

学位論文題名

局所重複する訓練データを用いるテスト特徴法の効率的学習 (Using Locally Overlapping Training Data for Efficient Learning in Test Feature Classifier)

学位論文内容の要旨

従来、官能検査であるがゆえに人間の目視検査に依存していた品質管理も、近年ではセンサや画像処理、コンピュータビジョンなどの技術を組み合わせた品質検査装置により自動化される例が多い。例えば、k-NN や SVM,boosting などのパタン分類器による分類の精度が向上していることもこの要因の一つである。パタン分類アルゴリズムを用いた品質検査分類では、観測画像を検査員が査定して正解ラベル（教師信号）を付与し、訓練データとして登録する必要がある、これにより教師あり学習が可能となる。さらに学習型パタン分類器は現在数多く利用されているが、高い学習コストが実用における問題の一つとなっている。この問題を低減するためには、訓練データ数の削減、学習計算アルゴリズムの効率化などのアプローチがある。また上記のような実産業分野における品質検査分類などにおいては、教師（専門家）が生産ラインから日々得られる検査データをラベル付する際、特に画像計測する特徴量においては、しきい値などの最適なパラメタ設定において試行演算回数が多いため高いコストを必要とすることが問題となる。

本論文では、従来よりノンパラメトリックなパタン識別手法として提案されてきたテスト特徴法 (Test Feature Classifier 以下 TFC) においてこれら学習コストや教師ラベル付コストなどの問題を解決するために学習高速化手法、能動学習を適応させた投票に基づくデータ選択手法、そして特徴量が追加された場合に効率的に学習させる手法の提案を行う。

第 1 章では、研究背景と目的を示し、本研究で対象とする分類問題について述べるとともに分類問題における課題を示し、本研究の目的を明確にする。

第 2 章では、本研究の開発の基本となる分類器であるテスト特徴法 (TFC) とその拡張手法である逐次テスト特徴法 (successive Test Feature Classifier 以下 sTFC) の基本アルゴリズムについて述べる。

第 3 章では、第 2 章で述べたテスト特徴法における学習過程でそれぞれの部分特徴空間上で存在するデータを独立したものとしてとらえていたのに対して、部分特徴空間ごとの関連性を見出し、データにおいても別空間での重複状況により任意の部分特徴空間において重複を判断し、計算コストを削減するアルゴリズムを提案する。簡単な数値例を用いてアルゴリズムを示し、さらに理解を深めるために空間上でのデータの推移を追いながらアルゴリズムの説明を行う。高速化アルゴリズムの学習計算時間の比較実験により、本手法の有効性を確認する。また英字文字認識用データセット、貝雄雌判別用データセットを用いて他のデータによる効果の違いについて実験的に検証する。

第 4 章では、sTFC の拡張として次元追加型 TFC について述べる。第 2 章で述べた TFC は、特徴量の一部を追加変更した場合に TFC を再構成する必要がある。この問題を解決するために本論文では、特徴量追加に伴う部分特徴空間の変化を解析し、再探索すべき特徴空間を示す。次元追加型 TFC の比較実験により本手法の有効性を確認する。

第 5 章では、sTFC において教師ありから教師なし学習へと切り替える自律的学習について検討す

る。パターン認識における学習型分類器において教師データを作成するためのラベル付作業はコストが高く問題視されている。第4章で一つの解決策として能動学習を取り入れることを提案。本章では、もう一つアプローチとして教師なし学習に注目する。教師なし学習では、分類クラスを識別器によって自動で作成するため機械的な分類クラスが生成される。そこで教師あり学習において識別クラスを設定し、十分な学習を経て教師なし学習へと切り替える。本章では、分類器の成長度合いを表す指標として成熟度を定義し、また切り替えにより自律的学習の可能性を示す自律度を定義する。教師あり学習から教師なし学習へ切り替えるタイミングの検討を行うため成熟度を推測するための手法として PTF 特徴空間の学習による変化量を示し、変化量と成熟度の関連性について実データを用いた実験にて検証する。

第6章では、sTFC において能動学習の枠組みを組み込み、データ選択する手法について述べる。sTFC では、データ選択型追加戦略として正識別追加、誤識別修正追加が提案されているが、教師ラベル付きのデータ選択手法となり、ラベル付コストの削減を実現できない。本研究では、逐次学習の追加戦略として能動学習を取り入れることにより学習に必要なデータを選択する手法を提案する。一般的に学習に必要なデータは、識別問題の決定境界付近に存在するデータといわれており、TFC においては識別決定境界を線型的に決定することはできない。そこで、TFC の識別過程における投票を用いて識別決定境界付近にあるデータを推定する手法を提案する。ROC 曲線を識別器において描くことにより選択データの妥当性を示す。データ選択手法の sTFC の性能比較実験により本手法の有効性を示す。

