

4M-7

次元モデルに基づく表情認識 – Adaboost を用いて

小塩 ひさよ[†] 小林 知博[†] 月安 聡[‡] 伊藤 昭[†] 寺田 和憲[†][†]岐阜大学工学部応用情報学科 [‡]岐阜大学院工学研究科

はじめに

表情の認識方法には主に2つの方法がある。基本的な表情6カテゴリ(喜び, 悲しみ, 恐怖, 嫌悪, 驚き, 怒り)に分類する方法と, ある次元空間上の点として表情を分類・認識する方法である。我々は後者に注目し, 基本カテゴリでは捉え難い微妙な表情に対する認識を目指している。ここでは表情を分類する次元の定義を快-不快[気分が良い/悪い], 覚醒-睡眠[活動的である/ない], 攻撃-防御[対象に向かう/対象から逃れる]の3次元として, この次元値を機械学習がどの程度認識できるかを調べた。

データベース

学習と試験に用いる画像データは, JAFFE データベース[1]の一部の画像(JAFFE データ)と, 独自に作成したデータ(DVD データ)である。JAFFE データは, 各カテゴリから同一人物でない10枚ずつ70枚を無作為に選出した。DVD データは, 基本6表情に加え様々な表情について検証するために今回新たに作成したもので, TVドラマや映画の映像から表情がうまく表出されていると思われる画像を, できるだけ様々な表情を含むようにして60枚切り出したものである。

各画像の次元値については, 情報学科学部学生に各画像についての次元値を5段階で評価してもらい, 一番多数の人が付けた評価値をその画像の次元値とし, これを学習データとした。

学習アルゴリズム

機械学習の手法には Haar フィルタによる LookUpTable を用いた Adaboost[2]で学習を行う。Adaboost とは, 単純な Classifier を組み合わせで精度のよい Classifier を構築する Boosting の手法の一つである。

【AdaBoost アルゴリズム】

サンプル空間: X ラベル空間: Y サンプル集合: $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$
 $x \in X, y \in Y$ Weak Classifier: $h: X \times Y \rightarrow [-1, 1]$ 教師信号: $Y[l] = \begin{cases} 1 & l \in Y \\ -1 & l \notin Y \end{cases}$ サンプルの重みの初期化: $D_1(i, l) = 1/(mk)$
 $i = 1 \dots m, l = 1 \dots k, k = |Y|$ For $t = 1 \rightarrow T$ r_t を最大にするような h_t を選ぶ

$$r_t = \sum_{i,l} D_t(i,l) Y_i(l) h_t(x_i, l)$$

ここから, $\alpha_t = \frac{1}{2} \ln \left(\frac{1+r_t}{1-r_t} \right)$

サンプルの重みの更新

$$D_{t+1}(i, l) = \frac{D_t(i, l) \exp(-\alpha_t Y_i[l] h_t(x_i, l))}{Z_t}$$

 Z_t : 正規化要素

得られる Strong Classifier :

$$H(x) = \arg \max_l \left(\sum_t \alpha_t h_t(x, l) \right)$$

【Weak Classifier】

ここで用いる WC は Haar フィルタの出力と分類クラスから成るサンプルの分布統計値である。Haar フィルタは矩形で等分された領域のピクセル値の差分を出力する。その出力範囲を等分し, 分類クラス毎の画像の出力分布をつくることで, 等分値 × クラス数のテーブル (LUT) を作る。Haar フィルタのサイズと位置の組み合わせの数だけ WC ができる[2]。

実験内容

【サンプル画像】

機械学習のために以下の前処理を行う。

- 1) 両目の端と口の下端から顔領域を切り出す。
- 2) 両目が水平になるように回転する。
- 3) 縦横比 1 : 1 (36 × 36pixel) に正規化する。
- 4) 2値画像に変換する。

Facial expression recognition based on a dimension model – using Adaboost algorithm

H. Ozio[†], T. Kobayashi[†], S. Tsukiyasu[‡], A. Ito[†], K. Tereda[†][†]Department of Information Science, Gifu University[‡]Graduate School of Engineering, Gifu University

さらに、サンプル数を増やすため、 $\pm 5\%$ の回転と $\pm 5\%$ の拡大・縮小を行う。2値化を行うにあたって、DVD データは光源が一定でないため、簡易な補正を加える。

【学習内容】

学習は次元軸毎に5段階（1～5）の次元値に対して行う。今回使用するサンプルの次元値の分布は偏りがあるため、JAFFE データと DVD データを合わせ、次元値毎に同数を用いて学習させる。以下、「快 不快」の次元軸に対する学習内容を記す。データの意味は次の通り。

- ・ **A 学習** : 両データからかく次元値同数のサンプルを無作為に選んだもの。
- ・ **B 試験** : 全データから A を除いたもの。
- ・ **C 試験** : B のうち JAFFE データ。
- ・ **D 試験** : B のうち DVD データ。

