

# Particle Swarm Optimization Based Algorithm for Conditional Probability Neural Network Learning

Junjie Xu, Min Jiang

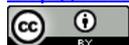
Fujian Key Lab of the Brain-Like Intelligent Systems, Xiamen University (XMU), Xiamen Fujian  
Email: xujunjie@stu.xmu.edu.cn, minjiang@xmu.edu.cn

Received: Mar. 12<sup>th</sup>, 2016; accepted: Mar. 25<sup>th</sup>, 2016; published: Mar. 30<sup>th</sup>, 2016

Copyright © 2016 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Conditional probability neural network (CPNN) has special advantage in pattern classification problems. However, how to find the optimal parameters of the CPNN to achieve better performance is an extraordinary challenge. Considering the structure feature of CPNN, we proposed a new training method based on particle swarm optimization (PSO). This method utilizes PSO to optimize the structure of CPNN and label distributions by introducing Hellinger distance between different label distributions. We applied the improved CPNN on facial age estimation. The experimental results showed that this network could increase recognition accuracy significantly.

## Keywords

Age Estimation, Label Distribution Learning, Conditional Probability Neural Network, Particle Swarm Optimization

---

## 基于粒子群优化的条件概率神经网络

徐骏捷, 江敏

厦门大学福建省仿脑智能系统重点实验室, 福建 厦门  
Email: xujunjie@stu.xmu.edu.cn, minjiang@xmu.edu.cn

收稿日期: 2016年3月12日; 录用日期: 2016年3月25日; 发布日期: 2016年3月30日

## 摘要

条件概率神经网络在进行模式分类时具有独特的优势, 然而如何对其进行有效的训练, 从而找到最优参数却是一个困难的问题。在考虑条件概率神经网络的结构特点之后, 本文提出了一种基于粒子群优化的条件概率神经网络的训练方法。我们将这种基于粒子群优化的条件概率神经网络用于人脸年龄估计, 实验结果表明这种网络能够显著地提高识别的准确率。

## 关键词

年龄估计, 标签分布学习, 条件概率神经网络, 粒子群优化

## 1. 引言

年龄估计因其在身份认证、人机接口、视频检索以及机器人视觉中存在着潜在的应用, 因而倍受关注。年龄估计有着广泛的应用前景, 主要包括: 1) 更人性化的 HCI: 通过自动识别用户的年龄, 可以为用户提供更合适其年龄的个性化的用户界面和服务。2) 图像及视频检索: 限定年龄范围, 从数据库中检索到需要的人脸图像或视频。3) 人口信息采集: 用自动年龄估计系统在公共场合收集有价值的人口信息。4) 安全门禁控制: 通过自动检测年龄, 实现安全网络或系统门禁控制。

自动年龄估计面临的主要难题就是年龄相对于其他面部因素的特殊性:

- 1) 老化的过程不可控。导致样本数据收集很费力
- 2) 独特的年龄格局。每一个人的基因, 健康情况, 生活方式都不同, 造成年龄格局不同。
- 3) 年龄格局是时间的函数。老化的过程必须按照时间顺序, 某时刻的脸部状态能够影响到后面的年龄的面部状态, 但是不能影响到年轻时刻的。

国际上, Young 和 Niels [1]是最早提出年龄估计的人。他们早在 1994 年就提出过通过人脸图像进行年龄估计。但他们的工作相对较为简单。他们把年龄粗略地分成: 小孩、年轻人和老年人三种。Hayashi 等人[2]研究了基于 Hough 变换的皱纹纹理和人脸图像肤色分析的年龄和性别识别方法。2003 年, Iga 等人用支持向量机开发了一个用于估计年龄的分类器。Lanitis 等人[3]提出一种基于脸部外观的统计模型。他们比较了不同分类器, 例如 kNN、MLP、SOM 的性能, 并且认为机器几乎可以和人一样估计出人的年龄。Nakano 等人[4]提出利用脸部和脖子上皱纹纹理的边缘信息来进行年龄估计。Zhou 等人[5]提出用 Boosing 的方法做为回归方法进行年龄的估计, 并用实验表明该方法比基于 SVMs 的方法还要好。Geng 等人[6]提出衰老模式子空间的方法, 通过学习一些代表性的子空间来建模衰老模式。

