Ordinal Regression with Multiple Output CNN for Age Estimation

CVPR 2016, Zhenxing Niu, Mo Zhou, Le Wang, Xinbo Gao, Gang Hua.

汇报人：赖贵宇
汇报时间：2020年11月29日

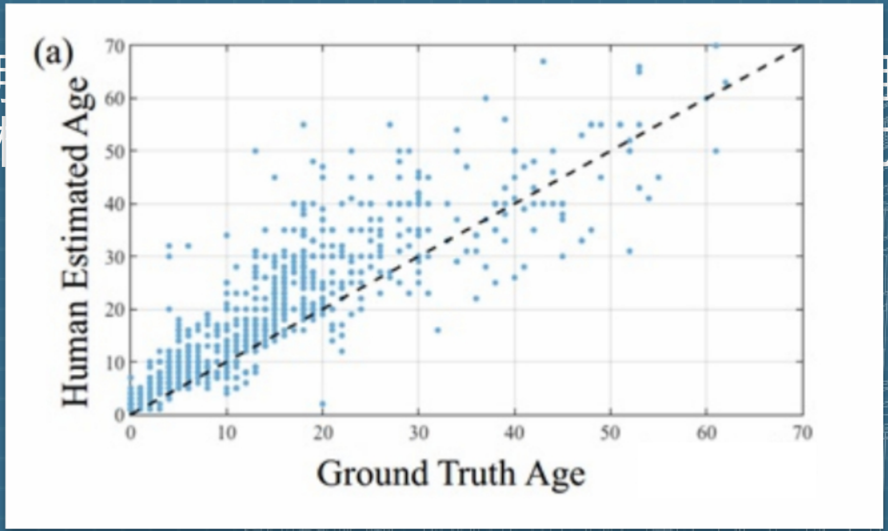
01 问题背景

- 年龄估计（Age Estima
  提供一张图像，自动识别出

- 技术应用
  视频监控、产品推荐、人机交互、市场分析、用户画像、年龄变化预测……

- 技术难点

人脸年龄与骨头形状、五官位置、肌……会受到光照、拍摄角度、姿态、表情等随机……对年龄的观察估计都会有很大的偏差。



02 常见思路

- **Classification**

  将年龄预测任务作为一个多分类任务处理，即1~100共100个类的多分类任务。但这样的方式忽略了年龄本身的连续性，独立地考虑各个类别间的差异。

- **Regression**

  将年龄当成连续值进行回归预测。但由于年龄与人脸很多随机因素有关，直接使用回归处理往往会造成过拟合。

- **Ranking**

  将年龄预测任务当作多个二分类比对子问题来求解，通过寻找当前年龄标签在年龄序列中的相对位置来确定最终的年龄预测值。

03 相关工作

● Local features and metric regression

● AAM model: simultaneously model the shape and texture of facial images

● Discriminative manifold learning and quadratic regression
● RankBoost
● OR-SVM

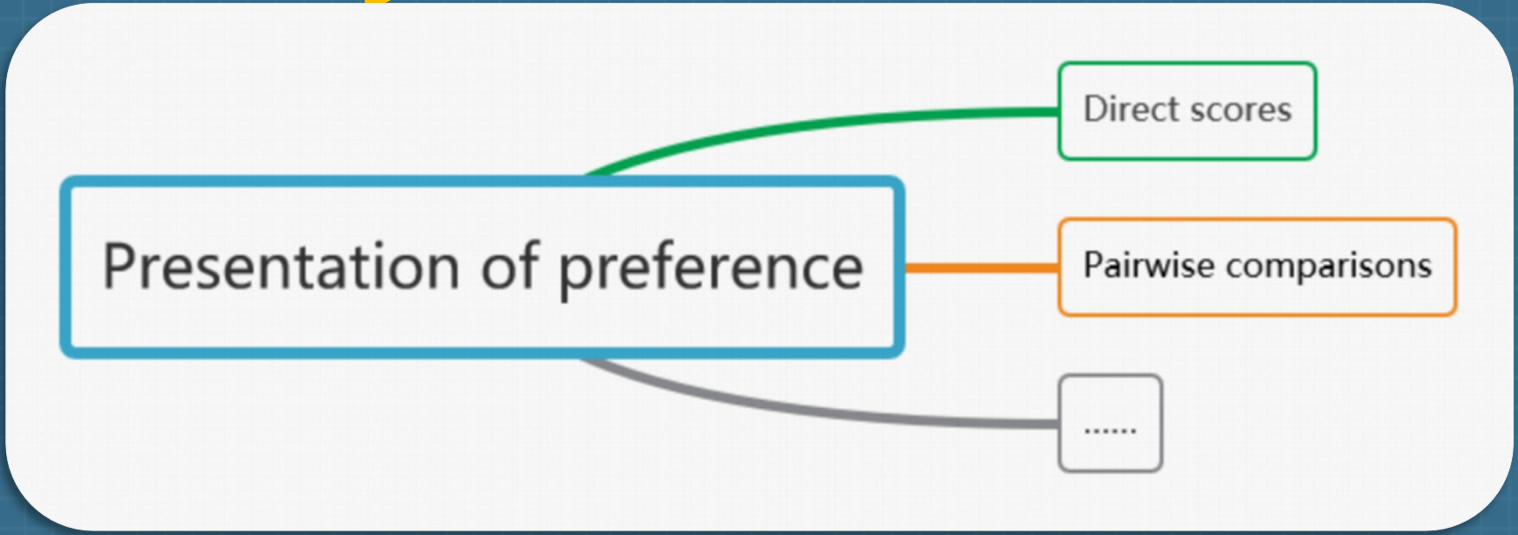
● Deep Learning way: using CNN to extract features and then feeding features to another regression for the final age estimation

