

## 欠陥分類のためのテスト特徴法における逐次学習効率化

北海道大学 ○松尾 祥和 北海道大学 高氏 秀則 北海道大学 金子 俊一  
要旨

本研究では、テスト特徴分類器 (Test Feature Classifier 以下, TFC) に基づく逐次学習において効率よくデータを追加するために、追加データを選別する方法を提案する。また教師あり学習から教師なし学習へ切り替える際の指標となる「自律度」を定義し、これと相関が高い分類器の成熟度の推測方法として、判定器群間の距離を利用する手法を提案する。

### 1. 序論

パタン認識の実応用問題の一つとして、製品製造ライン上の欠陥自動分類が挙げられる。欠陥自動分類では、日々の製造プロセスにおいて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。逐次的に得られるデータを利用し、高性能な分類器を実現する手法として逐次学習アルゴリズムが従来から数多く提案されている。統計的パラメタ推定・同定における逐次最小自乗法、確率近似法 [1], カルマンフィルタ [2], LVQ[3] などがある。

本研究の目的は、我々が従来より提案してきた TFC に基づく逐次学習アルゴリズムである sTFC [4] において、逐次的に得られるデータによる教師なし学習によって分類器を効率よく高性能化することである。TFC においては、専門家による査定により分類されるべきクラス情報が付加されたデータ (以下, 査定データ) を追加データとして必要とする。この査定コストを軽減するため、クラス情報が付加されていないデータ (以下, 未査定データ) による分類器の教師なし学習が望まれている。本研究では、sTFC における教師なし学習のための新しいデータ追加戦略として、PTF の投票率に基づくデータ選別を利用した部分追加戦略を提案する。また、教師あり学習から教師なし学習へ切り替える際の指標となる「自律度」を定義し、これと相関が高い分類器の成熟度の推測方法として、PTF 間距離を利用する手法を提案する。

### 2. 投票率に基づく部分追加戦略

学習時の訓練パタンの追加方法について述べる。文献 [4] には 4 つの追加戦略が提案されている。本論文では、新しく学習に有効であるデータを選別し追加する方法を提案する。この方法は、未査定データをデータ追加前に作成した TFC により識別し、識別過程において出てくる PTF の投票率に基づき学習に必要なデータ、つまり他クラス間の境界付近にあるデータのみを選別し追加する方

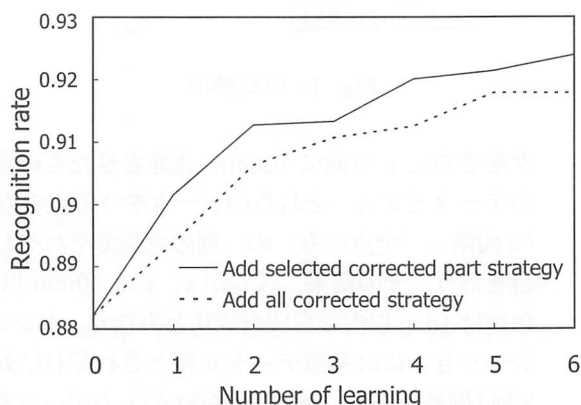


Fig. 1: 投票率 30 ~ 70% のデータを修正追加

法である。

### 3. 投票率に基づく部分追加実験

評価データを 1,100 個とし、残り 3,900 個のデータを 7 分割し逐次学習を行った。境界付近のデータが、分類に重要なデータであると考え、追加データのうちデータ追加前に作成した TFC で識別し、投票率が 30 ~ 70% のデータのみを修正後に追加する実験を行った。投票率が 30 ~ 70% のデータ (50% 近くのデータ) は、どちらのクラスにも同じ程度投票される事より、どちらかのクラスに判定しづらい、つまり境界付近のデータであると考えられる。

Fig.1 に全修正追加との比較結果を示す。Fig.1 より、投票率に基づき追加データを選別したほうが、全修正追加よりも識別率が高く、効率よく学習できるのがわかる。今回の実験では、投票率を、一つのクラスに 30% 以上 70% 以下の範囲に投票されたデータのみを用いたが、どのようなデータにおいてもこの範囲が使えるかどうかかわからない。よってさらに多くの実験を行ってみる必要がある。

