

表情识别算法研究进展与性能比较

崔凤焦

(北京科技大学 计算机与通信工程学院 北京 100083)

摘要:表情是人类情感在面部的表达方式,包含了诸多有用的人类情感和心理活动信息,而表情识别则是研究分析这些信息并进行正确分类的工作。目前,表情识别已成为互联网及相关行业的关注热点,在新兴的智能家居、情感机器人等方面具有较好的应用前景。为此,在分析已有研究成果的基础上,基于 Cohn-Kanade 库表情,选用卷积神经网络、支持向量机和 Adaboost 三类算法作为研究对象,通过算法结构设计和参数优化分别得到三类算法相对较优的算法结构,并根据识别过程和识别结果进行了三类算法的对比分析。实验结果及其分析表明,卷积神经网络对 Cohn-Kanade 表情库的识别效果最好,而 Adaboost 的处理时间最短,支持向量机的识别效果介于两者之间;表情识别算法的研究及其性能分析为表情识别的实际应用提供了有益的借鉴与参考。

关键词:人脸表情识别;卷积神经网络;支持向量机;Adaboost;人机交互

中图分类号: TP391.4

文献标识码: A

文章编号: 1673-629X(2018)02-0145-05

doi: 10.3969/j.issn.1673-629X.2018.02.031

Research and Performance Comparison of Facial Expression Recognition Algorithm

CUI Feng-jiao

(School of Computer and Communication Engineering, University of Science and Technology,
Beijing 100083, China)

Abstract: As a expression of human emotions in face, emotion contains a lot of useful information about human emotions and mental activity. Emotional recognition is to be analysis of the information and classify them correctly. Presently, emotion recognition has become a new focus of Internet and related industries, which has a good application prospect in emerging smart home, emotional robots and others. Therefore, on the basis of analyzing existing research, the convolutional neural network, support vector machine and Adaboost are taken as object based on Cohn-Kanade expression library. The optimal structures of each algorithm are obtained respectively through design of algorithm structure and parameter optimization and a comparison on them is made according to the recognition process and recognition result. Experiment shows that convolutional neural network has the best recognition result, while Adaboost has the minimal processing time and support vector machines somewhere in between. Research and performance analysis of facial expression recognition algorithm provides a reference for the application of facial expression recognition.

Key words: facial expression recognition; convolutional neural network; support vector machine; Adaboost; human-computer interaction

0 引言

1970年,研究者开始关注表情识别并进行了相关研究。GALTON于1888年和1920年在《Nature》上发表了两篇论文,主要是表情识别在身份识别方面的应用,这是最早关于表情识别应用的文章。1971年,Ekman P和Friesen W N研究了6种基本表情(高兴、悲伤、惊讶、恐惧、愤怒、厌恶),并采集上千幅不同人的表情组成了一个表情库^[1],这是一个系统的图像库。自表情识别的出现到现在已取得了一定的进展^[2-3],

但由于对表情的研究相对复杂,而且会受到光照、角度等外因的影响,发展相对比较缓慢。目前,市场上也缺乏较好的应用系统。此外,由于现有表情库都是在特定条件下采集的,使得研究具有一定的局限性。

在已有研究的基础上,采用卷积神经网络(convolutional neural network, CNN)、支持向量机(support vector machine, SVM)和Adaboost三种算法对Cohn-Kanade表情库进行了分类识别,并结合参数优化、算法结构设计、算法融合等对算法进行了对比分析,为下

收稿日期: 2016-07-28

修回日期: 2016-12-03

网络出版时间: 2017-10-19

基金项目: 国家自然科学基金重点项目(61432004)

作者简介: 崔凤焦(1990-),女,硕士研究生,研究方向为表情识别、机器学习。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20171019.1559.004.html>