条件概率神经网络(Conditional Probability Neural Network, CPNN) [7], 是一种三层的神经网络模型, 它接收目标变量和条件特征向量作为输入, 输入两者的条件概率, 使用 KL 散度来作为损失函数, 并通过误差反向传播来更新网络的权重。CPNN 可以利用相邻的年龄信息协助学习, 有效地解决年龄估计问题。然而 CPNN 存在如下问题: 首先, 反向传播算法容易陷入局部最优。反向传播算法是一种局部寻优算法。它利用误差对于权值系数的一阶导数信息来获取下一步的权值调整方向, 以求达到误差最小。实际问题的求解空间往往是极其复杂的多维曲面, 存在很多局部极小点, 使得陷入局部极小的可能性大大增加。其次, 难以决定标签分布的参数。合适的标签分布可以使得网络在学习一个类的时候从周围获得良好的特征帮助学习, 而范围过大可能会对学习造成干扰。据此, 本文提出了使用粒子群优化算法帮助寻找条件概率神经网络的最优参数, 从而有效的解决人脸年龄估计问题。

文章结构组织如下。第二节先介绍了标签分布的概念, 然后将问题模型化, 并介绍了以 Hellinger distance 作为损失函数的 CPNN 和其参数学习的方法, 第三节介绍了运用在 CPNN 上的 PSO 算法来优化标签分布的参数和 CPNN 的权值。第四节是实验部分, 我们给出了实验结果, 并对其进行了分析。最后, 我们总结了文中所提及的研究工作, 并对未来将要进行的研究进行了展望。

## 2. 背景知识

### 2.1. 问题描述

首先我们给出标签分布的定义。

**定义 1. (描述度)** 给定一个实例  $x$ , 定义描述度  $d_x^y \in [0, 1]$ , 表示  $y$  描述  $x$  的程度。

对于一个实例  $x$ , 实际年龄为  $a$ , 那么  $a$  的领域范围内的描述度对  $a$  均有贡献。对于年龄  $a$  的标签分布, 应该满足如下 2 条假设:  $a$  的描述度在标签分布中是最大的, 以确保在类描述中实际年龄起到主导地位; 其他年龄的描述度随着离  $a$  越远而越小, 保证了相邻的年龄对实际年龄的贡献度比其他的更高一些。显然对于这两条假设, 高斯分布和三角分布都满足要求。

令  $X = \mathbf{R}^q$  代表  $q$  维输入空间,  $Y = \{y_1, y_2, \dots, y_c\}$  代表有限的标签集, 标签数量为  $c$ , 训练集  $S = \{(x_1, D_1), (x_2, D_2), \dots, (x_n, D_n)\}$ , 这里  $x_i \in X$  代表一个实例,  $D_i = \{d_{x_i}^{y_1}, d_{x_i}^{y_2}, \dots, d_{x_i}^{y_c}\}$  代表实例  $x_i$  所对应的描述度, 标签分布学习的目标为条件概率  $P(y|x)$ 。假设条件概率  $P(y|x)$  为含参模型  $P(y|x; \theta)$ , 这里  $\theta$  为模型的参数, 因此标签分布的学习目标为找出参数  $\theta$  使得给定一个  $x_i$ , 模型可以产生一个尽可能相似于  $D_i$  的标签分布。用 Hellinger distance [8] 来描述, 定义如下。

**定义 2. (Hellinger 距离)** 对于两个分布  $P = (p_1, p_2, \dots, p_n)$  和  $Q = (q_1, q_2, \dots, q_n)$ , Hellinger 距离

$$H(P, Q) = \sum_{i=1}^n (\sqrt{p_i} - \sqrt{q_i})^2$$

因此模型如下:

$$\theta^* = \arg \min_{\theta} \sum_i \sum_j \left( \sqrt{d_{x_i}^{y_j}} - \sqrt{P(y_j | x_i; \theta)} \right)^2 \quad (1)$$

### 2.2. CPNN 算法描述

条件神经网络由输入层, 一层隐层和输出层构成。CPNN 与传统 BP 神经网络存在不同: 输入变量中不仅包含实例  $x$ , 还包含实例的一个标签  $y$ ,  $x$  与  $y$  可以来自不同的领域, 拥有不同的取值范围。如在年龄估计问题中,  $x$  代表面部特征, 每个元素取值通常在  $[0, 1]$  之间,  $y$  代表面部所对应的年龄, 取值范围通常在  $[0, 100]$  之间。CPNN 的输出为条件概率  $P(y|x)$ 。

