

称号及び氏名	博士（経済学）西口 真央
学位授与の日付	平成29年3月31日
論文名	「必須コントラストパターンを利用した分類モデルに関する研究」
論文審査委員	主査 森田 裕之 副査 荒木 長照 副査 中山 雄司

論文要旨

本稿では、特徴的パターンの1つであるコントラストパターンを利用した、新たな分類モデル構築手法 **CAECP** を提案している。特徴的パターンを利用した分類モデルは、高い精度で予測可能であることに加え、モデルに利用されたパターンから有用な知識を抽出できる点が大きな長所として知られている。したがって、単に予測するだけではなく、複数のクラス間の特徴や差分を分析し、次のアクションにつなげることが必要なビジネスシーンでよく利用される傾向にある。本研究は、特徴的パターンベースの分類モデルの改善および、モデルのビジネス領域への適用を試みる応用研究に位置付けられるものである。

特徴的パターンをベースとする代表的な分類モデル構築手法は、**CAEP** や **CACP** など多くの手法が存在する。どの手法でも、モデルに利用するパターンにスコアを与え、そのスコアに従いトランザクション **ID**(以下 **TID** と呼ぶ) に対してクラス投票を行うことでクラスを決定するという非常にシンプルなモデル構築手順でありながら、高い予測精度のモデルを構築できることが実証されている。しかし、いくつかの事例研究の中で、これらの手法は問題によってはモデル構築に非常に多くのパターンを必要とすることが明らかになった。これはモデルの可読性の観点からは望ましいことではない。そこで本稿で提案する分類モデル構築手法である **CAECP** は、予測精度を落とさずに、モデルに利用するコントラストパターンを極力削減することで、より可読性を高めることを目的としている。また、モデルに利用された少数のコントラストパターンを効率的に把握するための新たな視覚化手法 **CRPD** も提案している。

本稿の構成は以下のとおりである。1章は序章として、本研究の目的や工夫点、既存研

究や本稿の構成などを概観している。2章では、本研究を説明するための基礎的内容や既存研究を説明し、3章で提案手法の詳細を記載している。4章では、3つのベンチマークデータセットに対する計算実験と、2つの実ビジネスデータに対して提案手法を適用し、有効性を確認している。

1章ではまず、特徴的パターンベースの分類モデルが誕生した経緯や、特徴的パターンの応用研究の広がりなどを説明し、本研究の位置付けを説明する。その後、既存の特徴的パターンベースの分類モデルの構築手法や報告事例などを紹介している。そして既存研究の中で明らかになった既存手法の問題点を考察し、既存研究を踏まえての本研究の改善箇所を明確にする。最後に本稿の構成をまとめている。

2章は、本稿で取り扱うデータ形式や特徴的パターンの定義など、基礎的な内容を説明したうえで、2つの代表的な特徴的パターンベースの分類モデルである **CAEP** と **CACP** について詳細に議論する。さらに、これらの分類モデルを適用した応用研究から明らかになった課題の1つとして、問題によってはモデルに利用するパターンが多くなってしまい、モデルの可読性が損なわれていることを指摘し、その課題を解決するために従来のパターン選択基準の他に、新たなパターン採用基準が必要であることを説明する。2章ではさらに、このような課題に取り組んでいる既存研究も紹介し、良い点や悪い点を整理する。

3章は、本稿で提案する2つの手法を順に説明している。提案するパターン選択手法は、クラス内の **TID** がどの程度パターンにカバーされているかと、**1TID** あたり何個のパターンに該当するかなどを停止条件として、ある選択状態で最も分類精度を高めるパターンを1つずつ選択する手法である。また、簡単な出力モデルを例として、視覚化手法 **CRPD** の作図方法を説明する。**CRPD** は、コントラストパターンの特徴である対称性を強調した図である。さらに、パターン間やアイテム間の関連性や、パターン長の区別などを行うことで、従来の図より多くのパターン情報を含み、かつ直感的な理解が可能なように工夫を凝らしたパス図となっている。

4章ではまず、3つの有名なベンチマークデータに対して **CAECP** を適用し、既存手法との比較実験を行っている。実験結果により、問題によっては採用パターン数を **10%** 以下にまで抑えることに成功した。また、2つのデータセットに対しては正答率も上昇する結果となった。以上の結果から、提案手法は正答率を保ちつつ、採用パターン数を減らすことが可能であることが示された。4章ではつづいて、ある製造小売業のデータに対して、購買額が高い顧客のうち、長期的に購買を継続するか、またどの程度アクティブであり続けるかといった観点から、2つの分類問題を設定した。実験結果は、既存手法より高い予測精度を出しつつ、採用パターン数も従来の手法より大幅に削減することができた。また、これらの分類問題を解くにあたり、基礎集計を踏まえて様々な説明変数を作成し、**CRPD** によって描画されたモデルを考察することで、実ビジネスに役立つ有用な知見を得ることができた。4章ではまた、複数のスーパーマーケットの **POS** データに対して手法を提案している。この実験では、各店舗レベルで、注力すべき商品分類を決定するためのヒントとなる指標の提案

