

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 20 日現在

機関番号：10103

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25330315

研究課題名(和文) ラフ集合による関係性マイニング 感性データ分析の新展開

研究課題名(英文) Rough set-based interrelationship mining - A new approach for Kansei data analysis -

研究代表者

工藤 康生 (KUDO, Yasuo)

室蘭工業大学・工学(系)研究科(研究院)・准教授

研究者番号：90360966

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：本研究は、従来のラフ集合データ分析では困難であった2つの属性間の比較に基づく特徴の抽出を行うことで、ラフ集合による関係性マイニング手法を確立することにより、属性間の関係性をも抽出する新たな感性データ分析手法を構築することを目的とする。本研究では関係性マイニングの提唱と理論整備を行い、また実データに対して関係性マイニングを行う際に必要となる、大規模データに対する縮約計算システムを実装した、更に、関係性マイニングによる感性データ分析を行った。本研究により、ラフ集合による感性データ分析の新たなアプローチを示すことができたと考える。

研究成果の概要(英文)：In this study, we proposed a new approach for Kansei data analysis based on rough set-based interrelationship mining. We firstly proposed the concept of the interrelationship mining that enables us to extract characteristics based on comparison of attribute values between different attributes. In the interrelationship mining, the relationship between attribute values are explicitly described by newly introduced attributes called interrelated attributes. Addition of the interrelated attributes for the interrelationship mining causes increase of the number of attributes. We then constructed rough set-based attribute reduction systems for datasets with many objects and attributes. We also discussed to apply the rough set-based interrelationship mining to Kansei data analysis. Consequently, our study contributed to provide a new approach for rough set-based Kansei data analysis.

研究分野：感性情報学

キーワード：関係性マイニング ラフ集合 感性情報処理

1 . 研究開始当初の背景

ラフ集合を用いたアンケートデータ等の分析手法は、データマイニング技術の感性情報処理への応用として近年注目されている。ラフ集合によるデータ分析では、対象間の識別不能性（互いに区別がつかないという関係）に基づき行われるが、分析結果としてデータから得られる規則性は、個々の属性とその値の組み合わせで表現可能な規則性のみに限られるため、例えば「サンプル使用前の評価より使用後の評価の方がよい」など、2つの属性間の比較に基づく特徴を直接抽出することは困難であり、このような属性間の相互関係をも考慮したラフ集合データ分析は行われていなかった。

2 . 研究の目的

本研究では、感性工学の分野で注目されているラフ集合を用いて、購買履歴やアンケート調査結果等、ユーザの感性を反映したデータに対する新たな分析手法を構築することを目的とする。特に、従来のラフ集合データ分析では困難であった、2つの属性間の比較に基づく特徴の抽出を行うことで、ラフ集合による関係性マイニング手法を確立することにより、属性間の関係性をも抽出する新たな分析手法を構築することを目的とする。

3 . 研究の方法

申請者は既に、2つの属性間の関係性を新たな属性として表現する基本アイデアを発表済みであるため、本研究ではこの基本アイデアを踏まえ、以下の(1)から(3)までの研究を行う。

(1) ラフ集合による関係性マイニング手法の理論整備

2つの属性間の関係性を新たな属性として表現する基本アイデアを更に発展させ、理論的側面を整備すると共に、属性間の関係性を見出すアルゴリズムの開発、関係性を表す新たな属性をも含めた縮約計算、決定ルール抽出手法の開発及びこれらを実現するための要素プログラムの実装等を通じて、ラフ集合による関係性マイニングの基盤を構築する。

(2) 関係性マイニングを用いた感性データ分析システムの構築

ラフ集合による関係性マイニング手法に基づく、感性データ分析システムを計算機上に実装し、データマイニングのベンチマークデータ等を用いた評価実験、及び関係性マイニングを用いない従来の感性データ分析手法との比較実験を通じて、実装したシステムの性能を検証する。

(3) 感性データ分析システムによるデータ分析およびその応用

感性データ分析システムを用いて、実際のアンケートデータ等、人間の感性を反映した実データに対する分析を行い、評価実験を通じてその有効性を検証する。また、

関係性マイニングに基づく感性データ分析システムの応用可能性について検討する。

4 . 研究成果

(1) 研究の主な成果

ラフ集合による関係性マイニングの基本アイデアを示し、その理論整備を行った。具体的には、関係性マイニングの基本的な考え方の提案（学会発表[10]）、関係性マイニングにより得られる相対縮約および決定ルールの理論的性質の整備（学会発表[5,6,8]）、関係性マイニングにより得られた決定ルールの表現能力の検討（学会発表[2]）、欠損値を含むデータに対して関係性マイニングを適用するための理論面の拡張（雑誌論文[1]）を行った。

ラフ集合によるデータ分析では、分析対象となるデータを表1のような決定表と呼ばれる表形式で表し、データを正しく分類するために最小限必要となる属性を抽出する縮約計算や、条件属性（表1では性別、年代、使用前、使用后）の値の組み合わせと決定属性（表1では購入）の値との間の規則性を記述する決定ルールの抽出に基づく。