隐层输出节点的网络激活函数可以表示为

$$f(x; \theta) = \sum_{i=1}^{M_2} \theta_{31m} G \left( \sum_{k=0}^{M_1} \theta_{2mk} x_k \right) \quad (2)$$

这里  $G$  为 sigmoid 函数,  $M_l$  是第  $l$  层的节点个数,  $\theta_{lmk}$  代表在第  $l$  层的第  $m$  个节点与第  $l-1$  层的第  $k$  个节点之间的链接权重。

当输入同时包含  $x$  和  $y$  时, 神经网络输出变为:

$$p(x, y) = \exp(c(\theta) + f(x, y; \theta)) \quad (3)$$

这里

$$c(\theta) = -\ln \left( \sum_y \left( \int \exp(f(x, y; \theta)) dx \right) \right) \quad (4)$$

代表输出节点的偏重, 保证了  $\int p(x) dx = 1$ 。

因此条件概率计算如下:

$$p(y|x;\theta) = \frac{p(x,y;\theta)}{p(x;\theta)} = \frac{p(x,y;\theta)}{\sum_y p(x,y;\theta)} = \frac{\exp(f(x,y;\theta))}{\sum_y \exp(f(x,y;\theta))} \quad (5)$$

根据 Sarajedini 等人[9]的方法,  $p(y|x;\theta)$  可以被写作如下所示:

$$p(y|x;\theta) = \exp(b(x;\theta) + f(x,y;\theta)) \quad (6)$$

比较 Equation (6)和 Equation (5), 归一因子可以重写为:

$$b(x;\theta) = -\ln\left(\sum_y \exp(f(x,y;\theta))\right) \quad (7)$$

回顾 Equation (1), 模型的损失函数为:

$$T(\theta) = \sum_i \sum_j \left( \sqrt{d_{x_i}^{y_j}} - \sqrt{\exp(b(x_i;\theta) + f(x_i, y_j;\theta))} \right)^2 \quad (8)$$

种令  $\exp(b(x_i;\theta) + f(x_i, y_j;\theta)) = k_{i,j}$ , Equation (8)关于  $\theta$  的梯度为:

$$\frac{\partial T(\theta)}{\partial \theta} = \sum_i \sum_j \left( k_{i,j} - \sqrt{d_{x_i}^{y_j}} \sqrt{k_{i,j}} \right) \left( \frac{\partial b(x_i;\theta)}{\partial \theta} + \frac{\partial f(x_i, y_j;\theta)}{\partial \theta} \right) \quad (9)$$

最后, 使用 RPROP 算法[10]作为权值的更新算法。RPROP 算法是一种从改变学习速度的方面着手的对反向传播的改进算法, 其中一个特性就是根据符号序列更新步长而不是梯度值。在 RPROP 算法中有两个值进行了计算: 1) 更新值的符号, 2) 更新值的大小。更新值的符号计算如式(10)所示。

$$\Delta \omega_{ij}^{(t)} = \begin{cases} -\Delta_{ij}^{(t)}, & \text{if } \frac{\partial J}{\partial \omega_{ij}} > 0 \\ +\Delta_{ij}^{(t)}, & \text{if } \frac{\partial J}{\partial \omega_{ij}} < 0 \\ 0, & \text{else} \end{cases} \quad (10)$$

这里  $\frac{\partial J}{\partial \omega_{ij}}$  为当前步。更新值得大小如式(11)所示。

$$\Delta_{ij}^{(t)} = \begin{cases} \eta^+ * \Delta_{ij}^{(t-1)}, & \text{if } \frac{\partial J}{\partial \omega_{ij}}^{(t-1)} * \frac{\partial J}{\partial \omega_{ij}}^{(t)} > 0 \\ \eta^- * \Delta_{ij}^{(t-1)}, & \text{if } \frac{\partial J}{\partial \omega_{ij}}^{(t-1)} * \frac{\partial J}{\partial \omega_{ij}}^{(t)} < 0 \\ \Delta_{ij}^{(t-1)}, & \text{else} \end{cases} \quad (11)$$

这里  $0 < \eta^- < 1 < \eta^+$ 。最终权值更新公式如式(12)所示。

$$\omega_{ij}^{(t+1)} = \omega_{ij}^{(t)} + \Delta \omega_{ij}^{(t)} \quad (12)$$

