

称号及び氏名 博士（工学） 桑島 功

学位授与の日付 平成 21 年 3 月 31 日

論文名 「進化型多目的最適化による識別器および  
相関ルールの獲得に関する研究」

論文審査委員 主査 石淵 久生  
副査 市橋 秀友  
副査 松本 啓之亮

## 論文要旨

近年、計算機の高性能化とデータベースの大型化に伴い、大規模な数値データを取り扱うことが一般的になり、自動的に価値のある情報を抽出することの必要性が高まっている。データマイニングとは、そのような大規模データから非自明で有用な知識を獲得する手法の総称である。相関ルールマイニングは、その中で最も多く用いられる手法の1つである。相関ルールマイニングでは、「もし A が存在するならば B も存在する」という If-then 形式で知識を獲得する。相関ルールマイニングは、その簡潔性、利便性から、売上データの解析など様々な実問題に応用されてきた。近年では、相関ルールマイニング手法をパターン識別問題に適用した手法が提案されている。

相関ルールマイニング手法は、一般に信頼度と支持度に閾値を設けることでルールを抽出する。閾値の設定によっては、抽出されるルールが非常に多くなる。抽出されたルールを用いて識別器を設計する場合、そのままでは識別器としての最適化が成されていないため、識別性能が高くないという問題が存在する。また、識別器が過度に複雑で、人間がそれを解釈することが難しいという問題も存在する。

これらの問題を解決する手法として、遺伝的ルール選択が考えられる。遺伝的ルール選択は、遺伝的アルゴリズムにより識別器に必要なルールのみを選択する手法である。遺伝的ルール選択では、精度を最大化し、複雑性を最小化するために適応度関数に加重和を用いることが考えられる。この加重和適応度関数では、重みを事前に設定することが難しいという問題がある。もし、満足する解が見つからなかった場合には、重みを設定し直して、再度実行する必要がある。このような問題点に対処するためには、精度と複雑性を同時に最適化することが必要となる。

数多くのルールから識別器を獲得する場合、多大な計算時間を必要とするという問題も存在する。この問題を解決するために、相関ルールマイニングの際に、候補ルールの絞込みを行うということが考えられる。そのような候補ルールの絞込みとしてパレート最適ルールのみを使用する方法がある。パレート最適

ルールは、信頼度と支持度の点でパレート最適ルールであり、信頼度と支持度以外の様々なルール評価基準を最大化することが知られている。

パレート最適ルールのマイニングに着目した場合、使用可能な従来手法として **Apriori** という手法がある。この手法は、計算負荷軽減のために支持度やルールの条件数に制限を設ける必要がある。この問題に対して、支持度やルールの条件数に制限を設けず、直接、信頼度と支持度を最適化し、パレート最適ルールをマイニングできる高速な手法の開発が望まれる。

進化型多目的最適化は、進化型計算の多点探索という特徴を利用して、複数の目的を同時に最適化することが可能な手法である。つまり、上述した識別器の設計においては、精度と複雑性を同時に最適化することができる。また、関連ルールのマイニングにおいては、ルールの信頼度と支持度を同時に最適化することができる。

本論文では、進化型多目的最適化を用いて、ルールに基づく識別器、および、関連ルールの獲得について論述した。第2章以降の各章の内容は以下の通りである。

第2章「関連ルールマイニングの後処理としての遺伝的ルール選択による識別器の獲得」では、関連ルールマイニングにより抽出されたルールに対する遺伝的ルール選択の適用について議論した。遺伝的ルール選択では、識別器の精度最大化と複雑性最小化を目的とした加重和適応度関数を使用し、単一目的最適化の枠組みの中で、高性能かつ単純な識別器の獲得を行った。ルールの条件部集合として区間集合とファジィ集合の2つを使用し、それぞれについて遺伝的ルール選択の効果を調査した。

数値実験では、遺伝的ルール選択を行うことで、ルール数が大幅に削減されることを示した。また、学習用データに対する識別性能が、信頼度と支持度の閾値設定に依存することなく改善されることを示した。さらに、評価用データに対しても、識別性能が改善することが多いことを示した。ただし、加重和適応度関数を使用する点において、利用者の選好に応じた解を獲得するためには、適応度の重みの設定が難しいことも示した。

第3章「進化型多目的最適化による識別器の獲得とパレート最適ルール」では、進化型多目的最適化を用いたルール選択により識別器の獲得を行い、獲得された識別器に含まれるルールの特徴について議論した。進化型多目的最適化アルゴリズムとして実装が容易で高性能な **NSGA-II** を使用し、様々な精度と複雑性を持つ識別器の獲得を行った。また、獲得された識別器に含まれる各ルールの特徴を、信頼度と支持度の観点から調査した。

実験結果から、進化型多目的最適化を用いることで、重みの再設定を行うことなく、複数の識別器を1回の実行で獲得できることを示した。特に、ルールの条件部にファジィ集合を用いた場合には、使用者に解釈可能な形で識別器を提示できることを示した。また、獲得された識別器に含まれるルールを調査することで、識別器に含まれるルールの多くは、信頼度最大化と支持度最大化の2つの目的でパレート最適、または、それに近いルールであることを示した。得られた知見に基づき、パレート最適ルールのみから識別器を獲得することを提案した。そして、パレート最適ルールから獲得された識別器と、全ルールから獲得された識別器の比較を行った。実験結果から、パレート最適ルールを用いた場合には、学習用パターンに対する識別性能は悪化するが、評価用パターンに対しては大きく変化しないことを示した。さらに、パレート最適ルールを用いた場合には、大幅に計算時間が短縮されることを示した。

