

テスト特徴法のための逐次パターン学習に関する研究

学位論文内容の要旨

パターン分類の技術はパターン解析の分野で最も重要なもののひとつであり、日々、より高速で、より頑強な分類法の開発が研究されている。パターン分類の実応用問題のひとつとして、半導体検査・評価システムがある。半導体製造ラインにおいては、製造ウェハに発生する微細欠陥の種別特定及び原因対策を迅速に行う必要がある。欠陥自動分類では、日々の製造プロセスで検出された欠陥を、分類器の性能を向上させるための新しいデータとして利用することができる。このように、常に新しい訓練データが与えられそれらを有効に利用することによりできるだけ高性能な分類器を実現したいような実応用分野、あるいは多くの訓練データの一括処理コストが大きく段階的に学習をすすめ十分な性能を確保したい場合などにおいては、逐次学習のアルゴリズムは必須のものである。

我々は以前より、訓練データの特徴から一意に決定される分類に有効な特徴量の組み合わせであるプライムテスト特徴（以下、PTF）を用いる分類器として、テスト特徴分類器（Test Feature Classifier 或いは TFC）を提案してきた。TFC は、既知データの誤分類が無いことが保障されている、比較的少ないデータ数で分類器を構成することが可能である、などの特長を持ち、数々の応用問題に適用されその有効性が確認されている。しかし、分類器の構成にかかる処理に、やや計算コストを要している。また、データの一部のみが変更された場合でも分類器を元から再構成しなければならない、という問題がある。

本論文では、ノンパラメトリックなパターン分類法として考案された TFC の逐次学習に関する特性について検討する。逐次的に与えられる未知パターン群を識別した後、その結果を訓練パターンとして分類器に追加して次のより高い性能の分類器を構成しようとする場合を考える。TFC 構成にかかる処理はデータからの PTF 抽出がそれに相当する。まず、データが逐次的に与えられる場合に毎回最初から計算しなおすのではなく、既に判定されている PTF を効果的に利用して計算コストを抑えることができるアルゴリズムを提案する。次に、実問題の逐次パターン学習における、増加する訓練データの効率的保存問題、追加すべき訓練データ候補が多数得られる場合の選択問題（どの訓練データをついかすればよいか）などを実データをを用いた実験によって検証する。また、PTF の特性に基づいた重み係数による重み付き投票処理を提案する。これにより、各 PTF の機能がより効果的に発揮され分類器の性能向上が期待できる。さらに提案した重み付き TFC を逐次学習アルゴリズムへ拡張する。実データによる実験結果により本手法の有効性を確認する。

第 1 章では、研究の背景と位置づけを述べる。ここでは本研究で対象とする分類問題について述べると共に、そこにおける逐次パターン学習の課題を示し、本研究の目的を明確にする。

第 2 章では、組み合わせ特徴量を定義し、未知パターンの識別法である TFC について述べる。この章では、同手法の定式化を示し、簡単な数値例を用いて TFC の学習過程、識別過程の説明をおこな

う. また, TFC の特徴をまとめ, TFC と同様に登録された訓練パターンと未知パターンの距離計算によって識別をおこなう分類法である k 近傍法との比較について述べる.

第 3 章では, 逐次学習型 TFC について述べる. 2 章で述べた TFC は, データの追加や削除によって訓練データの一部が変更された場合には分類器を元から再構成しなければならない. この問題を解決するために本論文では TFC に基づく逐次学習アルゴリズムを提案する. 逐次学習型 TFC のアルゴリズムを示し, 処理の流れの理解を助けるために数値例を挙げて説明をおこなう. 逐次学習の計算時間比較実験により, 本手法の有効性を確認する. また, 多くのパターン分類器の性能評価にもちいられてきた英文字認識セットをもちいて逐次学習型 TFC の性能について, 追加戦略による違い, 他の分類法との比較などを通して検討する. また, 実画像分類への応用として, 半導体ウェハ上の欠陥画像からの特徴量に基づいて, 逐次学習の効果を実験的に検証する. さらに, 所望される分類性能を実現するための必要な訓練データ数について, シミュレーションデータによる実験をおこない検討する.

第 4 章では, 重み付き投票処理の導入による重み付きテスト特徴法についての定式化を示す. 従来の TFC では, 投票の際に全てのテスト特徴から均一のスコアを投票していた. しかし, 実際にはテスト特徴上でのデータの分布はそれぞれ違っており, それによって各テスト特徴の識別性能も違ってくる. 本研究では代表的なアンサンブル学習法の一つである Boosting のアイデアを基に TFC に基づく重み付き投票処理を提案する. ただし, TFC において要素分類器であるテスト特徴はそれぞれについて訓練パターンの誤分類が無いことが保障されているため, Boosting のように誤分類尺度に基づく重み付き投票処理を適用することができない. そこで訓練パターンの分布から各テスト特徴の期待される識別性能を推定し, それを基に重み付き投票処理をおこなう. TFC に重み付き投票処理を導入することによってテスト特徴の機能をより効果的に発揮することができる. 重み付き TFC の有効性の確認と重み係数に関する考察を実データを用いた実験によっておこなう. さらに, 我々がこれまでに提案してきた逐次学習型 TFC を基に, 重み付き TFC を逐次学習処理が可能なアルゴリズムへ拡張する. 実システムを設計する上で, 提案したアルゴリズムが有効であることを実験により示す.

