

Boost 学習に基づく特徴量の貢献度の評価

Feature Importance Evaluation based on Boosting

土屋 成光 †

Masamitsu Tsuchiya

藤吉 弘巨 †

Hironobu Fujiyoshi

1. はじめに

近年、画像認識分野では一般に、物体やシーンを認識するために、画像から特徴量を計算し、それを入力とした識別器を構築する方法が採られている。識別器には、非線形な識別境界が構築可能なものが用いられている。特に、識別時の高速性や、弱識別器の変更による問題への柔軟性から Boosting 手法 [1] を用いるものが増加している。Viola らは、空間と時間変化に着目し、Haar-like wavelets を用いた AdaBoost による人の検出法を提案している [2]。また、Hoiem らは Colour, Texture, Location, 3D Geometri に大別される特徴を入力として Boosted decision trees を構築し、画像中の三次元構造を復元する手法を提案している [3]。また、同様の特徴量を用いた車両、人認識が提案されている [4]。

人や自動車の認識等の問題では、同一対象における見えの違い、車種等の種別の多様性に影響を受けない特徴量を選択する必要から、入力特徴の選定は非常に重要な問題である。特徴量に対する評価方法としては、実際に識別器に特徴を追加、削除した際の識別率の比較、全サンプルの分散をより大きく持つ特徴軸を抽出する主成分分析 (PCA)、クラス間の分離をより容易にする特徴軸を抽出する判別分析 (MDA) などが知られている。しかし、実際に入力特徴を増減しての識別率の比較は、識別器構築の際に選択するサンプルの偏りに大きく影響を受けてしまう、特徴量の数だけ識別実験を行う必要があるなどの問題が存在する。また、PCA, MDA は全サンプルデータに対し評価を行う手法であるため、サンプルの偏りなどの問題は発生しないが、その際得られる評価値は、固有値問題により得られた新たに張られた特徴軸に対するものであり、元の特徴量に対する評価値とするためには変換を行う必要がある。

一方、Boosting による識別器は、そのアルゴリズムによって有用な特徴量を多く、不要な特徴量を少なく判断に利用するよう構築される。これは、Boosting による識別器自身による特徴評価である。そこで、我々は Boost 学習により実際に得られる弱識別器とその重みを用いて、簡易に非線形な識別境界の構築に際して有用な特徴評価を行うことを提案する。従来の特徴評価方法と異なり、本手法を用いた特徴評価の有用性、精度は構築する識別器の性能が自ずと保証しており、弱識別器を変更することで様々な問題に柔軟に対処できる。次に、得られた特徴評

価値と PCA, MDA との比較実験を行い、提案する貢献度との関係を示す。最後に、貢献度の利用法として、貢献度を用いた Boosting 識別器の調整を提案する。

2. Boost 学習による特徴量の貢献度

本論文では、Boost 学習を行う際の弱識別器を各特徴量と関連付け、学習の結果得られる重みを考慮することで、実際に識別に対してその特徴量が貢献した度合いを求める。弱識別器には 1 特徴量の閾値判別を用いる。

2.1. AdaBoost による学習

本研究では、各識別対象毎に特徴量の評価を行うため、単純に 2 クラス識別器 $H(x)$ を $H^{ClsA}, H^{ClsB}, H^{ClsC}, H^{ClsD} \dots$ のように各クラスに対して構築し、各クラス毎に学習過程で選択された弱仮説に着目する。まず、各クラスに対して、対象クラスかそうでないかを出力とする 2 クラス識別器 H^c を構築する。 c は対象クラスのラベルである。このとき、他クラスの positive サンプルを自クラスにおいて negative サンプルとして学習する。

AdaBoost の学習による弱識別器 h の生成は、まず、学習サンプルから両クラスのヒストグラムを特徴量 p 毎に求め、次に、それら全てに対して誤識別率を最小とする閾値 th を探索することで行う。学習回数 t 回目における弱識別器 $h_t^{p,th}(x)$ は、Fig.1 中 (C) のように $t-1$ 回目の学習結果に基づいて重み付けされたデータに対して、特徴量集合 P 全てについて閾値探索を行い、最良の識別率が得られる特徴量 p とその閾値 th を採用する (1 式)。

$$h_t^{p,th}(x) = \underset{(0 \leq th, p)}{\operatorname{argmin}} \left\{ \sum \delta_K(y_i, \operatorname{sgn}(x_i^p - th)) \right\} \quad (1)$$