### 3. 基于 PSO 的 CPNN

#### 3.1. PSO 概述

借助于生物觅食运动的启迪, Kennedy 和 Eberhart 提出了 PSO 算法[11]。PSO 算法通过个体间的协

作与竞争, 实现复杂空间中最优解的搜索。PSO 先生成初始种群, 即在可行解空间中随机初始化一群粒子, 每个粒子都为优化问题的一个可行解, 并由目标函数为之确定一个适应值(fitness value)。每个粒子将在解空间中运动, 并由一个速度决定其方向和距离。通常粒子将追随当前的最优粒子而动, 并经逐代搜索最后得到最优解。在每一代中, 粒子将跟踪两个极值, 一为粒子本身迄今找到的最优解  $pbest$ , 另一为全种群迄今找到的最优解  $gbest$ 。

虽然进化算法对那些大而复杂、误差梯度信息很难获取或根本不可用的问题特别有吸引力, 但与 BP 算法的局部搜索能力相比, 进化算法在接近问题最优解附近的解空间搜索就比较慢, 其局部搜索能力较弱。把进化算法与 BP 算法相结合可以增强算法的整体搜索能力, 这种混合算法的学习过程分两步: 首先, 利用进化算法的全局搜索能力找到一个较好的神经网络初始权值; 其次, 在进化算法搜索到的初始权值点处利用 BP 算法训练神经网络, 最终搜索到神经网络的最优连接权值。这种混合算法主要是利用了进化算法全局搜索的优点和 BP 算法局部搜索速度快的特点, 而又避免了进化算法局部搜索的弱点和 BP 算法容易陷入局部极小点的缺点。

### 3.2. PSO-CPNN

假设 CPNN 网络结构中含有  $n$  个隐层节点, 粒子被设计向量  $(\theta, \sigma)$  的形式,  $\theta$  表示 CPNN 网络的权值,  $\sigma$  为标签分布的参数。

PSO-CPNN 算法流程图如下所示。

---

#### 算法 1 PSO-CPNN

---

**输入:** 输入训练集  $S = \{(x_i, y_i)\}_{i=1}^n$ , 粒子群大小  $q$  和迭代次数  $T$ 。

**输出:**  $p(y|x; \theta)$

1. 初始化粒子的起始位置  $P^t = \{(\theta_i, \sigma_i)\}_{i=1}^q$  和起始速度  $v^t$ , 当前迭代次数  $t=0$ 。

2. while  $t < T$

3.     for  $k=1$  to  $q$

4.         for  $s=1$  to  $n$

5.             生成标签分布  $L_s(\sigma_i) = \{D_s^i\}_{i=\alpha}^\beta$ 。

6.             end

7.             计算  $L(\sigma_i)$  和  $p(y|x; \theta_i)$  的 Hellinger 距离。

8.             更新历史最优粒子和全局最优粒子。

9.             更新粒子速度  $v^{t+1}$ 。

10.             粒子位置  $P^{t+1}$ 。

11.             end

12.             设置  $t \leftarrow t+1$ 。

13.     end

14. 使用已经训练好的  $\sigma_t$  来生成标签分布。

15. 使用 Hellinger 距离作为损失函数, 用训练好的  $\theta_t$  作为 CPNN 网络的初始值, 接着用 RPROP 算法来训练 CPNN 的参数  $\theta^*$ 。

---

## 4. 实验

### 4.1. 设计与说明

本文中使用的数据集为 FG-Net Aging Database [12]。此数据集中含有 82 个人共 1002 张面部图片, 其中每人有 6~18 幅标有实际年龄的图像, 数据集中的年龄变化范围在 0~69 岁之间。除了年龄特征之外, 图片中的面部还带有姿势, 表情等特征。

在本文的实验阶段, 一些现有专门用于面部年龄估计问题的算法用于与本文的算法进行比较, 如: OHRank [13], AGES [14], WAS [15]和 AAS [16]。同样的, 传统的分类算法也进行了比较, 如: kNN, BP, C4.5, SVM, ANFIS [17]。