第 5 章では, TFC のクロスバリデーションによる性能評価の高効率化について述べる. データの分割数を n としてクロスバリデーションによる評価をおこなうと, n 種類の訓練データから TFC を構成する必要がある. この操作を逐次学習型 TFC をもちいて高効率に計算する手順を示す. また, シミュレーションデータを用いた実験により, 分割数と効率化率の関係について検討する.

第 6 章は本論文の結論である.

学位論文審査の要旨

主 査 教 授 金 子 俊 一

副 査 教 授 山 下 裕

副 査 教 授 五十嵐 一

副 査 准教授 田 中 孝 之

学 位 論 文 題 名

テスト特徴法のための逐次パターン学習に関する研究

パターン分類技術はパターン解析分野で最も重要なものの一つであり、日々、より高速で頑強な分類法が研究されている。パターン分類の実応用問題の一つとして半導体検査・評価システムがある。半導体製造ラインでは製造ウェハに発生する微細欠陥の種別特定及び原因対策を迅速に行う必要がある。欠陥自動分類では、日々の製造プロセスで検出された欠陥を、分類器の性能を向上させるための新しいデータとして利用できる。このように、常に新しい訓練データが与えられそれらを有効に利用することによって高性能な分類器を実現したい実応用分野、あるいは多くの訓練データの一括処理コストが大きく段階的に学習をすすめ十分な性能を確保したい場合には、逐次学習のアルゴリズムは必須である。

本研究では、分類に有効な特徴量の組み合わせであるプライムテスト特徴 (PTF) を用いるノンパラメトリックな分類法として提案されているテスト特徴分類器 (TFC) の逐次学習に関する特性について検討している。実問題の逐次パターン学習における、増加する訓練データの効率的保存問題、追加すべき訓練データ候補が多数得られる場合の選択問題などを実データによって実験的に検証している。また、TFC に重み付き投票処理を導入することによって性能を向上させている。さらに提案した重み付き TFC を逐次学習アルゴリズムへ拡張している。実データによる実験結果により本手法の有効性を確認している。

第 1 章では、本研究で対象とする分類問題について述べると共に、そこにおける逐次パターン学習の課題を示し、本研究の目的を明確にしている。

第 2 章では、組み合わせ特徴量を定義し、TFC について述べている。この章では、同手法の定式化を示し、簡単な数値例を用いて TFC の学習過程、識別過程の説明を行っている。また、TFC の特徴をまとめ、 k 近傍法との比較についても述べている。

第 3 章では、逐次学習型 TFC について述べている。逐次学習型 TFC のアルゴリズムを示し、逐次学習の計算時間比較実験により提案手法の有効性を確認している。また、逐次学習型 TFC の性能について、追加戦略による違い、他の分類法との比較などを通して検討している。また、実画像分類への応用として、半導体ウェハ上の欠陥画像からの特徴量に基づいて、逐次学習の効果を実験的に検証している。さらに、所望される分類性能を実現するための必要な訓練データ数について、シミュレー

ションデータによる実験をおこない検討している。

第4章では、重み付き投票処理の導入による重み付きテスト特徴法についての定式化を示している。訓練パタンの分布から PTF の期待性能を推定し、それを基に重み付き投票処理をおこなう。重み付き TFC の有効性の確認と重み係数に関する考察を実データを用いた実験によって行っている。さらに、逐次学習型 TFC を基に、重み付き TFC を逐次学習処理が可能なアルゴリズムへ拡張している。実システムを設計する上で、提案したアルゴリズムが有効であることを実験により示している。

第5章では、TFC の交差検証法による性能評価の高効率化について述べている。分割数 n の交差検証法による評価をおこなう場合、 n 種類の訓練データから TFC を構成する必要がある。この操作を逐次学習型 TFC を用いて高効率に計算する手順を示している。また、シミュレーションデータによる実験で、分割数と効率化率の関係について検討している。

第6章は結論である。

第3章では、まず、独自のパタン分類法である TFC に基づく逐次学習アルゴリズムについて述べており、その有効性を実験により示している点を評価した。次に、TFC の逐次学習に関する特性を実験的に明らかにしており、実応用における逐次 TFC の可能性を示している。また、逐次 TFC を半導体欠陥画像分類へ適用し、実問題として十分な結果を得ていると考える。これらのことから、提案手法が実応用可能な技術であることが確認できる。

第4章では、各 PTF 上でのデータの分布から各 PTF の期待性能を推定し、それらを利用した PTF による重み付き投票処理をおこなうことによって TFC の性能を向上させることができることを実験的に示している。また、重み係数を効率的に計算する手順を示している。さらに提案手法の逐次学習への拡張を行っており、実験によりその有効性を検証している。ここで提案されている重み付き投票処理は、特に分類器が未成熟な学習初期の段階で特に有効であり、実システムを設計する上で有意な結果であることを確認した。

以上を要するに、著者は実産業を対象としたパタン認識分野において新しい技術を提案しており、検証実験により有用性を確認するものであり、情報科学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 (情報科学) の学位を授与される資格あるものと認める。