

特徴量選択型 SVM を用いた欠陥画像分類

A Support Vector Machine Approach to Defect Classification by Selecting Image Characteristics

余部 治昭†
Haruaki Amabe

長尾 智晴†
Tomoharu Nagao

1. はじめに

近年、半導体の微細化が進むにつれて製造プロセスでの欠陥管理が重要視されており、問題となる欠陥を検出、分類するための欠陥検出装置が必要とされ、様々な検査アルゴリズムが研究されている。しかし、現状のシステムでは欠陥を認識するための画像処理を行う際、専門家が問題ごとに手でプログラムを組む必要があり、これによって認識性能に大きく影響し、コストもかかってしまう。最近の研究では高い汎用能力を示す様々な学習法が提案されているが、それらを適用しても人間の識別能力には及ばない。

このような問題を解決するため筆者らは進化的画像処理を提案し、様々な問題に対して有効性を示してきた。進化的画像処理とは、生物の“進化”という考え方を工学における最適化に置き換え、このプロセスに“進化”の考え方を取り入れた進化的計算法を画像処理の問題に適用したものである。

そこで本研究では、この進化的計算法とパターン認識の学習法を組み合わせる手法を考案する。パターン認識の学習法として近年注目を集めているサポートベクターマシン (Support Vector Machine; SVM)[1]によって欠陥画像の分類器を構成し、進化的計算法として広範囲な問題に対して適用され、有効性が示されている遺伝的アルゴリズム (Genetic Algorithm; GA) を用いて SVM の入力となる画像の特徴量の中から有効なものを選択し、より高精度な分類を行う。

2. 特徴量選択型 SVM を用いた欠陥画像分類

2.1 サポートベクターマシン

一般に画像を分類する際、画像から特徴量を抽出し、識別器を学習することによって分類する手法が多く用いられている。識別器としてはニューラルネットワークが用いられることが多いが、出力結果が初期値に依存するため結果がランダムであることや、学習サンプルに対しては学習結果を保障するものの、未学習サンプルに対しては結果が保障されないという欠点がある。そこで本研究では識別器として、現在知られている様々なパターン認識手法の中でも最も優れた学習モデルの一つと考えられている SVM を用いた。

SVM はノンパラメトリックな手法であり、マージン最大化という基準を設けることで、識別境界の位置を決定する明確な基準が定義でき、未学習サンプルに対しても高い汎用能力がある。また、カーネルトリックと呼ばれる方法

を用いて非線形の識別関数を構成できるように拡張され、汎用能力が大きく向上している。

本稿では欠陥画像から抽出した特徴量を入力とし、SVM を用いて欠陥分類を行った。

2.2 GA による特徴量選択

一般にパターン認識で扱うデータは高次元であるものが多い。そのため、高次元ベクトルをそのまま識別器で学習した場合、

- ・ 計算コストがかかる。
- ・ 意味をもたない特徴量 (ノイズ) によって、最良の分類が行えない。

などの問題点がある。そのため、有効な特徴量だけを選択し次元数を削減する必要があると考えられる。しかし、ここで用いる特徴量によって分類器の精度に大きく影響を与えるが、人間が判断するのは大変難しい。また、分類する画像によって異なってくるため、分類器ごとに毎回人間の手に選択しては安定した精度も得られない。

そこで本研究では、進化的計算法として注目され、盛んに研究が行われている GA を用いて最適な特徴量を自動的に選択し、SVM によって分類器を構成する手法を用いた。このようにパターン認識の学習法と進化的計算法を組み合わせることで、画像処理の熟練者のような人手を必要とせず、より高精度で高速な分類器を獲得することができると考えられる。本手法の基本的な処理の流れを図 1 に示す。

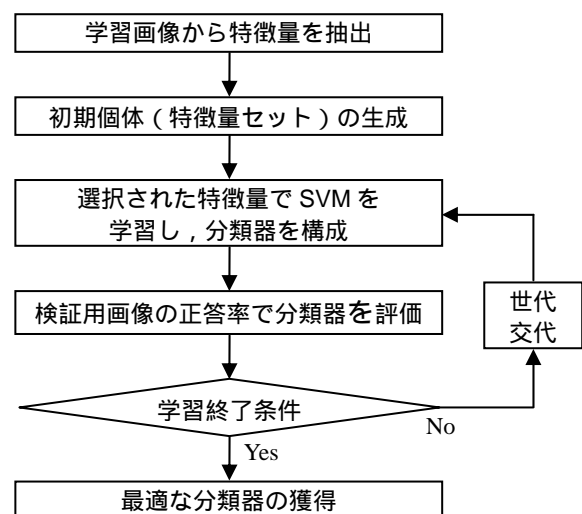


図 1 本手法の処理手順

上記のように、まずは画像から欠陥部分を切り出し、学習画像から特徴量を抽出する。次に SVM の入力である特徴量を GA によって選択し、最適な分類器を構成する。

3. 実験

3.1 実験条件

本実験では、ある製品の製造ラインにおける欠陥検査の際に撮影された欠陥画像を、指定の 6 種類に分類する実験を行った。その簡単な模式図を図 2 に示した。なお、これらの画像は、あらかじめ欠陥であると判断されたものだけを用いている。