AdaBoost では、識別関数 $h_t(x)$ が誤識別を起こしたデータを重視して再学習を行う。この処理をラウンド数 T 回反復した後、生成された識別器群の識別関数のアンサンプルによって最終的な識別関数を生成する。最終的な識別関数 $H_T(x)$ は次式で表される。

$$H_T(x) = \sum_{t=1}^T \alpha_t h_t(x). \quad (2)$$

ここで、 α_t とは t 番目の弱識別器による識別結果の信頼度を表し、最終的な識別に対して弱識別器 h_t の結果が影響する度合いである。以上のアルゴリズムを Fig.1 に示す。

2.2. AdaBoost による簡易なマルチクラス識別

マルチクラス識別は、2 クラス識別器を結合して実現する。まず、各クラスを対象として、ラベル y をデータ

† 中部大学

Algorithm The AdaBoost algorithm

1. Input
2. **Initialize:** $w_1(i) = 1/n (i = 1 \dots n)$, $h_0(\mathbf{x}) = 0$
3. **Do for** $t = 1, \dots, T$.
 - $\epsilon_t(h) = \sum_i^n I(y_i \neq h(\mathbf{x}_i))w_t(i)$
 - (a) $\epsilon_t(h_{(t)}) = \min \epsilon_t(h)$
 - (b) $\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t(h_{(t)})}{\epsilon_t(h_{(t)})} \right)$
 - (c) $w_{t+1}(i) = w_t(i) \exp(-\alpha_t h_{(t)}(\mathbf{x}_i) y_i)$
4. **Output:** Final hypothesis with weights α_t

$$\text{sign}(H_t(\mathbf{x})), \text{ where } H_T(\mathbf{x}) = \sum_{t=1}^T \alpha_t h_t(\mathbf{x})$$

図 1: AdaBoost アルゴリズム

がそのクラスに属する場合+1, 属しない場合-1 と設定し, 識別器 H^c を前述の AdaBoost 学習アルゴリズムにより構築する. これらの 2 値識別器に対するデータセットは全て共通である. ここで, 未知入力パターンがどのクラスに属しているかを識別するには, 各クラス毎に学習した 2 値識別器を統合する必要があるため, 各識別器における重み α を次式のようにその総和で正規化する.

$$\alpha'_t = \frac{\alpha_t}{\sum_{i=0}^T \alpha_i} \quad (3)$$

正規化した重み α' を用いて各クラス識別器の出力 $H^c(x)$ を求め, 最大のクラス c を識別結果とする.

$$\text{Class} = \underset{c}{\operatorname{argmax}} H^c(\mathbf{x}) = \sum_{t=1}^T \alpha'_t h_t(\mathbf{x}) \quad (4)$$

2.3. 特徴量の貢献度

弱仮説として選択された特徴量と, その弱仮説の正規化後の重みである α' は, その識別対象クラスの識別に重要な要素である. そこで, 学習後の AdaBoost において選択された各特徴量の弱仮説から, 貢献度 C_p を以下のように定義する.

$$C_p = \sum_{t=1}^T \alpha'_t \cdot \delta_K[P(h_t), p] \quad (5)$$

ここで, $P(h_t)$ は, h_t に採用された特徴量を求める関数である. この C_p を各クラス識別器において選択された弱仮説群から求めると, 特徴量 p がその識別器の識別能力にどの程度貢献するかを示す指標となる. このような特徴量の貢献度を予め求めることで, より多くの学習サンプルを用いて長時間の学習を必要とする高性能な他の識別器 (例えば ANN) を構築する際の特徴選択の指針とできる. この際, 汎化性能がより高いとされる Boost 学習 (LogitBoost, η -Boost 等) を用いることで, それぞれの特徴を反映した評価値を得ることが期待できる.