一步的研究工作提供基础和参考。

1 算法综述

1.1 卷积神经网络

CNN 是对传统神经网络的改进,它引入了卷积和降采样 (pooling) 技术。1988 年, Lecun Y 等^[4]第一次成功训练了多层 CNN。2012 年, Hinton G E 等^[5]将 CNN 应用于 ImageNet 图像库, 相对之前的研究, 其错误率降低了 6% (前五个由 25% 的错误率降为 17%)。CNN 在手写字符识别^[6]、人脸识别^[7]、图像边缘检测^[8]、车牌字符识别^[9-10]、交通信号识别^[11]、身份证号码识别^[12]、人脸表情识别^[13] 等领域取得了很好的效果^[14-15]。

CNN 每层的输出特征图, 是上层特征图被具有学习性的卷积核卷积后通过激励函数得到的, 如式 (1) 所示:

$$X_j^l = f(\sum_{i=M_j} X_{ij}^{l-1} * k_{ij}^l + b_j^l) \quad (1)$$

f 表示特征图选择集合; X 表示卷积核; k 表示 CNN 的结构层数; b 表示 CNN 输出图各自的偏置量。

网络结构的子采样层中输出特征图的数目和输入特征图的数目一致, 但是输出特征图较小, 子采样层形如式 (2):

$$X_j^l = f(\beta_j^l \text{down}(X_j^{l-1}) + b_j^l) \quad (2)$$

其中, $\text{down}()$ 为子采样函数, 子采样实际上是对输入图像进行区域求和的过程, 区域大小为 $n * n$ 。图像经过采样处理会在两个维度缩小 n 倍, 所以输出图像比输入图像小。每个输出有一个特定的乘性偏置和加性偏置, b 为输出图像的加性偏置。

1.2 支持向量机

SVM^[16-17] 是一种模式识别方法, 由 VAPNIK 等在 1998 年提出, 它的实现目标是达到结构风险最小的同时为 2 种及以上的样本数据找到一个最优的分类面。SVM 和卷积神经网络都是学习性算法, 不同之处在于, SVM 采用数学上的方法实现, 优化技术也是数学实现。

SVM 算法做分类, 关键要选择合适的核函数。为了实现在核函数的广泛应用, 采用松弛系数 (gamma, g) 和惩罚系数 (c) 来校正。

SVM 算法原理如图 1 所示。

1.3 Adaboost

Adaboost 运用迭代思想, 训练某个训练集的多个弱分类器, 最终级联为一个强分类器。2001 年, Viola P 和 Jones M^[18] 使用 Adaboost 算法对人脸进行检测, 首次提出了“积分图”, 并建立了真正意义上的检测系统: Haar 分类器 = Haar-like 特征 + 积分图方法 + Ada-

boost + 级联。算法过程如下:

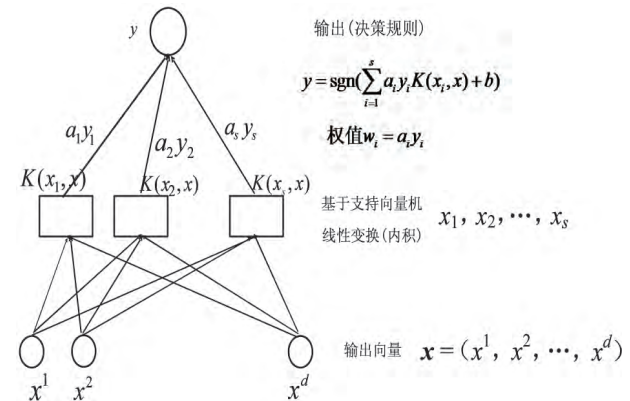


图 1 SVM 算法原理示意图

(1) 对要训练的样本集进行标定: $(x_1, y_1), (x_2, y_2), \dots, (x_L, y_L)$, $g_j(x_i)$ 为第 i 个样本的第 j 个 Haar-Like 特征, $x_i \in X$ 为训练样本, $y_i \in Y = (-1, 1)$ 对应真假样本, T 为循环的最大次数;

(2) 给权重设定默认值 $w_{i,j} = 1/2m, 1/2n$, 其中 m 为正样本数, n 为负样本数, 样本总数 $L = m + n$;

(3) 进行 T 轮训练, $\text{For } t = 1, 2, \dots, T$:

(a) 所有权重的样本进行归一化处理:

$$w_{i,j} = \frac{w_{i,j}}{\sum_{i=1}^L w_{i,j}} \quad (3)$$

(b) 由第 j 个 Haar-Like 特征得到初级分类器 h_j ,

阈值 θ_j 和偏置 p_j 可以使误差 $\varepsilon_i = \sum_{i=1}^L w_{i,j} |h_j(x_i) - y_i|$ 的值最小, 而

$$h_j(x) = \begin{cases} 1 & g_j(x) < p_j \theta_j \\ 2 & \text{其他} \end{cases} \quad (4)$$

式 (4) 可由 p_j 决定, 但只可选择 ± 1 两种情况。

(c) 满足最佳阈值条件下, 从已训练好的简单分类器中找到 ε_j 最小的 h_j ;

(d) 进行样本的权重更新: $w_{i+1} = w_i \beta_i^{1-\varepsilon_i}$, 其中 $\beta_i = \varepsilon_i / (1 - \varepsilon_i)$ 。如果 x_i 被 h_j 正确分类, 则 $\varepsilon_i = 0$, 反之 $\varepsilon_i = 1$ 。

(4) 强分类器为:

$$H(x) = \begin{cases} 1 & \sum_{i=1}^T \alpha_i h_i(x) > 0.5 \sum_{i=1}^T \alpha_i \\ 2 & \text{其他} \end{cases} \quad (5)$$

其中, $\alpha_i = \ln(1/\beta_i)$, 它的值与 h_i (第 t 轮的分类规则) 的预测错误相关, α_i 是对 h_i 的评价, h_i 的重要性与 α_i 的大小成正比。

上述算法的迭代过程共循环 T 次, 每次权重分布都不相同, 由此对正样本确定一个新的权重分布 P , 在 P 上得到一个新的弱分类器。所以, T 次循环共有 T 个弱分类器, 最后得到的强分类器是训练好的所有

弱分类器权值的平均。

2 实验结果与分析

选用的样本均来自 Cohn-Kanade 库,实验图片总数为 2 126 张,其中包括 anger(381 张)、disgust(302 张)、fear(280 张)、happy(312 张)、sadness(240 张)、surprise(380 张)、common(102 张)、contempt(129 张) 8 类表情。

从 Cohn-Kanade 表情库中取不同测试者的一部分图片作为样本集,其中 60% 作为训练集,40% 作为测试集;实验循环 15 次,并取它们的平均值作为最终的实验结果。

2.1 人脸检测

在进行表情识别之前,要先进行人脸区域检测和归一化处理,把非面部区域的部分去掉(例如头发),减小非人脸区域以及人脸角度偏离所造成的不必要的误差,以提高识别精度,裁剪图片统一为 64×64 。

提取人脸数据,然后进行特征提取,提取维度为 4 096。采用 LBP 进行全局特征提取,检测窗口为若干个 16×16 的小区域。对于每个区域中的某个像素点 i ,将其周围的 8 个像素点进行比较,值大于 i 的计为 1,小于 i 的计为 0,最后得到 8 位二进制数,即得到 i 的 LBP 值;计算所有小区域的直方图,即 LBP 值的概率,把 LBP 的统计直方图作为表情图的特征向量,并对直方图做归一化。最后,把所有区域的统计直方图合并为一个特征向量,即人脸的特征,提取的特征为 4 096 维,然后训练分类器进行表情分类。

2.2 基于卷积神经网络的表情识别

2.2.1 卷积神经网络结构设计

CNN 的网络结构主要由网络层、卷积层、子采样层和输出层组成。在进行结构设计时,网络层总数和每层的神经元个数都是考虑的主要因素。其中输出层有 8 个(因为共 8 种表情)神经元。

识别正确率随着 CNN 网络层数的增加而提高,但是层数越多 CNN 的结构就会越复杂,训练权值参数的效率就会降低。采用 7 层卷积神经网络结构,输入层为 64×64 的图像,卷积核的大小为 5×5 ,3 个卷积层 C_1, C_3, C_5 ,2 个子采样层 S_2, S_4 ,1 个全连接 F_6 和 1 个输出层。卷积神经网络结构如图 2 所示。