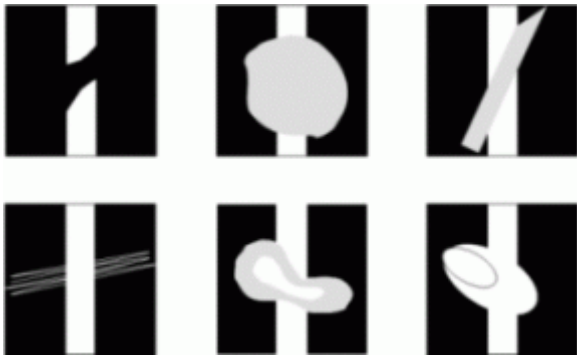


図 2 実験に用いた画像の模式図

上記のような画像 528 枚のうち学習画像を 96 枚、検証用画像を 218 枚、未知画像 214 枚として実験を行った。

GA の条件は個体数 50、世代交代モデルは MGG[3]、適応度の評価は個体ごとに SVM の分類器を構成し、未学習の検証用画像に対する正答率で行った。学習終了条件はエリート数が 5000 世代不変で終了とした。また、SVM の学習回数は 3000 回とした。使用した特徴量は平均階調値など、階調値を用いたものと、画像を 2 値化しラベリング処理をした後の面積や周囲長など、計 39 種を用意した。

3.2 実験結果

GA による特徴量選択を用いた SVM での欠陥画像分類の実験結果を表 1 に示す。なお、実験は 5 回行い、正答率の高いものから順に並べた。

表 1 欠陥画像分類の正答率

	特徴量選択なし (全 39 種)	特徴量選択	
		正答率	特徴量数
1 回目	84.9%	94.0%	16 種
2 回目	-	94.0%	16 種
3 回目	-	94.0%	16 種
4 回目	-	93.1%	16 種
5 回目	-	93.1%	17 種

ここで、特徴量選択で得られた 3 つの最良の分類器について、エリート個体の適応度の推移を図 3 に示す。

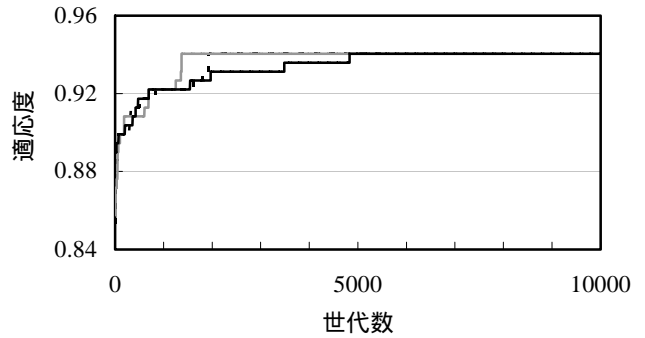


図 3 適応度の推移

以上の結果から、GA による特徴量選択を用いた SVM の分類器は、特徴量選択をしない分類器に比べて 9.1% 高い正答率を得られた。使用率の推移を見ても、進化の過程で常に選択される特徴量や全く選択されない特徴量なども多数見受けられることから、GA によって特徴量の取捨選択が進んでいるといえる。

また、未知画像に適応した結果正答率は 85.0% となり、検証用画像より低い数値ではあるが、特徴量選択前より正答率は上がっており、得られた分類器は未知画像に対しても有効であるといえる。以上のことから、本手法は有効であると考えられる。

4. おわりに

本研究では、GA による特徴量選択を用いた SVM での欠陥画像分類の実験を行い、その有効性を示した。使用する特徴量を選択することで、ノイズとなるような入力を除くことができたと考えられる。今回の実験では未知画像に適用した結果、汎用性もあることも確認することができたが、検証用画像と比べると正答率は低い値となった。このことから、過学習となった可能性がある。今後は、Boosting のような手法も実験し汎用能力を上げていく必要がある。また、本手法では一つの画像セットに対し一つの特徴量セットを選択し、各分類器の入力としたが、分類器ごとに異なる特徴量を選択することによって、さらに高精度な分類器の構築を目指していきたい。

参考文献

- [1] V.N.Vapnik: "Statistical Learning Theory", John Wiley & Sons, (1998)
- [2] 栗田多喜夫, 西田健次: "カーネル学習法とその画像認識への応用", 電子情報通信学会技術研究報告, Vol.104, No.291, pp17-24, (2004)
- [3] 佐藤浩, 小野功, 小林重信: "遺伝的アルゴリズムにおける世代交代モデルの提案と評価", 人工知能学会誌, Vol.12 No.5, pp.1-10, (1997)
- [4] 盧炳沃, 尹哲皓, 柳英基, 吳春錫: "ニューラルネットワークを用いた PCB の欠陥検出および分類に関する研究", 精密工学会誌, Vol.67, No.10, pp1621-1626, (2001)