

機関番号：24403

研究種目：基盤研究（C）

研究期間：2008 ～ 2010

課題番号：20500210

研究課題名（和文） ファジィクラスタリングに基づく高精度識別器

研究課題名（英文） High Performance Classifier Based on Fuzzy Clustering

研究代表者

市橋 秀友 (ICHIHASHI HIDETOMO)

大阪府立大学・工学研究科・教授

研究者番号：30151476

研究成果の概要（和文）：一般に高性能識別器は、調整できる自由パラメータを持っていて、識別器の汎化能力を高めている。監視カメラは防犯上の理由から多くの駐車場に設置されているが、カメラ方式の検知システムは検知精度の問題から屋外駐車場用としては実用化されていない。今回開発された新システムは平成21年に実際に導入された。また、屋外駐車場（屋上）における約2ヵ月間のテストでの検出精度は99.6%で当初の目標を大幅に上回る性能が得られた。訓練データが10倍になった時の訓練時間は、提案のFCMCでは10倍になるが、LibSVMでは約100倍になることを示した。

研究成果の概要（英文）：Generalization capability of classifiers is generally improved by introducing free parameters. The performance of the camera-based vehicle detection system is improved by using the FCM classifier. The new system was introduced to an underground parking lot in 2009. The system was also tested at an outdoor (rooftop) parking lot for a period of two months and achieved the detection rate (sensitivity) of 99.6%. The classification accuracy on test set, training time and testing time (i.e., detection time) of FCMC are compared with LibSVM by varying the number of training samples.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	1,600,000	480,000	2,080,000
2009年度	1,000,000	300,000	1,300,000
2010年度	1,000,000	300,000	1,300,000
総計	3,600,000	1,080,000	4,680,000

研究分野：知能情報学

科研費の分科・細目：情報学 ・ 感性情報学・ソフトコンピューティング

キーワード：クラスター分析，識別器，パターン認識，画像検索，パターン検出

## 1. 研究開始当初の背景

クラスター分析(クラスタリング)は、パターン識別の前処理として用いられるのが一般的であるが、我々はファジィクラスタリングに分散共分散構造を考慮することで高精度なパターン識別器が構成できることに着目した研究を進めていた。決定木やルールベース手法、また確立された高性能識別器である k-NN 法、さらに現在最も汎化性能が優れ

ているとされているサポートベクトルマシン (SVM) に比べても、提案の識別器はほとんどのベンチマークテストで優位な結果を得ている。確率統計分野の混合分布モデルとの対比からその位置づけを明らかにする研究も行っており、ファジィ理論の応用研究として数値的にその有効性を明らかにできる数少ない研究となることを目指した。

## 2. 研究の目的

クラスタリング研究をもとに、新たな識別器を提案し優れた汎化性能を示すことを明らかにする。そのパラメータ最適化に遺伝的アルゴリズム(GA)や粒子群最適化(PSO)などの進化的手法を取り入れ、識別性能の改善と性能評価を行う。その応用として実データでのSVMとの比較を広範囲に行い、有効性を明確にする。高次元データを扱える機能を生かして画像の説明文や文字データを用いない画像検索での有効性を明らかにする。画像のタグ付けへの応用や文書データの識別への応用も検討する。

## 3. 研究の方法

(1) FCM クラスタリング法を用いる識別器の三つの自由パラメータに加えて、さらにクラスター中心ベクトルの長さをパラメータとする場合とクラスターの混合比率をパラメータとする場合の比較を行った。

クラスタリングアルゴリズムは識別器の第1フェーズで用いられ、FCM識別器(FCMC)と呼ばれる。FCM 識別器は二つのフェーズから成り、第1フェーズではクラス毎にクラスタリングを行い、第2フェーズでは評価用データの識別及びメンバシップ関数の自由パラメータの最適化を行う。一般に高性能識別器は、調整できる自由パラメータを持っている。例えば、サポートベクターマシン(SVM)にはマージンやカーネルと呼ばれるパラメータがある。これらのパラメータが何らかの最適化手法で選択されることで、識別器の汎化能力を高めている。FCM 識別器には複数の自由パラメータがあり、パラメータと誤識別率の関係は単峰形の関数ではない。そこで、パラメータ探索の簡便な手法として粒子群最適化法(PSO)を適用した。