表1. 決定表の例

ユーザ	性別	年代	使用前	使用后	購入
U1	女性	30	3	3	
U2	女性	20	3	5	
U3	男性	10	4	5	
U4	女性	40	4	4	
U5	男性	20	3	3	x
U6	女性	30	4	3	x

関係性マイニングでは更に、任意の条件属性 a, b および属性値間の任意の2項関係 R を用いて、対象 x における属性 a の値 $a(x)$ と属性 b の値 $b(x)$ との間での関係 R の成立/不成立を、関係性属性と呼ばれる新たな属性 aRb を用いて明示的に記述することにより、2つの属性間の比較に基づく特徴も、ラフ集合の枠組みで抽出することを可能にする（学会発表[10]）。関係性属性を含む決定表の例を表2に示す。

表2. 関係性属性を含む決定表の例

	性	年	前	後	前後	購
U1	女	30	3	3	1	
U2	女	20	3	5	1	
U3	男	10	4	5	1	
U4	女	40	4	4	1	
U5	男	20	3	3	1	x
U6	女	30	4	3	0	x

表2において、関係性属性「前後」は属性「使用后」の評価値が属性「使用前」の評価値以上であるときに値1を、そうでないときに値0を取る新たな属性である。このような関係性属性を含む決定ルールとして、表2

からは例えば

(性別=女性) ((前 後)=1) (購入=)
などの決定ルールが得られる。関係性属性を含む相対縮約から得られた決定ルールは、ルールの汎用性の指標である被覆度の平均値が、通常の決定ルールと少なくとも同程度以上となる(学会発表[2])。そのため、関係性マイニングにより関係性属性を含む相対縮約を抽出し決定ルールを求めることで、より有用な規則性をデータから抽出することが可能となる。

関係性マイニングによるデータ分析では、属性間の関係性成立の有無を新たな属性(関係性属性)を用いて記述するため、必然的にデータ中の属性の個数が増加する。よって、関係性マイニングによるデータ分析を行うためには、大規模データに対する縮約計算の高速化が必須である。これを踏まえ、申請者の従来研究であるランダムラフ集合モデルに関する研究(平成23-24年度若手研究(B), 課題番号23700244)から得られた知見に基づき、Bazan et al.による縮約計算の統計的アプローチであるGeneralized Dynamic Reduct (GDR)の抽出の並列化(学会発表[4])およびその高速化(学会発表[3])を行った。

GDR抽出手法の並列処理を計算機サーバ上で実装し、更にその高速化を行った。その性能を計算機実験により評価した結果を表3に示す(学会発表[4])。実験ではUCI Machine Learning Repositoryから取得した3種類のデータを使用した。その詳細を表4に示す。

表3. GDR抽出の計算時間(単位:秒)

データ名	高速化前	高速化後
Letter Recog.	917	456
Nomao	2396	29
Diabetes	10226	782

表4. 使用したデータ

データ名	データ件数	属性数
Letter Recog.	20000	16
Nomao	34465	119
Diabetes	101766	49

実験の結果、3種類のデータすべてで、計算時間の短縮が確認できた。特に、属性の個数が多いNomaoデータセットでは、計算時間を高速化前の約1%程度まで短縮することができた。合わせて、3種類のデータすべてで、高速化を行う前と行った後で、得られるGDRは同一であることも確認した。上述の通り、関係性マイニングでは関係性属性を新たに用いるため、必然的に属性の個数が増加する。実験の結果より、本研究で実施したGDR抽出の並列化およびその高速化により、関係性マイニングによる感性データ分析を十分に現実的な計算時間で実施するための計算基盤が整備できた。

関係性マイニングによる感性データ分析およびその応用として、関係性マイニングの遺伝子発現データ分析への適用について検討した(学会発表[9])。また、関係性マイニングを用いた情報推薦手法について基本的なアイデアを示し(学会発表[7])、更に関係性マイニングを用いた楽曲推薦システムを試作し、実験に基づく評価を行った(学会発表[1])。

関係性マイニングを用いた情報推薦手法では、推薦を受けるユーザは推薦対象の商品等のごく一部に対して5段階評価等の価値判断を行い、その結果を関係性属性を用いて記述する。関係性属性として表現されたユーザの嗜好のパターンを用いて、例えば類似した嗜好パターンを持つ他のユーザを探し、そのユーザが高く評価する商品を推薦結果とすることにより、情報推薦を行うことが可能である(学会発表[1])。関係性マイニングによる情報推薦手法の概念図を図1に示す。

ユーザによる価値判断

札幌:4, 旭川:3, 函館:5, 登別:5

ユーザの嗜好パターンを抽出

札幌<旭	札幌<函	札幌<登	...
0	1	1	...

嗜好パターンの類似するユーザを検索

札幌:3, 旭川:3, 函館:4, 登別:4, 洞爺:5

札幌<旭	札幌<函	札幌<登	...
0	1	1	...