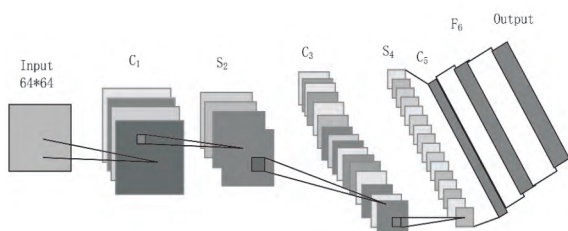


图2 卷积神经网络结构

2.2.2 识别结果

卷积神经网络神经元输出的特征图如图 3 所示。

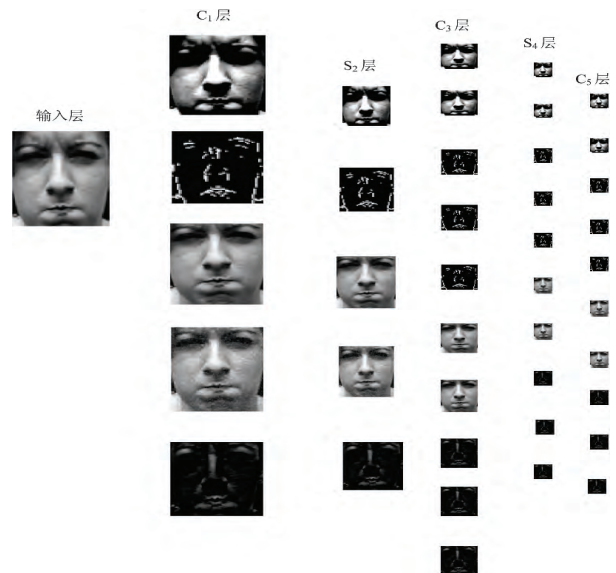


图3 卷积神经网络神经元输出特征
CNN 的识别结果如表 1 所示。

表1 CNN 识别结果

表情	训练样本数	测试样本数	识别率/%
anger	228	153	80.15
common	61	41	78.90
contempt	77	52	72.56
disgust	181	121	76.81
fear	168	112	70.64
happy	187	125	69.31
sadness	144	96	75.83
surprise	228	80	75.14

注: 平均识别率为 74.92%。

2.3 基于支持向量机的表情识别

2.3.1 基于 SVM 的表情识别实现方案设计

SVM 算法处理二类问题,而 8 种表情是多分类问题,所以要采取一定的方法使得 SVM 可以识别 8 种表情。采用以下 2 种分类方式使 SVM 实现 N 分类:一种是 $1vs(N-1)$;另一种是 $1vs1$ 。选择 $1vs1$ 的方法,训练 $N \times (N-1)/2$ 个分类器,判断某点属于分类器 i 还是属于分类器 j 。在 N 类问题中,有 N 个一对一的两类 SVM 分类器,任意两类之间都有 1 个分类超平面。故 8 类表情,共构造 28 个分类超平面。

2.3.2 支持向量机的参数设置

在 MATLAB 平台调用 Libsvm-3.20 工具箱,选择 SVM 的不同 c, g 参数进行识别调试,并比较识别结果,以寻找最优参数。其中, c 是惩罚系数,表示对误差的宽容度,值越高,越不能容忍误差的出现; g 是选择径向基函数作为核函数后,该函数自带的一个参数,

隐含地决定了数据映射到新的特征空间后的分布。参数选择结果对比, c 选为 4 左右、 g 为 $2^{(-8)}$ 左右时为最佳参数选择。

2.3.3 识别结果

SVM 的识别结果如表 2 所示。

表 2 SVM 识别结果

表情	训练样本数	测试样本数	识别率/%
anger	228	153	71.45
common	61	41	65.63
contempt	77	52	62.24
disgust	181	121	73.81
fear	168	112	70.14
happy	187	125	68.31
sadness	144	96	65.89
surprise	228	80	65.73