3. 貢献度の評価

3.1. 識別問題

本論文では, 実際に特徴評価を行う識別問題として, [5] の特徴セットとデータを用いた屋外環境下における

物体識別問題について行う. [5] で挙げているのは, 屋外環境下で撮影された映像における移動体を自動車 (VH) / 人 (SH) / 複数の人 (HG) / 自転車 (BK) の各クラスに識別する問題である. 本問題の識別対象は, 上記の 4 クラスとし, 識別器への入力特徴として, 形状に着目した特徴量 (2 種類), テクスチャに着目した特徴量 (4 種類), 時間情報に着目した特徴量 (1 種類) を用いる (Table.1).

表 1: 入力特徴セット [5]

特徴	ベクタ数
形状	3
縦横比と主軸の傾き (AS)	2
複雑度 (CS)	1
テクスチャ	8
垂直方向エッジ (V)	2
水平方向エッジ (H)	2
右上がり方向エッジ (R)	2
左上がり方向エッジ (L)	2
時間	1
オプティカルフローの分散 (OF)	1

学習には, CU データベース [7] より (VH:自動車 / SH:人 / HG:複数の人 / BK:自転車) について各 10 シーケンス, 計 40 シーケンスの動画像を用いる. 1 シーケンスから 40 フレーム分の特徴量を抽出し, 5 シーケンス分 200 フレームの画像を学習サンプルに用いる. 残りの 5 シーケンス分の 200 フレームは評価用のオープンデータに用いる.

3.2. 特徴量と弱識別器の重みの頑健性

本論文で提案した手法は, Boost 学習時の弱識別器の持つ重みを特徴毎に分類することで得られる. しかし, 逐次学習された弱識別器全てより得られた特徴貢献度が本当にその学習データに対して最適であるかを検討しなければならない. Jun-Su らは, AdaBoost 学習はあくまで逐次学習であり, 大域的最適解ではなく, 学習時の重みは最小限且つ最適なものではないと述べている [6]. Jun-Su らは, 進化的アルゴリズム (EA) により再評価した学習後の弱識別器の重みに基づいて, 弱識別器群から不要なものを選択, 削除することで識別性能を維持したまま識別を高速化できるとしている. そこで, EA により再評価した弱識別器の重みに対して貢献度を求め, 学習直後の重みによる貢献度と比較を行う.

EA による再評価実験 学習後の弱識別器とその重みに対して, EA を用いて再評価を行ったもの 3 世代 (1, 25, 50) の貢献度と学習直後の貢献度を比較する. 各世代の重みより求めた貢献度の相関係数を Table.2 に示す. 表より, EA により世代を重ねても, その貢献度は 0.89 程

度の相関性を持つことがわかる。これは、識別に対して大域的最適化を行うことで、弱識別器単位では重みが大きく変動するが、特徴量単位で見れば殆ど変化は無いことを示している。ここから、学習時に得られた弱識別器とその重みの組を貢献度を用いることに問題はないといえる。

表 2: 各世代に対する相関行列

	学習直後	1 世代	25 世代	50 世代
学習直後	1.00	0.94	0.93	0.90
1 世代	0.94	1.00	0.89	0.89
25 世代	0.93	0.89	1.00	0.92
50 世代	0.90	0.89	0.92	1.00

3.3. PCA, MDA との比較

PCA, MDA によって得られる特徴量評価と提案手法の値を比較する。PCA は、対象主成分を持つ特徴空間全体への寄与度として λ を算出する。この値を評価値として用いることを考える。しかし、PCA によって得られた寄与度は、元の特徴軸である説明変数集合 x に対してのものではなく、合成された新たな特徴軸となる主成分 x' に対する寄与度であるためシステムへの入力特徴に対する評価を直接示すことはできない。そのため、主成分がどれだけそれぞれの軸の成分を含んでいるかを因子負荷量より評価し、主成分の寄与度を元の特徴軸のスコアへ変換する。因子負荷量 $L_{i,j}$ は、第 i 特徴ベクトルと第 j 主成分ベクトルとの相関係数である。各特徴のスコア C_p^{pca} を以下のように定義する。ここで、 N は成分数である。

$$C_p^{pca} = \sum_{j=1}^N \lambda_j L_{p,j} \quad (6)$$

また、MDA の評価値は元の特徴軸の判別への重要度を表す偏 F 値を用いる。

比較実験 AdaBoost 学習による貢献度と PCA, MDA それぞれで得たスコアを比較する。[5] の問題に対する AdaBoost 学習による貢献度 (a), PCA で得られたスコア (b), MDA で得られたスコア (c) を Fig.2 に示す。図中の Correl. は、(b), (c) それぞれの (a) との相関係数である。