(2) さらにメンバシップ値を区間 $[a, b]$ に制約することでセミハードクラスタリングのアルゴリズムを提案する。そしてクラスタリング結果を識別器に用いてその性能をベンチマークデータで比較した。識別器には、K-L 情報量正規化 FCM 法や修正 FCM 法を考慮した。

(3) 監視カメラは防犯上の理由から多くの駐車場に設置されているが、駐車・空車を判別するために使用されているのは稀で、ほとんどが超音波や赤外線センサーを利用したシステムとなっている。カメラ方式のシステムは監視カメラとしての機能と車両管理のための機能を同時に持たせることができるが、検知精度の問題から屋外駐車場用としては実用化されていなかった。し

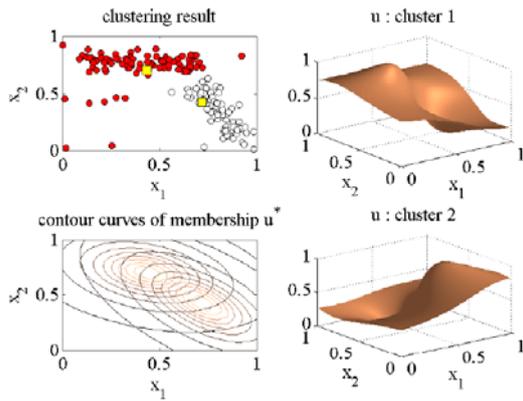
かし、今回実用化に成功した。

(4) セミハードクラスタリングに基づくファジィc-平均(FCM)識別器を大規模データに適用する際の課題や改善法について検討した。大規模データには、訓練データが大量である場合と特徴量(変数)の次元数が非常に大きい場合がある。データが大量である場合にはランダムなサンプリングでデータ数を削減することができるが、識別精度が悪くなることがある。特に利用可能な既知データに対する精度は低下する。提案のFCM 識別器は分散共分散行列を用いるために、データ件数が増えても行列のサイズは変わらない。そこで、多次元データを取り扱えるシステムを開発した。

(5) ファジィc平均識別器はクラスタリングを基にした識別器であり、訓練データが大量であればクラスター数を多くすることで訓練データに対する精度を向上できる。一方、テストデータに対する性能(汎化性能)はクラスター数を増やしても必ずしも改善されない。特に比較的少量の訓練データではクラスター数を多くするとオーバーフィットするだけでなく、各クラスターに含まれる訓練データが少なくなるために、共分散行列やクラスター中心が正確に求まらず性能が悪くなる。そこでクラスター数を8まで多くして、訓練データ数が変化することによるLibSVMとの性能比較を行った。

## 4. 研究成果

(1) ベンチマークデータを用いた幾通りかの分割による交差確認法(CV法)での比較から、再代入誤識別率(1-CV)を最小化する方法が有効であることを示した。また、自由パラメータにクラスターの混合比率、または中心ベクトルの変更割合を加えた場合の比較結果を報告した。提案FCM 識別器は、k 最近傍法(k-NN)よりも優れ、高性能な識別器として知られたSVMにほぼ等しい汎化性能を示した。また10-CV法の評価用データに対する識別精度はSVMよりも優れた結果が得られた。