对比试验的参数设置如下, 在 OHRank 中, 使用了绝对值损失函数与 RBF 核函数。在 AGES 中, aging pattern subspace 的维度设置为 20。在 AAS 中, 在 appearance cluster training step 中的误差阈值设置为 3。在 kNN 中, k 设置为 30 并使用 Euclidean distance 来寻找邻点。BP 神经网络隐层节点数设置为 100, 并使用 sigmoid 函数作为激活函数。C4.5 在 SVM 中使用 RBF 核函数。ANFIS 中隶属函数设置为 2 个。Cpnn 中使用两种分布: 高斯分布, 标准差设置为 1, 2, 3, 4。三角分布, bottom length 设置为 2, 4, 6, 8。

年龄估计的性能指标使用 MAE (Mean Absolute Error), 即模型输出值为  $f(y_i | x_i)$ , 对应真实值为  $y_i$ , 则  $MAE = \frac{1}{n} \sum_{i=1}^n |f(y_i | x_i) - y_i|$ 。使用此指标可以很好的反映实际年龄和预测年龄之间的差距。

算法的测试阶段使用十折交叉验证, 即将数据集分成十分, 轮流将其中 9 份作为训练数据, 1 份作为测试数据, 进行试验。每次试验都会得出相应的 MAE。最后将 10 次的结果的 MAE 的平均值作为对算法精度的估计。

### 4.2. 结果与分析

实验结果如表 1 所示, PSO-CPNN 显示的是最好的结果, 可以看出, 标签分布算法的性能优于其他算法。原因主要是标签分布学习算法可以学习相邻年龄的信息。

进一步看, 高斯分布优于三角分布, 因为高斯分布利用所有相邻的年龄, 三角分布只利用了三角形内的年龄。然而, CPNN 对参数的选择比较敏感, 选择不同的参数对结果影响较大。图 1 显示了选择不同参数时 MAE 的变化。通过运用本文中的方法, 确定了较优的参数, 从而使得 MAE 降到了一个较低的值。

为了比较 KL 距离和 Hellinger 距离的性能, 单纯的 CPNN 算法也进行了, 如图 2 所示, Hellinger 距离在大部分情况下优于 KL 距离。

为了验证 CPNN 适用于训练样本较少的情况, 逐渐减少训练样本数量。通过每次减少一半的训练样本, CPNN 的 MAE 曲线如图 3 所示, 并与 OHRank 和 AGES 方法比较, 即使减少训练样本, CPNN 的 MAE 并不会增加很快。相反的, 当训练数据集减少的时候 CPNN 与 PSO-CPNN 表现得更加地稳定。但是, 当数据减少的比较多的时候, 由于模型复杂性更高的原因, PSO-CPNN 的性能劣于 CPNN。并且, 在对标签分布进行优化的时候, 当标签分布越接近均匀分布时, 两个年龄相异的标签分布的相似度就越近。图 4 便显示了 PSO-CPNN 算法中 MAE 的变化。

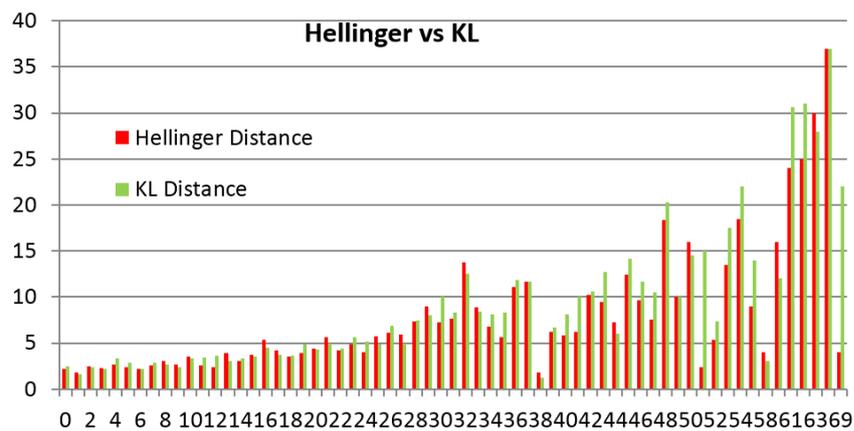
最终, 模型对应高斯分布的方差值如图 5 所示。

## 5. 总结和展望

本文中提出了一种新的用于年龄估计的神经网络 PSO-CPNN, 其特点是利用了粒子群优化算法来寻找条件概率神经网络的最优参数。从宽泛的意义上讲, 这种模型属于进化神经网络(evolutionary neural network, ENN)的一种, 而我们所提出这种算法克服了传统条件概率神经网络不能有效处理缺失数据集的

