファジイシステムシンポジウム、2017
⑤ H.Sakai、Association rule-based modal analysis for various data sets with uncertainty、International Symposium on Management Engineering、2016
⑥ H.Sakai、A survey on rough set-based non-deterministic information analysis、International Symposium on Management Engineering、2016
⑦ 酒井浩、プライバシー保護を考慮したアンケートの提案、人工知能学会全国大会、2016
⑧ 酒井浩、ルール生成による欠損値の推定について、第 32 回ファジィシステムシンポジウム、2016
⑨ H.Sakai、A survey on rough non-deterministic information analysis: Table data analysis with other fuzziness、FUZZY51: The 51st Anniversary of Fuzzy Sets、2015（招待講演）
⑩ H.Sakai、A survey on rough set-based non-deterministic information analysis and the next topics、The 24th International Conference of the Forum for Interdisciplinary Mathematics、2015
⑪ 劉臣籍、アプリオリ法による多様な不完全データからの ルール生成について、第 31 回ファジィシステムシンポジウム、2015
⑫ H.Sakai、On the definability of a set and rough set-based rule generation、IIAI-AAI2014 conference、2014
⑬ H.Sakai、Toward the enhancement of the getRNIA system for rough-set based data analysis、SCIS-ISIS2014 conference、2014
⑭ H.Sakai、Rough non-deterministic information analysis and decision support: A survey、17th Czech-Japan Seminar on Data Analysis & Decsion Making、2014
⑮ M.Wu、並列化機能による NIS アプリオリの高速化、第 30 回ファジイシステムシンポジウム、2014
⑯ 山口直人、相関ルール生成における１つの問題点について、第 25 回ソフトサイエンスワークショップ、2014

〔図書〕（計０件）

〔産業財産権〕

○出願状況（計０件）

○取得状況（計０件）

〔その他〕
ホームページ等
Software Tools for RNIA (Rough set Non-deterministic Information Analysis)
http://www.mns.kyutech.ac.jp/~sakai/RNIA/

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

(4)研究協力者
中田　典規　（NAKATA, Michinori）
城西国際大学・経営情報学科・教授

Kao-Yi Shen
Chinese Culture University・
Department of Banking and Finance・
Associate Professor