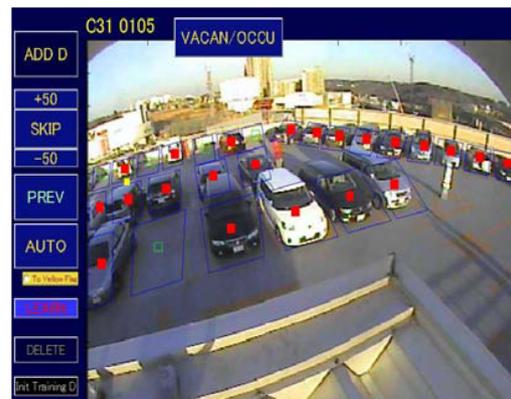
図：提案法でのクラスタリング結果

(2) 標準ファジィ c 平均法の目的関数に定数を加え，そこから導かれるメンバシップ関数にセミハードクラスタリングでの共分散行列を用いた．識別性能を最適化するためのパラメータ探索には粒子群最適化法 (PSO) を用いた．提案識別器の識別精度は SVM にほぼ等しく，k-NN に比べて優れた結果が得られた．

表：SVM との汎化性能の比較

	テスト用データの誤識別率				
	$r$	$\beta$	FCM 識別器 100 回	$t$	SVM 10 回
Iris	4	0.8	$3.04 \pm 2.47$	-	$3.4 \pm 3.4$
Breast	0	0.6	$2.99 \pm 1.04$	○	$3.6 \pm 1.0$
Ionosphere	3	0.9	$4.54 \pm 1.54$	-	$4.6 \pm 1.7$
Liver	5	0.9	$31.50 \pm 3.78$	-	$29.6 \pm 3.2$
Pima	3	0.8	$24.75 \pm 2.11$	♠	$22.7 \pm 2.2$
Sonar	8	1.0	$16.39 \pm 5.10$	⊙	$25.0 \pm 6.6$
Wine	13	0.9	$1.86 \pm 1.87$	-	$2.2 \pm 2.1$
Australia	2	1.0	$14.57 \pm 2.54$	-	$13.7 \pm 1.8$
German	0	0.9	$24.59 \pm 1.85$	-	$24.1 \pm 1.4$
Heart	0	0.6	$16.90 \pm 3.00$	-	$15.3 \pm 4.8$

(3) ファジィ c 平均クラスタリングに基づく識別器を応用してカメラ方式での車両検知システムの精度を改善した．セミハードクラスタリングにより計算時間の改善を図り，検知性能を改善するようにメンバシップ関数のパラメータを粒子群最適化法 (PSO) により最適化した．今回開発された新システムは平成 21 年 10 月に東京銀座の地下駐車場に第 1 号機として導入された．また，屋外駐車場 (屋上) における約 2 ヶ月間のテストでの検出精度は 99.6 % で当初の目標を大幅に上回る性能が得られた．



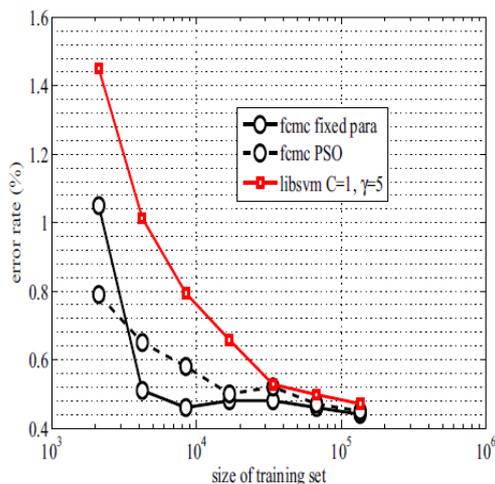
図：駐車場での車両検知システムの実用化

(4) 最も高性能な識別器の一つであるサポートベクターマシン (SVM) を用いた計算結果や改良 SVM の文献での報告との比較から，FCM 識別器は大量データの場合も SVM と同程度の識別精度を達成していて，訓練時間とテスト時間において大幅に優れた結果が得られた．

FCM 識別器は特徴量の次元数が比較的小さいか主成分分析 (PCA) などで圧縮して用いる場合には訓練データが大量でも短時間で訓練が収束する．しかし，特徴量の次元数が大きい場合には分散共分散行列のサイ