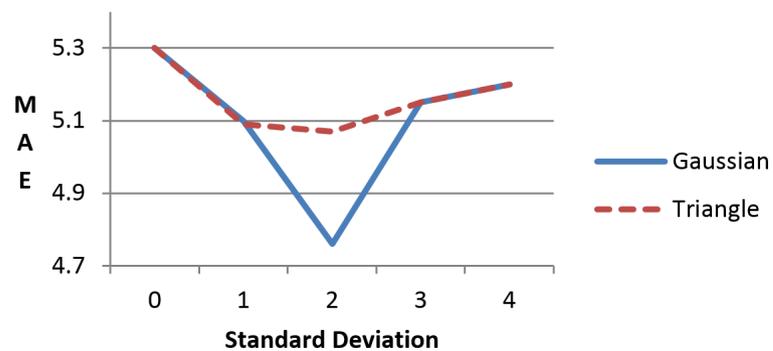
**Table 1.** MAE in different age estimators  
**表 1.** 不同模型的 MAE

| Method    |          | MAE   |
|-----------|----------|-------|
| My method | Gaussian | 4.39  |
|           | Triangle | 4.88  |
| CPNN      | Gaussian | 4.76  |
|           | Triangle | 5.07  |
| OHRank    |          | 6.27  |
| AGES      |          | 6.77  |
| WAS       |          | 8.06  |
| AAS       |          | 14.83 |
| kNN       |          | 8.24  |
| BP        |          | 11.85 |
| C4.5      |          | 9.34  |
| SVM       |          | 7.25  |
| ANFIS     |          | 8.86  |