## 4. 自律度と成熟度

本研究では、ある程度査定データによる学習を行い、その後未査定データを追加する手法を考える。未査定データによる学習結果は、データ追加前の分類器の状態や追加データに依存する。追加する未査定データを代えて実験を行い、識別率が上昇した回数によって自律度を定義する。全試行回数のうちの識別率が向上した割合を「自律度」とする。追加する未査定データに依存せずに常に識別率が上昇し始める点を自律度 100 とし、この状態を「分類器は自律した」とする。

パターン認識において、一般的に査定データを増やすにつれて識別率が上昇する。しかし査定データを追加しても、それ以上識別率の上昇を望めない状態がある。この状態に達したとき、この分類器は「成熟している」とする。「成熟している」点を基準として、いまどの程度成熟しているかを示す指標を「成熟度」とする。

## 5. TFC における成熟度の新たな指標

本研究において独自に定義した自律度を算出するためには、毎回データを追加するごとに教師なし学習を試みる必要があるため、計算コストがかかる。このため、他の方法により自律度を推定することが求められているが、これを実現することは困難である。そこで、まずは自律度と成熟度に相関があると考え、成熟度を推測する方法を考える。

TFC における逐次学習では、査定データが追加されるごとに PTF 抽出が行われる。分類器が成熟してくると識別率の変化が小さくなり、これに伴い分類器の更新による変化、つまり PTF の変化も小さくなり、そこから成熟度が推測できると考えた。本論文では、分類器の変化を PTF 集合の距離としてとらえることを提案する。PTF 集合の距離は、PTF を特徴空間上で表し、逐次学習においてそれぞれの学習回ごとに全 PTF の重心点を算出し、学習前と後の重心点の距離を測定する事により分類器の変化を表す。

## 6. 欠陥データによる実験

実際に金属製造メーカーの欠陥データを用いて実験を行った。クラスは 2。クラス特徴量は 20 で、評価データを 1,400 個。初期データを 360 個とし、毎回 360 個のデータを追加する。追加回数は 15 回とした。Fig.2 に、この実験を 10 回行った際の平均値を示す。

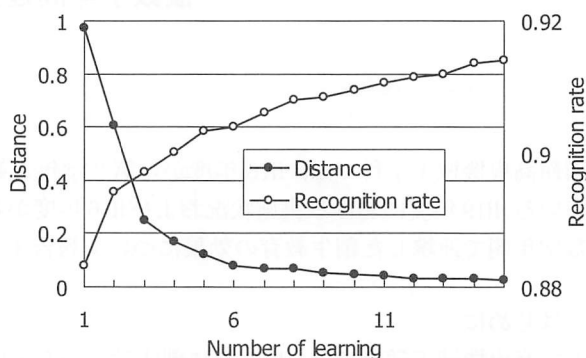


Fig. 2: 重心点間距離と識別率

Fig.2 より、識別率の上昇に伴って、PTF 集合の重心点間距離が減少している事がわかる。しかし、識別率は最後の学習においてもまだ上昇し続けるような兆候が見取れ、分類器が成熟しきっていない状態にあると考えられる。しかし、重心点間距離を見ると 0 近傍の値を示しているため、識別率とある程度の相関はあるものの、本データにおいて、成熟度を推測するには更なる工夫が必要である。

## 7. 結論

新しい追加戦略については、さらに実験を進めて実験例を増やし TFC の投票率と識別率の関係について考察を深める必要がある。成熟度については、今回の実験では PTF の距離を用いることを提案した。ある程度識別率との相関がとれているように見えるが、成熟度を推測するにはまだ工夫が必要である。

## 参考文献

- [1] A.P.Sage and J.L.Melsa : “System Identification”, Academic Press, New York, 1971.
- [2] G.C.Goodwin and R.L.Payne : “Dynamic System Identification”, Academic Press, New York, 1977.
- [3] T.Kohonen : “Self-Organizing Maps”, Springer, 1995.
- [4] 坂田幸辰, 金子俊一, 高木裕治, 奥田浩人 : “テスト特徴法に基づく逐次パターン学習と欠陥画像分類への応用”, 電学論誌 C, Vol.124, No.3, pp.689-698, 2004.