学習ステップは、学習データの認識率が9割に達した時点で終了し、作成された Classifier で、試験データの認識率を調べる。また、次元値を定めたアンケート結果にはばらつきがあるため、人の評価の一致率とも比較する。

結果と考察

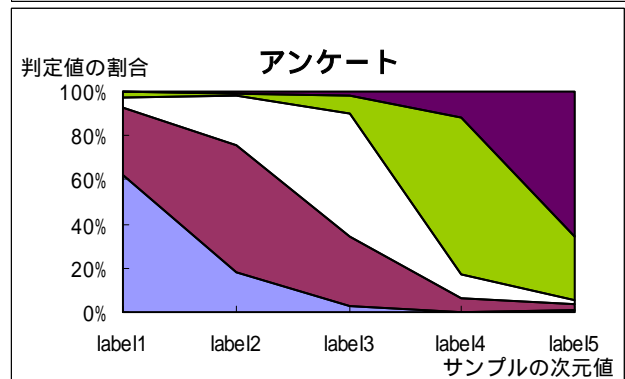
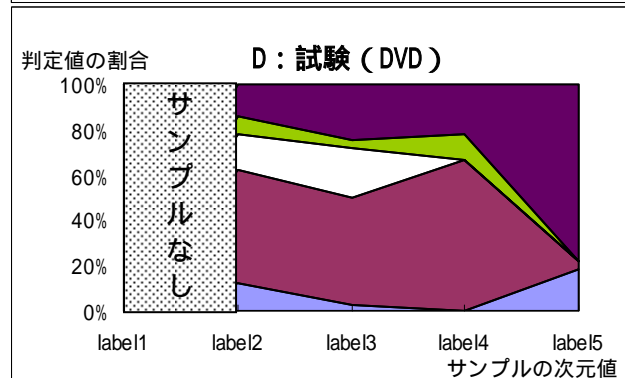
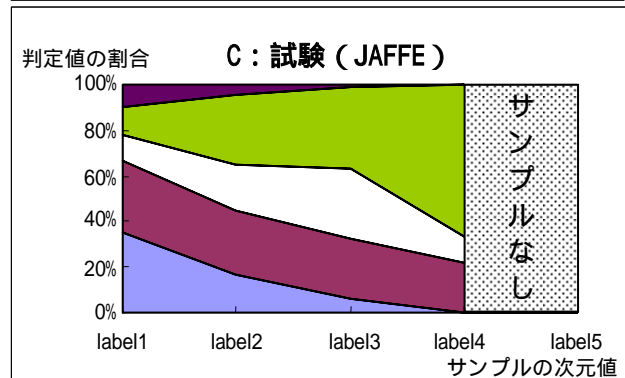
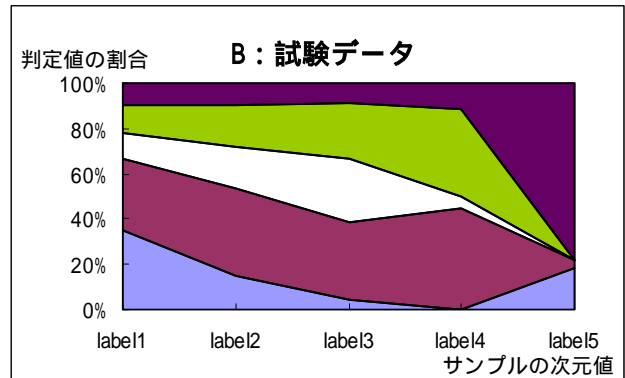
学習は 100 ステップで終了した。各データセットにおいて、ある次元値に属するサンプルに対する判定値の割合を図1に示す。どのグラフもおおむね斜めの形状をとっており、一定は次元値の学習が成功したと言える。D のグラフがとりわけ歪んでいるが、DVD データはメガネ付きや眉などに髪がかかった顔、正面でない顔も含むため、認識が困難であったと考えられる。また、一部サンプル数が極端に少ないために、1つのサンプルに対する影響が大きい。

まとめ

表情を表す次元値を機械学習でどの程度認識可能かを調べ、被験者の評価と一定程度一致する判別器を得ることができた。さらに妥当性を高めるために、より多くのデータや画像の補正法、また、それらに合わせた学習器の改良が必要である。

参考文献

[1]M. J. Lyons et al.: "Coding Facial Expressions with Gabor Wavelets", in IEEE Proc. Int'l Conf. on FG, Nara, Japan, 1998, pp.200-205.
 [2]Yubo WANG et al.:"Real Time Facial Expression Recognition with Adaboost", 17th Int'l Conf. on PR Vol.2 48 2-411, 2004.



ans1 ans2 ans3 ans4 ans5

図1 . 同次元値サンプル毎の判定値の割合