第7章は本論文の結論を述べる。

学位論文審査の要旨

主査	教授	金子俊一
副査	教授	小野里雅彦
副査	教授	工藤峰一
副査	准教授	田中孝之

学位論文題名

局所重複する訓練データを用いるテスト特徴法の効率的学習 (Using Locally Overlapping Training Data for Efficient Learning in Test Feature Classifier)

従来、官能検査であるがゆえに人間の目視検査に依存していた品質管理も、近年ではセンサや画像処理、コンピュータビジョンなどの技術を組み合わせた品質検査装置により自動化される例が多い。例えば、k-NN や SVM,boosting などのパタン分類器による分類の精度が向上していることもこの要因の一つである。パタン分類アルゴリズムを用いた品質検査分類では、観測画像を検査員が査定して正解ラベル(教師信号)を付与し、訓練データとして登録する必要がある、これにより教師あり学習が可能となる。さらに学習型パタン分類器は現在数多く利用されているが、高い学習コストが実用における問題の一つとなっている。この問題を低減するためには、訓練データ数の削減、学習計算アルゴリズムの効率化などのアプローチがある。また上記のような実産業分野における品質検査分類などにおいては、教師(専門家)が生産ラインから日々得られる検査データをラベル付する際、特に画像計測する特徴量においては、しきい値などの最適なパラメタ設定において試行演算回数が多いため高いコストを必要とすることが問題となる。

本論文では、従来よりノンパラメトリックなパタン識別手法として提案されてきたテスト特徴法(Test Feature Classifier 以下 TFC)においてこれら学習コストや教師ラベル付コストなどの問題を解決するために学習高速化手法、能動学習を適応させた投票に基づくデータ選択手法、そして特徴量が追加された場合に効率的に学習させる手法の提案を行う。

第1章では、研究背景と目的を示し、本研究で対象とする分類問題について述べるとともに分類問題における課題を示し、本研究の目的を明確にする。

第2章では、本研究の開発の基本となる分類器であるテスト特徴法(TFC)とその拡張手法である逐次テスト特徴法(successive Test Feature Classifier 以下 sTFC)の基本アルゴリズムについて述べる。

第3章では、第2章で述べたテスト特徴法における学習過程でそれぞれの部分特徴空間上で存在するデータを独立したものとしてとらえていたのに対して、部分特徴空間ごとの関連性を見出し、データにおいても別空間での重複状況により任意の部分特徴空間において重複を判断し、計算コストを削減するアルゴリズムを提案する。簡単な数値例を用いてアルゴリズムを示し、さらに理解を深めるために空間上でのデータの推移を追いながらアルゴリズムの説明を行う。高速化アルゴリズムの学習計算時間の比較実験により、本手法の有効性を確認する。また英字文字認識用データセット、貝雄雌判別用データセットを用いて他のデータによる効果の違いについて実験的に検証する。

第4章では,sTFCの拡張として次元追加型TFCについて述べる.第2章で述べたTFCは,特徴量の一部を追加変更した場合にTFCを再構成する必要がある.この問題を解決するために本論文では,特徴量追加に伴う部分特徴空間の変化を解析し,再探索すべき特徴空間を示す.次元追加型TFCの比較実験により本手法の有効性を確認する.

第5章では,sTFCにおいて能動学習の枠組みを組み込み,データ選択する手法について述べる.sTFCでは,データ選択型追加戦略として正識別追加,誤識別修正追加が提案されているが,教師ラベル付きのデータ選択手法となり,ラベル付コストの削減を実現できない.本研究では,逐次学習の追加戦略として能動学習を取り入れることにより学習に必要なデータを選択する手法を提案する.一般的に学習に必要なデータは,識別問題の決定境界付近に存在するデータといわれており,TFCにおいては識別決定境界を線型的に決定することはできない.そこで,TFCの識別過程における投票を用いて識別決定境界付近にあるデータを推定する手法を提案する.ROC曲線を識別器において描くことにより選択データの妥当性を示す.データ選択手法のsTFCの性能比較実験により本手法の有効性を示す.

第6章は本論文の結論を述べる.

これを要するに,申請者はこの学位請求論文において,分類問題における汎用手法であるsTFCの独自の改良版について提案し,その高速計算性能を明らかにしており,その実応用分野への貢献度は大きいと判断できる.また学術的・技術論的価値も少なくなく,よって著者は北海道大学博士(情報科学)の学位を授与される資格あるものと認める.