ズが大きくなり計算不可能となる。そこで、訓練データ数は比較的少ないが特徴量の次元数が非常に大きい場合に、特徴量の次元を圧縮することなく通常のパーソナルコンピュータでも計算可能な改良アルゴリズムを提案した。高次元データの例として COREL 画像データの分類問題を取り上げて、PCA によるデータ圧縮を用いる場合や文献で報告されている結果との比較を行った。

- (5) 大量訓練データを用いて訓練データ数が変化した場合の性能をクラス毎のクラスター数を 8 まで増やして比較した。訓練データが比較的少量の場合を考慮し、識別器はクラスター数を余り多くせず、逆に訓練データに対して最適化するパラメータ数を多くした場合の比較とする。世界的に実用可能なツールとして認められている LibSVM を用いて、訓練データ数が変化することによるテストデータの識別精度と訓練時間とテスト時間（検出時間）への影響を比較した。訓練データが 10 倍になった時の訓練時間は、FCMC では 10 倍になるが、LibSVM では約 100 倍になることを示した。



図：Parking データのテストデータでの誤識別率

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 19 件)

- ① 市橋秀友, 野津 亮, 本多克宏: セミハードクラスタリングとその識別器への応用, 日本知能情報ファジィ学会誌, 22, 3, 358-367 (2010). 査読有
- ② 市橋秀友, 堅田達也, 藤吉 誠, 野津 亮, 本多克宏: ファジィ c 平均識別器による駐車場のカメラ方式車両検知システム, 日本知能情報ファジィ学会誌, 22, 5, 599-608 (2010). 査読有

- ③ 市橋秀友, 長浦一哉, 野津 亮, 本多克宏: パラメータ化した FCM 識別器のベンチマークテスト, 日本知能情報ファジィ学会誌, 22, 5, 609-620 (2010). 査読有
- ④ 市橋秀友, 野津 亮, 本多克宏: 大規模データのファジィ c 平均識別器, 日本知能情報ファジィ学会誌, 22, 6, 792-803 (2010). 査読有
- ⑤ 市橋秀友, 本多克宏, 野津 亮: 多数のパラメータを用いるファジィ c 平均識別器の訓練データ数による性能比較, 日本知能情報ファジィ学会誌, 23, 2, 254-263 (2011). 査読有

[学会発表] (計 12 件)

- ① 市橋秀友, 野津 亮, 本多克宏: セミハードクラスタリングとその識別器への応用—大量データでの SVM との比較—, 第 26 回ファジィシステムシンポジウム, 2010 年 9 月 13 日, 広島大学 (広島県)
- ② 市橋秀友, 堅多達也, 藤吉 誠, 野津 亮, 本多克宏: ファジィ c 平均識別器による駐車場のカメラ方式車両検知システム, 第 26 回ファジィシステムシンポジウム, 2010 年 9 月 13 日, 広島大学 (広島県)
- ③ 市橋秀友, 野津 亮, 本多克宏: FCM 識別器のトリプレット, 第 25 回ファジィシステムシンポジウム, 2009 年 7 月 15 日, 筑波大学 (茨城県)
- ④ 市橋秀友, 本多克宏, 野津 亮, 太田圭一, 桑本奈歩: 粒子群最適化を用いるファジィ c 平均識別器, 第 24 回ファジィシステムシンポジウム, 2008 年 9 月 3 日, 阪南大学 (大阪府)

## 6. 研究組織

### (1) 研究代表者

市橋 秀友 (ICHIHASHI HIDETOMO)  
大阪府立大学・工学研究科・教授  
研究者番号: 30151476

### (2) 研究分担者

本多 克宏 (HONDA KATSUHIRO)  
大阪府立大学・工学研究科・准教授  
研究者番号: 80332964

野津 亮 (NOTSU AKIRA)  
大阪府立大学・工学研究科・助教  
研究者番号: 40405345