このユーザが高評価の「洞爺」を推薦

図1. 関係性マイニングによる情報推薦手法(観光地推薦の事例)

(2)得られた成果の国内外における位置づけおよびインパクト

研究期間全体を通しての成果として、ラフ集合による関係性マイニングの理論的基礎を整備し、大規模データからの縮約計算・データ分析システムを構築したこと、および関係性マイニングによる感性データ分析の事例を示したことにより、関係性マイニングに基づく感性データ分析の新たなアプローチを示すことができたと考えている。

研究開始当初の背景で述べた通り、2つの属性間の比較に基づく特徴を直接抽出することは困難であり、このような属性間の相互関係をも考慮したラフ集合データ分析は行われていなかった。これに対して、本研究による成果を用いることで、項目間の比較に基づく特徴をも抽出することが可能となったため、例えばアンケートデータでの設問間の回答内容の比較に基づく特徴の抽出など、より詳細なデータ分析が行えるようになった。よって、本研究の成果は、ラフ集合理論の感性情報処理への応用面に進展をもたらすと期待できる。

(3)今後の展望

ラフ集合による関係性マイニングの考え方をを用いて、サンプル間の比較に基づくユーザの感性的価値判断を踏まえた情報推薦手法の具体化を目指す。特に、サンプル間の比較結果を関係性属性として表現することにより、ユーザの嗜好に含まれる矛盾をも感性的特徴として反映した情報推薦手法を構築する。本研究で得られた知見を踏まえ、関係性マイニングに基づく情報推薦の具体的なアルゴリズムの構築および実装、実験による検証などを行う予定である(嗜好の矛盾を許容した情報推薦 関係性マイニングによる新たなアプローチ, 平成 28 年度基盤研究 (C), 課題番号 16K00365)。

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 1 件)

Yasuo Kudo and Tetsuya Murai, Rough-set-based Interrelationship Mining for Incomplete Decision Tables, JACIII, 査読有, Vol.20, No. 4, to appear.

[学会発表](計 10 件)

比田啓貴, 工藤康生, 岡田吉史, 村井哲也, 関係性マイニングを用いた楽曲推薦手法の提案, 計測自動制御学会システム・情報部門学術講演会 2015, 2015 年 11 月 18 日, 函館アリーナ(北海道・函館市)。

Yasuo Kudo and Tetsuya Murai, On Representation Ability of Interrelated Attributes in Rough Set-based Interrelationship Mining, ISIS 2015, 2015 年 11 月 6 日, Mokpo (Korea)。

高橋智, 工藤康生, 村井哲也, ラフ集合における Generalized Dynamic Reduct の抽出手法の改良, 第 31 回ファジィシステムシンポジウム, 2015 年 9 月 4 日, 電気通信大学(東京都・調布市)。

高橋智, 工藤康生, 村井哲也, ラフ集合における Generalized Dynamic Reduct の並列抽出, 第 25 回ソフトサイエンスワークショップ, 2015 年 3 月 11 日, 下関市立大学(山口県・下関市)。

Yasuo Kudo and Tetsuya Murai, Some Properties of Interrelated Attributes in Relative Reducts for Interrelationship Mining, SCIS & ISIS 2014, 2014 年 12 月 5 日, 北九州国際会議場(福岡県・北九州市)。

Yasuo Kudo and Tetsuya Murai, Interrelationship Mining from a Viewpoint of Rough Sets on Two Universes, IEEE GrC

2014, 2014 年 10 月 22 日, 登別グランドホテル(北海道・登別市)。

工藤康生, 松井美佳, 村井哲也, ラフ集合による関係性マイニングを用いたレコメンデーションシステム, 感性フォーラム札幌 2014, 2014 年 2 月 8 日, 札幌市立大学サテライトキャンパス(北海道・札幌市)。

Yasuo Kudo and Tetsuya Murai, Decision Logic for Rough Set-based Interrelationship Mining, IEEE GrC 2013, 2013 年 12 月 14 日, Beijing (China)。

Yasuo Kudo, Yoshifumi Okada, and Tetsuya Murai, On a Possibility of Applying Interrelationship Mining to Gene Expression Data Analysis, AMT-BHI 2013, 2013 年 10 月 29 日, 前橋テルサ(群馬県・前橋市)。

工藤康生, 村井哲也, ラフ集合による関係性マイニングの構想, FSS 2013, 2013 年 9 月 9 日, 大阪国際大学枚方キャンパス(大阪府・枚方市)。

6. 研究組織

(1)研究代表者

工藤 康生 (KUDO, Yasuo)
室蘭工業大学・工学研究科・准教授
研究者番号: 9 0 3 6 0 9 6 6

(2)研究分担者

村井 哲也 (MURAI, Tetsuya)
北海道大学・情報科学研究科・准教授
研究者番号: 9 0 2 0 1 8 0 5

(3)連携研究者

岡田 吉史 (OKADA, Yoshifumi)
室蘭工業大学・工学研究科・准教授
研究者番号: 0 0 4 4 3 1 7 7