**Figure 1.** Flowchart of estimation of distribution algorithm

**图 1.** 分布估计算法流程图



**Figure 2.** Compare Hellinger with KL

**图 2.** Hellinger 距离与 KL 距离的性能比较

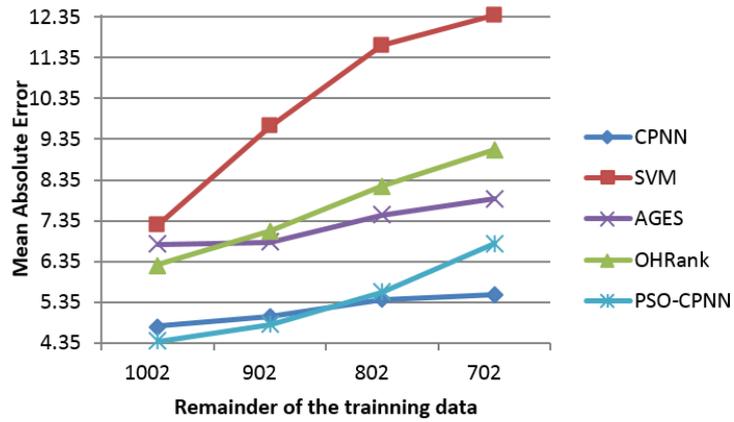


Figure 3. The performance on different estimators  
图 3. 不同模型性能比较

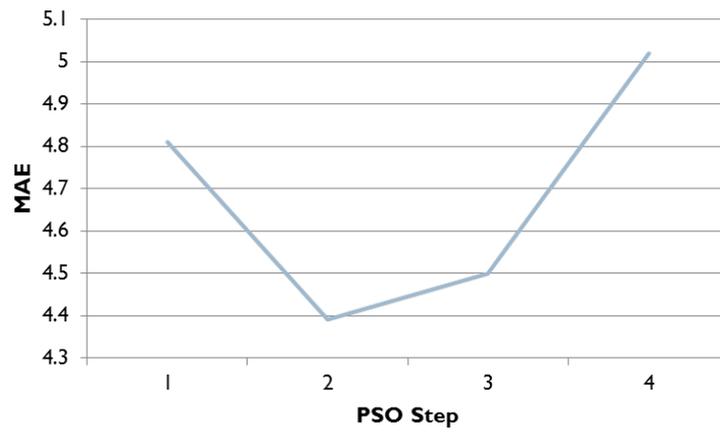


Figure 4. MAE variation of the PSO  
图 4. 算法过程中 MAE 变化情况

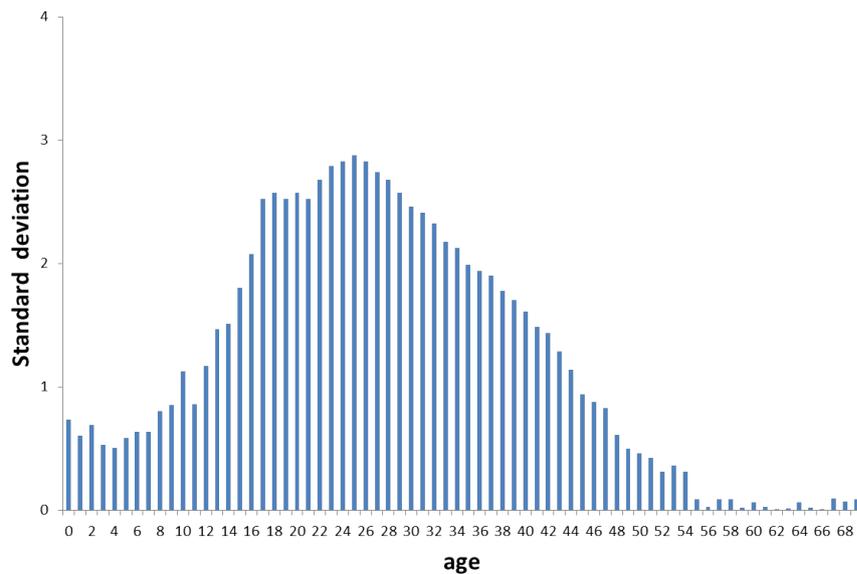


Figure 5. Standard deviations to different ages  
图 5. 不同年龄的标准差

缺点。实验表明, 本文中提出的方法的性能显著地高于其他用于年龄估计的算法。需要指出的是 PSO-CPNN 算法不仅能够用于优化条件神经网络的结构, 同时所有输出为分布形式的网络, 都可以使用本文所提出的方法来进行方法优化。

通过本文研究, 我们可以得出结论, CPNN 网络可以被粒子群算法优化, 由此, 我们相信结合不同的启发式算法, 来优化条件或深度网络的结构, 并将其应用到不同的领域, 将是我们未来的研究方向, 例如如何在深度网络中自动调参[18], 如何整合符号-神经元方法[19]-[21], 以及如何在机器人中运用本文的结果[22]-[25]等。

## 基金项目

此研究工作受到国家自然科学基金项目(No. 61003014, No. 61105026 和 No. 61273338)的资助。

## 参考文献 (References)

- [1] Kwon, Y.H. and da Vitoria Lobo, N. (1994) Age Classification from Facial Images. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Seattle, 21-23 Jun 1994, 762-767.
- [2] Hayashi, J., Yasumoto, M., Ito, H. and Koshimizu, H. (2002) Age and Gender Estimation Based on Wrinkle Texture and Color of Facial Images. *16th International Conference on Pattern Recognition*, **1**, 405-408.
- [3] Lanitis, A., Draganova, C. and Christodoulou, C. (2004) Comparing Different Classifiers for Automatic Age Estimation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, **34**, 621-628. <http://dx.doi.org/10.1109/TSMCB.2003.817091>
- [4] Nakano, M., Yasukata, F. and Fukumi, M. (2004) Age Classification from Face Images Focusing on Edge Information. *Knowledge-Based Intelligent Information and Engineering Systems*. Springer Berlin Heidelberg.
- [5] Zhou, S.K., Georgescu, B., Zhou, X.S. and Comaniciu, D. (2005) Image Based Regression Using Boosting Method. *10th IEEE International Conference on Computer Vision*, **1**, 541-548.
- [6] Geng, X., et al. (2006) Learning from Facial Aging Patterns for Automatic Age Estimation. *Proceedings of the 14th annual ACM International Conference on Multimedia*, ACM. <http://dx.doi.org/10.1145/1180639.1180711>
- [7] Yin, C. and Geng, X. (2012) Facial Age Estimation by Conditional Probability Neural Network. *Pattern Recognition*. Springer Berlin Heidelberg, 243-250. [http://dx.doi.org/10.1007/978-3-642-33506-8\\_31](http://dx.doi.org/10.1007/978-3-642-33506-8_31)
- [8] Hellinger, E. (1909) Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *Journal für die reine und angewandte Mathematik*, **136**, 210-271.
- [9] Sarajedini, A., Hecht-Nielsen, R. and Chau, P.M. (1999) Conditional Probability Density Function Estimation with Sigmoidal Neural Networks. *IEEE Transactions on Neural Networks*, **10**, 231-238. <http://dx.doi.org/10.1109/72.750544>
- [10] Riedmiller, M. and Braun, H. (1993) A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm. *IEEE International Conference on Neural Networks*, **1**, 586-591. <http://dx.doi.org/10.1109/icnn.1993.298623>
- [11] Eberhart, R.C. and Kennedy, J. (1995) A New Optimizer Using Particle Swarm Theory. *Proceedings of the 6th International Symposium on Micro Machine and Human Science*, Nagoya, 4-6 October 1995, 39-43. <http://dx.doi.org/10.1109/MHS.1995.494215>
- [12] Lanitis, A., Taylor, C.J. and Cootes, T.F. (2002) Toward Automatic Simulation of Aging Effects on Face Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**, 442-455. <http://dx.doi.org/10.1109/34.993553>
- [13] Chang, K.-Y., Chen, C.-S. and Hung, Y.-P. (2011) Ordinal Hyperplanes Ranker with Cost Sensitivities for Age Estimation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, 20-25 June 2011, 585-592.
- [14] Geng, X., Zhou, Z.H. and Smith-Miles, K. (2007) Automatic Age Estimation Based on Facial Aging Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **29**, 2234-2240. <http://dx.doi.org/10.1109/TPAMI.2007.70733>
- [15] Lanitis, A., Taylor, C.J. and Cootes, T.F. (2002) Toward Automatic Simulation of Aging Effects on Face Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**, 442-455. <http://dx.doi.org/10.1109/34.993553>
- [16] Lanitis, A., Draganova, C. and Christodoulou, C. (2004) Comparing Different Classifiers for Automatic Age Estimation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, **34**, 621-628.