も行なっている。その指標を用いて、ある店舗では刺身類を注力すべきであると定め、刺身類を積極的に購買している顧客とそうでない顧客を分類する問題を設定した。計算結果では、この問題でも正答率と利用パターン数の少なさの両方で既存手法を上回ることに成功した。

しかし、本稿で提案する手法は事前に設定すべきパラメータが多く、最適なパラメータ設定を発見するまでに時間がかかることが課題としてあげられる。実際にいくつかの実験結果では、まだ最適なパラメータ設定ではないような結果も見られた。今後の課題として、焼きなまし法などの局所探索アルゴリズムの適用を行うことで、パラメータ設定の手間をなるべく省くことができるように改善していきたい。さらに、より多くの実ビジネスデータに対して手法を適用し、実際のマーケティング施策につなげることで、より社会に貢献していきたいと考えている。

審査結果の要旨

本論文は、クラス分類問題に対して、特徴的パターンの1つであるコントラストパターンを利用した新たな分類モデル構築手法 CAECP を開発している。特徴的パターンを利用した分類モデルは、説明変数間の関連性を取り入れてモデルを構築することが可能であるため、他の説明変数の独立を前提としている分類モデルとはタイプが異なるモデルであるといえる。また、モデル構築に利用されたパターンから、決定木モデルのような可読性のあるルールを識別できることも、実用上の大きな長所であるといえる。

過去の関連研究としては、CAEP や CACP などの手法が提案されていた。これらの基本的なフレームワークは、特徴的なパターンを抽出し、特徴的であるほどクラス予測に貢献する値を大きくするようなスコアを算出し、識別対象のスコアの大小によってクラスを予測する投票モデルが主体である。このフレームワーク自体は、非常に優れているものの、実際のデータに対しては、モデル構築に非常に多くのパターンを必要とする事例が多く見られた。それはモデルの予測精度の向上からも、可読性の観点からも望ましいことではない。本論文では、予測精度を維持、または向上しながら、モデル構築に必要な十分な必須コントラストパターンを識別してモデル構築に成功し、結果として、可読性を高めることにも成功している。さらに、モデルに利用された必須コントラストパターンを効率的に表現する方法として、新たな視覚化手法 CRPD も提案している。既存研究では、モデル構築に様々な工夫がされていたが、最終的なモデルの視覚化についても新たな手法を提案していることは、この論文の大きな貢献の1つであるといえる。

論文の構成は、1章で本研究の目的や貢献について述べた後、2章で本研究を説明するための基礎的内容や既存研究との関係について説明している。3章では、提案手法である CAECP とその出力モデルの視覚化手法である CRPD について詳細を説明している。4章では、まず提案手法をこれまで他の研究論文でも使用され、既存手法のパフォーマンスが確認されている3つのベンチマークデータセットに対して計算実験を行い、既存手法のパフォーマンスと比較して、より少ない必須コントラストパターンを用いて、優れた結果であることを示している。その後、製造小売業とスーパーマーケットの異なる2つの実ビジネスデータに対して提案手法を適用し、同様に有効性を確認するとともに、CRPD による視覚化の効果についても確認している。

本論文の評価すべき貢献は、既存の手法の改良モデルではあるものの、その基本的な欠点を明確に指摘し、それを解決していること、そして単にデータに適用するだけでなく、複数のベンチマーク問題に対して適用して、その客観的な有用性を確認していること、実データにも利用可能であることも複数の実データに対して適用して確認していること、そして、モデルのパフォーマンスのみならず、実用性を重視する点から、モデルの視覚化手法についても開発し、実際の結果を示している点にある。残っている課題としては、提案

手法の設定すべきパラメータが複数あり、最適なパラメータ設定が困難な場合があることである。しかし、実際上はパラメータの目安は示されており、最適解にこだわらなければならないような場合を除いて、実際の利用上は大きな障害ではない。

本論文を作成するには、これまで国内学会の全国大会で4件、国際会議で3件の報告を行っていると共に、論文として査読付きが5本、査読なしが1本を発表している。またこれらのうち、5つは英語での報告または論文であり、英語能力も十分であると判断できる。研究内容に関しても、2015年10月に大阪で開催された **The 11th International Conference on Knowledge Management** で行った研究報告に対しては、学生の中から選ばれた賞であるが大阪府知事賞を受賞しており、対外的にも十分認められた研究内容であると言える。

以上から総合的に判断し、博士（経済学）の学位を認めるについて所定の水準に達しており、合格とすることが適当であると判断する。