03 相关工作

- **直接进行回归和分类的弊端**
  - 多分类问题把年龄标签看做是独立的，但实际上年龄是有序的
  - 在直接的回归模型中，人脸图像的头型、皮肤等不稳定因素容易导致模型过拟合
- **论文提出的新思路**
  - 将有序排序问题转换为一系列二分类问题，每个子问题只关注输入的图像的预测年龄是否比某个年龄值k大。

04 背景知识

● Ordinal Regression



MCDM (Multiple Criteria Decision Making)

- a finite set of alternatives $A = \{a_1, a_2, \ldots, a_m\}$
- n criteria (objectives) $F = \{g_1, \ldots, g_n\}$, with $g_i : R^m \longrightarrow R$
- G represents the objective space, for each pair of vectors x, y in G:

$$U(x) \geq U(y) \text{ if and only if } x \succsim y.$$

04

背景知识

● Ordinal Regression

$$U(x) \geq U(y) \text{ if and only if } x \succsim y.$$

However, the value function U of a DM is unkown.

A reverse search of the preference model from decision examples is done by so-called ordinal regression.

背景知识

● Ordinal Regression

Here come the question: how to obtain U(x) ?

A simple and well known way: UTA method

$$U[\mathbf{g}(a)] = \sum_{i=1}^{n} u_i[g_i(a)]$$

Subject to monotonicity and normalization constraints as follows:

$$\begin{cases} \sum_{i=1}^{n} u_i(\beta_i) = 1 \\ u_i(g_i(a)) \geq u_i(g_i(b)) \text{ if } g_i(a) > g_i(b) \\ \quad \text{and } g_i \text{ is a gain-type criterion,} \\ u_i(g_i(a)) \geq u_i(g_i(b)) \text{ if } g_i(a) < g_i(b) \\ \quad \text{and } g_i \text{ is a cost-type criterion,} \\ \qquad \forall a, b \in A, i = 1, \ldots, n \\ u_i(\alpha_i) = 0, \quad i = 1, \ldots, n. \end{cases}$$

Ordinal Regression

Usually, among the many sets of parameters of a preference model compatible with the preference information, only one specific set is used to give a recommendation on a set of alternatives.

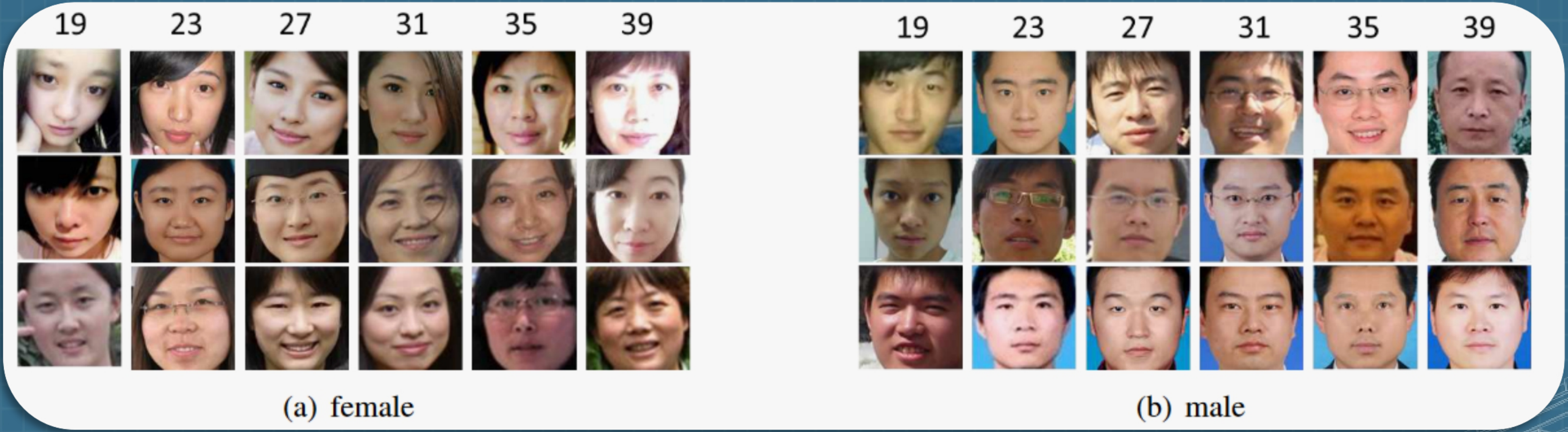
$$\begin{cases} \sum_{i=1}^{n} u_i(\beta_i) = 1 \\ u_i(g_i(a)) \geq u_i(g_i(b)) \text{ if } g_i(a) > g_i(b) \\ \quad \text{and } g_i \text{ is a gain-type criterion,} \\ u_i(g_i(a)) \geq u_i(g_i(b)) \text{ if } g_i(a) < g_i(b) \\ \quad \text{and } g_i \text{ is a cost-type criterion,} \\ \qquad \forall a, b \in A, i = 1, \ldots, n \\ u_i(\alpha_i) = 0, \quad i = 1, \ldots, n. \end{cases}$$

