

学 位 論 文 題 名

多値型テスト特徴によるパターン分類法に関する研究

学位論文内容の要旨

学習的なパターン認識の過程において訓練パタンの選択および特徴量の設計は重要課題であり、多くの認識システムにおいてその性能を大きく左右する。これに対して、訓練パターン数が比較的少ないかつ使用特徴量の特性が不明瞭という不良条件においても、訓練パターンを十分に区別することのできるテスト特徴を選択することにより効果的な識別が行える“テスト特徴分類法”が提案され、実問題などへの応用を通してその有効性が確認されている。しかしながら、元来論理的な解析を目的にして開発された同手法は2クラス問題を対象とする上に2値で表現された特徴量を扱っている。そこでは多値特徴量を扱う場合に、その特徴量を任意の2値表現にコード化する必要があり、それに伴う特徴次元数の増加による学習及び識別処理のための計算時間が増大し、比較的少ない種類の特徴量しか扱えないという利用しにくいという問題があった。

本研究の目的は、テスト特徴分類法における2クラス問題から多クラス問題へ、かつ、2値特徴から多値特徴へと扱えるようにアルゴリズムの拡張を行い、そのパターン認識のための実応用における可能性について検討することにある。これにより、クラス数の限定されないパターン認識手法としての一般性をもち、多値パターンをも直接扱えるようになり、上述の問題が解決される。テスト特徴分類法は識別に有効な特徴量の選択に基づく手法である。そこで、本研究では拡張したテスト特徴の応用として、テスト特徴およびテスト特徴分類法の基本構造を利用した複数のk-近傍分類法を組み合わせる手法を提案する。使用する訓練パターンおよび特徴量に依存する手法として指摘されるk-近傍分類法に対して、テスト特徴を導入することによりその識別性能の安定および向上を図る。

本論文では、テスト特徴分類法における多クラスおよび多値への拡張、およびテスト特徴のk-近傍分類法の組合せ識別への応用に関する定式化を行い、それらについてのシミュレーション実験による定量的評価に基づく特性解析を行う。さらに、提案手法を実際的な問題に適用することで、実パターン認識問題に対しても利用可能な一つの手法であることを明らかにする。

第1章では、研究の背景と位置づけ、目的および意義について述べる。

第2章では、2クラス問題および2値特徴を扱う従来のテスト特徴分類法を述べる。この章では同手法の定式化を示し、訓練パターンを誤認識しない特性、長所および短所について述べる。また、人工的なデータを用いて、本手法によって生成される識別境界線を図示し、それに関する特性を論じる。テスト特徴分類法は使用する n 次元特徴に対して $m(\leq n)$ 次の識別関数を生成することにより、訓練パタンの分布に依らない効果的な識別が行える。

第3章では、テスト特徴分類法の多クラス問題への拡張に関する定式化を行う。テスト特徴の定義および未知パターンに対する識別処理の流れの理解を補助するために簡単な数値例を挙げる。先に述べた訓練パタンの分布に依らない特性を調べるために、この章ではテクスチャ分類における部分クラス問題を取り上げる。部分クラス問題は、一つのクラスに複数の部分クラスを含むような問題である。任意の問題に対して不変性をもたないテクスチャ解析のための汎用的な特徴である同時生起行列を使用する。ただし、2値表現を行うために同時生起行列からランク特徴を定義する。

第4章では、これまでの同一性評価によるパタンの一致判定の代わりに、距離による類似判定を導入し、新たにテスト特徴を定義することにより、直接多値パターンにも対応できるようにアルゴリズムを拡張する。分離困難な問題として知られている人工データを用いてテスト特徴およびテスト特徴分類法の定量的な評価実験を行う。さらに、よく利用される2種類のパターン認識データベース、すなわち英文字画像認識および画像分割問題に対して認識実験を行い、提案手法の実問題における応用の可能性を検討する。

第5章では、特徴選択法としてのテスト特徴の応用を述べる。k-近傍分類法の識別性能の新しい改良法として、テスト特徴及びテスト特徴分類法の基本構造を利用した新しいパターン分類法を提案する。テスト特徴をk-近傍分類法へ導入することの有効性、及び提案手法の基本的な認識特性について述べる。第4章で使用したデータを用いた認識実験を行い、従来のk-近傍分類法およびランダム特徴選択法との比較実験を通して、テスト特徴の導入はk-近傍分類法の性能向上および安定をもたらす効果を示す。

第6章では、テクスチャ画像の分類問題を例に、テスト特徴分類法の実応用について述べる。対象は、地質調査で用いられる岩石画像である。ここでは第3章で述べたランク特徴を用いる。ただし、多値のままのランク特徴を使用する。任意の訓練パターン画像によって構築されたテスト特徴分類器を用いて、別の未学習の画像に対する分類実験を行い、手法の有効性を示す。

第7章では、テスト特徴分類法における課題についてまとめ、考察を行う。本研究では同一または距離のみの評価を用いてテスト特徴を定義しているが、多様な問題に対応するためにはそれら以外の評価パラメータを導入する必要がある。また、テスト特徴抽出の逐次的な学習処理の可能性、テスト特徴への重み係数の導入について議論する。