第4章「パレート最適ルールの条件緩和とその有効性」では、パレート最適ルールの条件を緩和した  $\epsilon$ -パレート最適ルールからの識別器の設計について議論した。パレート最適に近いルールを表現するために、 $\epsilon$ -優越性の考えに基づき、パレート最適ルールの条件緩和を行った。

数値実験では、 $\epsilon$ の値の設定と獲得される識別器の性能について調査した。 $\epsilon$ の値を適切に設定することにより、短時間で、パレート最適ルールよりも識別能力の高い識別器を獲得できることを示した。また、 $\epsilon$ -パレート最適ルールには、信頼度が1で支持度が低いルールが多く含まれることを示し、そのようなルールが含まれないように $\epsilon$ -パレート最適ルールの定義を修正した。修正の結果、同じ $\epsilon$ の値では、修正前よりも短時間で同程度の性能を持つ識別器を獲得できることを示した。

また、提案手法の有効性を示すために、様々なルール評価基準の値が高いルールから識別器を獲得した場合と、 $\epsilon$ -パレート最適ルールから獲得した場合との比較を行った。様々なルール評価基準について信頼度-支持度空間上での分布と識別性能の調査を行い、ルール評価基準を3つの集団に分類できることを

示した。

第5章「区間集合のファジィ化による識別器の性能改善」では、学習用パターンの分布を考慮した区間集合から台形型ファジィ集合を生成することで、識別器の性能改善を試みた。

数値実験では、2つのファジィ度の設定方法を調査した。1つは、識別器の学習前にファジィ度を設定し、途中で変化させない方法である。もう1つは、識別器の学習前にファジィ度を設定し、学習を行った後に変化させる方法である。実験結果から、学習前に設定する場合は、ファジィ度が低ければ学習用パターンに対する識別率が高いことを示した。一方、ファジィ度が、ある程度高い場合には、評価用パターンに対する識別率が高いことを示した。また、学習後にファジィ度を変化させる方法では、学習前にファジィ度を低い値に設定し、学習後にファジィ度を高い値に変化させることで、高い汎化能力が得られることを示した。

第6章「進化型多目的最適化によるパレート最適ルール獲得」では、進化型多目的最適化によるパレート最適ルールのマイニング手法を提案し、その性能について議論した。提案手法では、遺伝的機械学習におけるミシガンアプローチの枠組みで、個体の定義を行った。そして、進化型多目的最適化の目的関数を信頼度と支持度の最大化に設定することで、パレート最適ルールの獲得を行った。

提案手法では、NSGA-IIの他にMOEA/Dと呼ばれる進化型多目的最適化アルゴリズムを使用した。数値実験では、Aprioriアルゴリズム等の従来の相関ルールマイニング手法との比較を行い、MOEA/Dを用いた提案手法が、Aprioriアルゴリズム等よりも高速に実行可能であることを示した。さらに、提案手法により獲得されるルールは、最大ルール長の制約があるAprioriアルゴリズムと比較して、信頼度と支持度の値が高いルールが獲得できることを示した。また、問題特有の知識を組み込んだヒューリスティック初期化を行い、計算時間を大きく犠牲にすることなく、探索能力が改善されることを示した。今回用いたヒューリスティック初期化では、Aprioriアルゴリズムが初期ルール長を1に設定するのと同様に、初期世代でのルール長の期待値が1になるように設定した。他の様々な初期ルール長と比較することで、ルール長の期待値が1のヒューリスティック初期化の妥当性を示した。

第7章「結論」では、本研究で得られた結果についての総括を行った。

## 審査結果の要旨

本論文は、ルールに基づく識別器の獲得および相関ルールの獲得への進化型多目的最適化手法の応用について提案、調査したものである。ベンチマーク問題を用いた実験的解析により、以下の成果を得ている。

- (1) 相関ルールマイニングの後処理として遺伝的ルール選択により識別器を設計することで、識別に用いるルール数が大幅に削減されることを示した。また、学習用データに対する識別性能が、相関ルールマイニングの信頼度と支持度の閾値設定に依存することなく改善されることを示した。さらに、評価用データに対しても、識別性能が改善することが多いことを示した。
- (2) 進化型多目的最適化を用いたルール選択により識別器の獲得を行い、様々な精度と複雑性を持つ識別器が獲得可能であることを示した。また、獲得された識別器に含まれるルールを調査することで、識別器に含まれるルールの多くは、信頼度最大化と支持度最大化の2つの目的でパレート最適、または、それに近いルールであることを示した。
- (3) パレート最適ルールのみから識別器を獲得することを提案した。提案手法により、学習用パターンに対する識別性能は悪化するが、評価用パターンに対しては大きく変化しないことを示した。さらに、大幅に計算時間が短縮されることを示した。

- (4) 学習用パターンの分布を考慮した区間集合から台形型ファジィ集合を生成することで、識別器の性能改善が可能であることを示した。特に、台形型ファジィ集合のファジィ度を学習前に低い値に設定し、学習後に高い値に変化させることで、高い汎化能力が得られることを示した。
- (5) 進化型多目的最適化によるパレート最適ルールマイニング手法を提案し、Apriori アルゴリズム等の従来の相関ルールマイニング手法と比較して高速に実行可能であることを示した。また、問題特有の知識を組み込んだヒューリスティック初期化を行い、計算時間を大きく犠牲にすることなく、探索能力が改善されることを示した。

以上の研究成果は、知能情報工学分野における進化型多目的最適化によるデータマイニング手法の発展に貢献するところ大である。また、申請者が自立して研究活動を行うに必要な能力と学識を有することを証したものである。学位論文審査委員会は、本論文の審査および最終試験の結果から、申請者に対して博士（工学）の学位を授与することを適当と認める。