注: 平均识别率为 67.90%。

2.4 基于 Adaboost 的表情识别分类

2.4.1 Adaboost 分类器结构设计

采用 Adaboost 级联分类器进行面部表情识别时, 分别采用 3 级和 5 级级联, 并比较识别效果。实验过程从有两个特征的一个强分类器开始, 通过调整它的阈值达到最小错误率, 得到相对较好的人脸分类器。

2.4.2 识别结果

3 级和 5 级级联 Adaboost 分类器识别结果对比如表 3 所示。

表 3 3 级和 5 级 Adaboost 识别结果对比

表情	训练样本数	测试样本数	3 级识别率/%	5 级识别率/%
anger	228	153	63.27	65.18
common	61	41	64.75	67.32
contempt	77	52	68.43	68.70
disgust	181	121	70.71	71.01
fear	168	112	69.05	70.23
happy	187	125	72.31	72.89
sadness	144	96	69.79	71.25
surprise	228	80	73.82	73.72

注: 平均识别率分别为 69.02% 和 70.04%。

由表 3 可知, 5 级级联分类器的识别率较高。因此, 采用 5 级级联 Adaboost 分类器的识别结果作为最终的实验结果。

2.5 实验结果及算法性能分析

通过比较可知, CNN 的平均识别率最高, 即对 Cohn-Kanade 表情库的识别中, CNN 是三类算法中识别效果相对较优的算法。根据实验结果, 针对鲁棒性、参数设置和处理时间等对这三类算法进行性能比较。

鲁棒性: 即算法的稳定性, 在实验中 CNN 对图片畸变的承受能力是最强的, 主要因为 CNN 二次抽样的特点使其具有很好的鲁棒性。

参数设置: CNN 的参数设置相对复杂, 参数的初始值不能太小、权重参数以及设置梯度更新步长等都较复杂; SVM 主要是参数 c 、 g 和核函数的参数设置, 以及与对应的 Libsvm-3.20 工具箱参数的设置, 寻找最优 c 、 g , 比 CNN 的参数设置相对简单一些; Adaboost 基本不用调参。

处理时间: Adaboost 的处理速度是最快的, 研究表明 Adaboost 在训练样本充足时可适应特别复杂的分类面。

3 算法分析及其发展趋势

3.1 算法的优缺点分析

3.1.1 卷积神经网络

CNN 的算法结构比传统神经网络的要简单, 最大的特点是权值共享, 减少了计算时间, 而且 CNN 处理图像时可把原图直接输入网络, 特征提取也较简单, 因此会简化处理, 加快处理速度。目前广泛应用于人工智能、模式识别、人机交互和图像处理等方面, 多分类问题用 CNN 处理可得到更好的效果。

缺点: 实现相对复杂; 网络训练时间较长; 用 CNN 处理数据, 由于考虑其泛化性, 需要对处理的样本进行编号, 在实际的研究中, 很少有数据是有类标号的, 如果进行人工标号会大大降低工作效率。

3.1.2 支持向量机

SVM 是基于统计学的算法, 属于模式识别方法, 主要用于解决分类问题。SVM 是把数据从低维变换到高维, 并转换到线性问题来处理, 所以具有较高的效率和精度。

缺点: SVM 是小样本的学习算法, 在处理大量的样本数据时不能取得很好的效果; SVM 本身是解决二类问题的算法, 因此应用在多分类问题上有一定的局限性。因此, SVM 的改进方向主要针对多分类问题和大数据样本问题。

3.1.3 Adaboost

Adaboost 为迭代算法, 具有较高的分类精度。Adaboost 级联分类器是由几级弱分类器构成, 而构造弱分类的过程并不复杂, 所以 Adaboost 分类器的形成也不复杂, 构造极其简单; 此外, Adaboost 分类器中各级弱分类器都是独立的, 可以选择自己的方法进行结构组成, 由此可得 Adaboost 算法是为各级级联提供框架。Adaboost 的另一个优点是没有过拟合问题。Adaboost 同时也存在一些缺点: 噪声会影响其分类结果; 对样本的训练时间较长; 最终结果受制于弱分类器的