Fig.2 より、本手法、PCA, MDA による特徴スコアは全て形状特徴とテクスチャ特徴の有効性を示す結果となり、相関係数は共に高い。これらの結果は、一般的に知られている知識である、車両、人判別の際の形状、エッジ特徴の有効性と合致している。また、識別器より識別に用いている度合いが直接得られる本手法は、クラス間の分離を最大にする MDA とより相関性の高い値であることがわかる。これらのことから、本手法は識別のための特徴貢献度を評価する手法の一つとして用いることが可能であると考えられる。

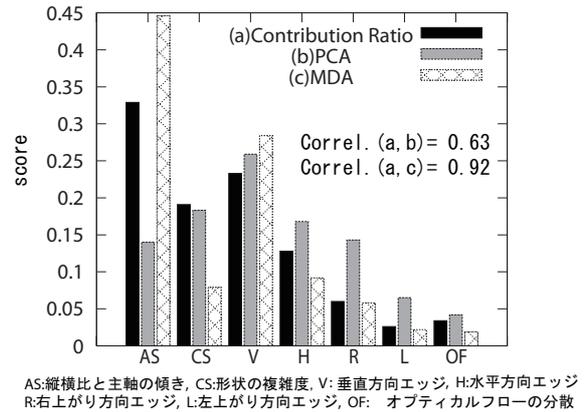


図 2: PCA, MDA との比較

4. 貢献度の利用

貢献度の利用法の一つとして、学習後に再学習することなくオープンデータへ調整により対応可能な識別器を提案する。

4.1. 学習後の識別器の調整

実際の物体識別において、学習サンプルデータと設置現場でのオープンデータの違いから識別性能が低下する問題がある。これは、学習サンプル取得時の環境とシステム納入時における環境の違いにより発生する。このような問題への対処には、学習サンプルデータにオープンデータを加え、識別器を再学習することが一般的である。しかし、システムを納入、設置する際の現場において、再学習による早急な対処は難しい。そこで、Fig.3 に示すような、貢献度を参考に、識別対象と各特徴量に対して識別感度を学習後に調整可能な識別器を考える。

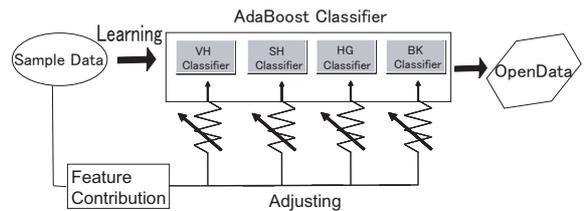


図 3: 特徴貢献度による調整可能な識別機

4.2. 学習後に調整可能な識別器

本論文の AdaBoost によるマルチクラス識別器は、1 特徴量の閾値判別の結果を重みつき多数決している。そこで、弱識別器の閾値を特徴量毎に調整することで、学習後の識別器をオープンデータ環境に対応させる。具体的には、(13) 式に示すように、AdaBoost 学習後に得られた弱識別器における閾値 th を Δ_{th} により調整することで、弱識別器の感度を変更する。

$$h_t(x) = \text{sgn} \{x - (th - \Delta_{th} \cdot \delta_K [P(h_t), p])\} \quad (7)$$

複数のクラス間で貢献度の高い特徴量を優先的に調整することで、オープンデータに対してより迅速な対応が可能となる。

4.3. 調整実験

貢献度を用いた調整実験を行う。[5] に対して構築した識別器はオープンデータに対して、人複数クラス (HG) の一部が人単体クラス (SH) に誤識別されることが多く、全体の識別性能を低下させていた。そこで、Fig.2(a) に着目し、SH, HG 間で最も貢献度が高い特徴である縦横比と傾き (AS) について HG クラスの識別感度を調整した。調整値 Δth を 0.0~1.0 まで変化させた際のオープンデータに対する識別率の変化を Fig.4 に示す。貢献度の高い AS を調整した場合を Fig.4.(a)、その他の特徴量を調整した場合を Fig.4.(b) に示す。