第8章は、本研究のまとめである。

学位論文審査の要旨

主査	教授	五十嵐	悟
副査	教授	土谷	武士
副査	教授	島	公脩
副査	教授	岸浪	建史
副査	助教授	金子	俊一

学位論文題名

多値型テスト特徴によるパターン分類法に関する研究

パターン認識の分野において認識法に関する研究が従来から行われており、様々な手法が提案されている。本研究は学習的な認識法の開発に関する次の2つの背景をもつ。ひとつは、訓練パターンが少ないという現実的なパターン認識の問題に対して、近年、ノンパラメトリックな立場からの対策法としてテスト特徴分類法が提案されているが、元来論理回路解析のために開発された同手法には、2クラス問題かつ2値で表現されたパターンを対象とするという制限があるため、汎用性が低く、また手法の特性が明らかにされておらず、実問題へ適用するには困難な場合がある。もうひとつは、そのノンパラメトリックな認識法の中で、識別器の設計が簡明であり有効な手法として幅広く利用されているk-近傍分類法に対して、その識別性能を向上させることを目的とする研究が少ないということである。両手法において、より複雑かつ多様なパターン認識の問題に対応するためには今後の改善が強く期待されている。

本論文は、(1) 上述のような状況にあるテスト特徴分類法に対して、多クラス問題及び多値パターンを扱えるようにアルゴリズムを拡張することにより、特定の問題に限定されない汎用性をもたせること（第3章及び第4章）、(2) k-近傍分類法に対して、拡張した多値型テスト特徴及びテスト特徴分類法の基本構造を利用した新しい組合せ分類法を提案することにより、k-近傍分類法の識別性能を向上させること（第5章）、について研究したものである。

本論文において得られた成果は次の通りである。

第2章において、従来の2クラス問題を対象とする2値型テスト特徴分類法を述べ、2値パターンの扱いにおける問題点などを指摘した。2クラス問題におけるシミュレーション実験を行い、テスト特徴分類法の特性を確認した。

第3章において、多クラス問題に対応するようにテスト特徴分類法の拡張を行った。第2章と同様のシミュレーション実験を行い、多クラス問題においても上で述べた特性を確認した。更に、その特性を活かした問題として、テキストチャ分類における部分クラス問題を取り上げた。多クラス問題における実画像を用いたテキストチャ分類実験では、コントラ

ストの違いによる部分クラスと、種類の違いによる部分クラスを設計した。代表的なパラメトリックな識別法との比較実験を通して、テスト特徴分類法は部分クラス問題に対して有効であることを明らかにした。

第4章において、多値パターンをも扱える多値型テスト特徴分類法への拡張を行った。ここでは、従来の2値型テスト特徴を判定するための同一性の評価に代えて、距離による類似性の評価を導入した。拡張した多値型テスト特徴の定量的な評価を行うために組合せ特徴における分離度を定義した。分類困難な問題として知られている人工データ及びよく利用される2つの実認識問題データベースを用いて、テスト特徴抽出の安定性、分離度によるテスト特徴の有効性、投票スコアの特性などに関する基礎的な実験を行い、多値型テスト特徴分類法の有効性を明らかにした。

第5章において、特徴選択法としての多値型テスト特徴の応用を行った。すなわち、 k -近傍分類法の識別性能を向上させることを目的とし、テスト特徴及びテスト特徴分類法の基本構造を利用した k -近傍分類器の組合せ識別法を提案した。まず、多値型テスト特徴を k -近傍分類器へ導入することの有効性を調べるための実験を行い、テスト特徴が k -近傍分類器の識別性能の向上に寄与していることを確認した。次に、第4章で述べた実験データを使用し、従来の k -近傍分類法や代表的な組合せ識別法との比較実験を行い、次の結果が得られた。(1) テスト特徴の導入は k -近傍分類器の識別性能を向上させるだけでなく、重要でかつ選定の困難な k の値による認識率のばらつきを小さく保持し安定させる効果をもたらした。(2) 提案手法の組合せ識別では、複数の k -近傍分類法を直列に実行しても、距離計算を工夫することにより、未知パターンの認識時間が単一の k -近傍分類法に比べて大きく増大しないようにすることができた。更に、実際的な応用として、独自のテクスチャ画像の分類問題を取り上げた。対象は、地質調査で用いられる岩石画像である。学習及び認識において、同じ画像を複数の小画像に分割して得られた合計400個の小画像を用いた場合、376個の未学習の小画像のうち、372個の小画像(98.9%)が正認識された。学習のものと違った未知画像に対しても97.0%の高い正認識率を得ることができた。

第6章において、テスト特徴に基づく分類法における逐次学習に関する特性を明らかにした。提案手法の逐次学習では、一般によく行われる訓練パターン単位による逐次学習の他に、クラス単位による逐次学習も可能である。これは認識システムに新しいクラスを定義・追加しなければならない場合に非常に有利な特性である。また、その逐次学習におけるテスト特徴数の変化の特性を利用して、過剰な学習の制御や例外的なパターンの予測などができる。これに関して、実データによる逐次学習に関する実験を行って明らかにした。

これを要するに、著者は、汎用性の高い多クラス・多値型テスト特徴分類法を新たに提示し、それに基づいて、 k -近傍分類法の識別性能を顕著に向上させるための手法を提示し、シミュレーション及び実データを用いた実験よりその有効性を実証したものであり、画像認識工学、画像情報処理技術の分野に対して貢献するところ大なるものがある。よって著者は、北海道大学博士(工学)の学位を授与される資格あるものと認める。