<http://dx.doi.org/10.1109/TSMCB.2003.817091>

- [17] Jang, J.-S.R. (1993) ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Transactions on Systems, Man, and Cybernetics*, **23**, 665-685. <http://dx.doi.org/10.1109/21.256541>
- [18] Jiang, M., Ding, Y., Goertzel, B., *et al.* (2014) Improving Machine Vision via Incorporating Expectation-Maximization into Deep Spatio-Temporal Learning. *International Joint Conference on Neural Networks (IJCNN)*, Beijing, 6-11 July 2014, 1804-1811.
- [19] Jiang, M., Zhou, C. and Chen, S. (2010) Embodied Concept Formation and Reasoning via Neural-Symbolic Integration. *Neurocomputing*, **74**, 113-120. <http://dx.doi.org/10.1016/j.neucom.2009.11.052>
- [20] Wu, Y., Jiang, M., Huang, Z., *et al.* (2015) An NP-Complete Fragment of Fibring Logic. *Annals of Mathematics and Artificial Intelligence*, **75**, 391-417. <http://dx.doi.org/10.1007/s10472-015-9468-4>
- [21] Jiang, M., Yu, Y., Chao, F., *et al.* (2013) A Connectionist Model for 2-Dimensional Modal Logic. *IEEE Symposium on Computational Intelligence for Human-Like Intelligence (CIHLI)*, Singapore, 16-19 April 2013, 54-59.
- [22] Jiang, M., Yu, Y., Liu, X., *et al.* (2012) Fuzzy Neural Network Based Dynamic Path Planning. *International Conference on Machine Learning and Cybernetics (ICMLC)*, Xi'an, 15-17 July 2012, 326-330.
- [23] Chao, F., Lee, M.H., Jiang, M., *et al.* (2014) An Infant Development-Inspired Approach to Robot Hand-Eye Coordination. *International Journal of Advanced Robotic Systems*, **11**. <http://dx.doi.org/10.5772/57555>
- [24] Chao, F., Chen, F., Shen, Y., *et al.* (2014) Robotic Free Writing of Chinese Characters via Human-Robot Interactions. *International Journal of Humanoid Robotics*, **11**, 1450007. <http://dx.doi.org/10.1142/S0219843614500078>
- [25] Jiang, M., Huang, W., Huang, Z. and Yen, G.G. (2015) Integration of Global and Local Metrics for Domain Adaptation Learning via Dimensionality Reduction. *IEEE Transactions on Cybernetics*, **PP**, 1-14.