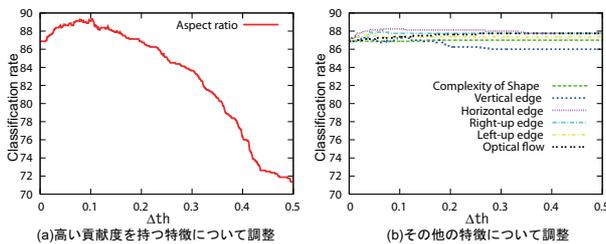


図 4: Δth vs 識別率

Fig.4.(a) より、貢献度の高い特徴量を調整することで識別性能を大きく向上させることができた。一方、4.(b) より、それ以外の貢献度の低い各特徴量について調整しても識別率に大きな変化は見られない。これらの結果より、貢献度の高い特徴量を優先的に調整することが、識別性能の向上に効果的であるといえる。また、本実験のケースでは、調整値 Δth の値を 0.1 付近に設定することで 2.5% 向上し、その後、識別性能は低下していくことがわかる。一般に、オープンデータに対する識別率は設置現場でしか得られないため、調整値の設定については、実際の設置現場で取得したデータに対して適応的に設定する必要がある。Table.3 に、貢献度に基づいた調整後 (a)、調整前 (b)、それぞれのオープンデータに対する識別性能を示す。何も調整しない (b) に対して、貢献度の高い特徴のみを調整した (a) は、HG クラスの識別性能を 15% 向上させつつ、他クラスへの悪影響が少ないため、総合的な識別率が向上していることがわかる。

表 3: オープンデータへの調整 [%]

	VH	SH	HG	BK	Total
(a) C_p に基づいた調整	100.0	83.0	88.0	88.0	89.1
(b) 無調整	100.0	86.0	73.0	88.5	86.7

5. おわりに

本論文では、Boost 学習より得られる弱識別器と重みから、新たな特徴量の貢献度を得る手法を提案した。また、主成分分析、判別分析により得られる統計量と比較することで、得られる貢献度が対象の判別に着目した値であることを示した。これにより、本手法の特徴評価は

統計量として意味を持つ値であるといえる。最後に、提案した貢献度の応用例として、学習後に調整可能な識別器への応用を行った。今後は、特徴間の共起性を表現した評価、複数の Boosting 手法間での比較を行う予定である。

文 献

- [1] Y. Freund and R.E. Schapire: "A decision-theoretic generalization of on-line learning and an application to boosting", Journal of Computer and System Sciences, Aug, 1, 55, 119-139, 1997.
- [2] P. Viola, M. J. Jones, D. Snow: "Detecting Pedestrians Using Patterns of Motion and Appearance", Proc. of the Ninth IEEE International Conference on Computer Vision (ICCV'03) - Volume 2, 2003.
- [3] D. Hoiem, A. A. Efros, M. Hebert: "Automatic photo pop-up". ACM Trans, Graph, 24(3): 577-584 (2005).
- [4] D. Hoiem, A. A. Efros, M. Hebert: "Putting Objects in Perspective", CVPR (2) 2006: 2137-2144.
- [5] M. Tsuchiya, H. Fujiyoshi: "Evaluating Feature Importance for Object Classification in Visual Surveillance", Proc. of ICPR 2006; 18th Int'l Conf. on Pattern Recognition (HongKong/China), pp. 978-981, 4 pages (Aug. 2006).
- [6] Jun-Su. Jang, Jong-Hwan. Kim: "Evolutionary Pruning for Fast and Robust Face Detection", Evolutionary Computation. CEC 2006. IEEE Congress on, p.1293-1299, 2006.
- [7] 小村剛史, 藤吉弘巨, 矢入郁子, 香山健太郎, 吉水宏: "歩行者 ITS のためのフレーム間差分による移動体検出法とその評価", 情報処理学会論文誌 CVIM, Vol.45, No.SIG13(CVIM10), pp.11-20, Dec. 2004.