选择。

3.2 算法的改进方向

根据各个算法的上述缺点,有针对性地提出三种算法的改进方向。

对 CNN 结构方面的改进,是直接把图像输入网络进行处理,下一步研究工作把原始数据进行相应的预处理后再输入网络,提高识别精度,减少样本的训练时间;CNN 各层映射结果一般是独立的,若用 PCA 或其他算法对 CNN 各层的映射结果进行降维融合,并作为最后的特征提取结果,可以提高识别率。

由于 SVM 在多类问题方面的应用有一定的限制,需要找到一定的方法来构造解决多类问题的 SVM 分类器,例如通过 1vs1 的分类方法来构造多类的 SVM 分类器。在今后的研究中可以采用其他方法,例如与其他算法结合,获得较好的分类效果。

Adaboost 的改进方向主要考虑弱分类器的权值分配和算法间的结合,其中需要通过大量的实验掌握权值的选择方法,弱分类器的权值分配决定了 Adaboost 分类器的性能。

4 结束语

文中对 CNN、SVM 和 Adaboost 三种算法的算法结构进行了对比分析和改进设计,以分别实现对 Cohn-Kanade 表情库的识别,并根据实验结果对三种算法的优缺点进行了对比分析,对表情识别领域的研究与实际应用具有一定的参考价值。

参考文献:

- [1] EKMAN P, FRIESEN W V. Constants across cultures in the face and emotion[J]. Journal of Personality and Social Psychology, 1971, 17(2): 124-126.
- [2] MEHRABIAN A. Communication without words[J]. Psychology Today, 1968, 2(4): 53-56.
- [3] 詹永杰, 龙飞, 卜轶坤. 基于独立子空间分析特征学习的表情识别[J]. 系统仿真学报, 2015, 27(10): 2316-2319.
- [4] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural Computation, 1989, 1(4): 541-551.
- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]// Advances in neural information processing systems. [s.l.]: [s.n.], 2012: 1097-1105.
- [6] 高学, 王有旺. 基于 CNN 和随机弹性形变的相似手写汉字识别[J]. 华南理工大学学报: 自然科学版, 2014, 42(1): 72-76.
- [7] 刘洋, 季桂树, 王刚. 表情识别中基于轴对称的人眼定位方法[J]. 计算机技术与发展, 2008, 18(12): 47-49.
- [8] 叶波. 基于细胞神经网络的图像边缘检测研究[D]. 重庆: 重庆大学, 2012.
- [9] 王李冬. 一种新的人脸识别算法[J]. 计算机技术与发展, 2009, 19(5): 147-149.
- [10] 谢文浩, 翟素兰. 基于加权稀疏近邻表示的人脸识别[J]. 计算机技术与发展, 2016, 26(2): 22-25.
- [11] 陈先昌. 基于卷积神经网络的深度学习算法与应用研究[D]. 杭州: 浙江工商大学, 2013.
- [12] 郑永森. 基于卷积神经网络的身份证号码识别研究与实现[J]. 计算机光盘软件与应用, 2015, 18(3): 13-14.
- [13] 徐鹏, 薄华. 基于卷积神经网络的人脸表情识别[J]. 微型机与应用, 2015, 34(12): 45-47.
- [14] 孔锐, 张冰. 光照变化条件下人脸识别方法研究[J]. 系统仿真学报, 2016, 28(3): 689-695.
- [15] 白小叶, 程勇, 曹雪虹. 基于光照归一化分块自适应 LTP 特征的人脸识别[J]. 计算机技术与发展, 2016, 26(5): 56-60.
- [16] 祝曙光, 胡晓峰, 司光亚, 等. 仿真实验与 SVM 相结合的数据分类方法研究[J]. 系统仿真学报, 2010, 22(3): 761-764.
- [17] 黄永明, 章国宝, 董飞, 等. 基于 Gabor、Fisher 脸多特征提取及集成 SVM 的人脸表情识别[J]. 计算机应用研究, 2011, 28(4): 1536-1539.
- [18] VIOLA P, JONES M. Robust real-time face detection[J]. International Journal of Computer Vision, 2004, 57(2): 137-154.