$$\max \varepsilon$$

$$(\text{PRVF}) \begin{cases} u_i(g_i(a_h)) \geq u_i(g_i(a_k)) \\ \quad \text{if } g_i(a_h) < g_i(a_k), i = 1,2,3,4; \; h, k = 1,2,3,4 \\ U(a_4) \geq U(a_1) + \varepsilon \\ U(a_1) \geq U(a_2) + \varepsilon \\ U(a_4) \geq U(a_2) + \varepsilon \\ u_i(4) = 0 \quad i = 1,2,3,4 \\ u_1(1) + u_2(1) + u_3(1) + u_4(1) = 1 \end{cases}$$

05 论文工作

● Problem Formulation



● 输入是一张图像x_i，输出是其对应的年龄标签y_i (取值$r_1, r_2, \ldots, r_K$)，标签间有序。任务是找到一个图像到年龄序数的映射。

05 论文工作

● Problem Formulation

● In order to measure the cost between predicted ranks and ground-truth ranks, the cost matrix is employed to do this task. Particularly, a popular choice for general ordinal regression problems is the absolute cost matrix:

		Ground-truth				
		1	2	......	99	100
Predicted	1	0	1	......	98	99
	2	1	0	......	97	98
	......	......	......	0	......	......
	99	98	97	......	0	1
	100	99	98	......	1	0

05 论文工作

● Approach

● 将有序回归问题转换为多个二分类子问题处理。

具体来说，有K个排序等级的有序回归问题可以被转换为K-1个二分类子问题。

在年龄预测问题中，年龄标签1~100，共100个排序级别，所转换出来的99个二分类模型的作用是分别判断当前图像的预测年龄是否大于某个级别。第一个二分类器用来判断预测年龄是否大于1岁，第二个二分类器用来判断预测年龄是否大于2岁……第99个分类器用来判断预测年龄是否大于99岁。

$$y_i^k = \begin{cases} 1, & \text{if}(y_i > r_k) \\ 0, & \text{otherwise,} \end{cases}$$

05

论文工作

• Approach

Input: training data $D = \{\boldsymbol{x}_i, y_i\}_{i=1}^N$ and testing images $D' = \{\boldsymbol{x}'_j\}_{j=1}^M$

- Loop for $k = 1, 2, \cdots, K-1$:
  - Build a distinct training data $D^k = \{\boldsymbol{x}_i, y_i^k, w_i^k\}_{i=1}^N$ for the k-th binary classification task according to Eq.1 and Eq.2.
- The learning of the multiple output CNN:
  - The proposed multiple output CNN is trained with $D^k, (k = 1, 2, \cdots, K-1)$ according to the proposed learning algorithm (refers to Sec.3.4).
- For each testing image $\boldsymbol{x}'_j \in D'$:
  - Forward $\boldsymbol{x}'_j$ to the trained CNN, and get the $K-1$ binary labels $f_k(\boldsymbol{x}'_j) \in \{0, 1\}, (t = 1, 2, \cdots, K-1)$;
  - Predict its rank $h(\boldsymbol{x}'_j)$ with previous binary labels $\{f_k(\boldsymbol{x}'_j)\}_{k=1}^{K-1}$ according to Eq.3.

Output: the predicted ranks for testing images $\{h(\boldsymbol{x}'_j)\}_{j=1}^M$.

$$h(\boldsymbol{x}') = r_q$$

$$q = 1 + \sum_{k=1}^{K-1} f_k(\boldsymbol{x}'),$$

电子科技大学
University of Electronic Science and Technology of China

05

论文工作

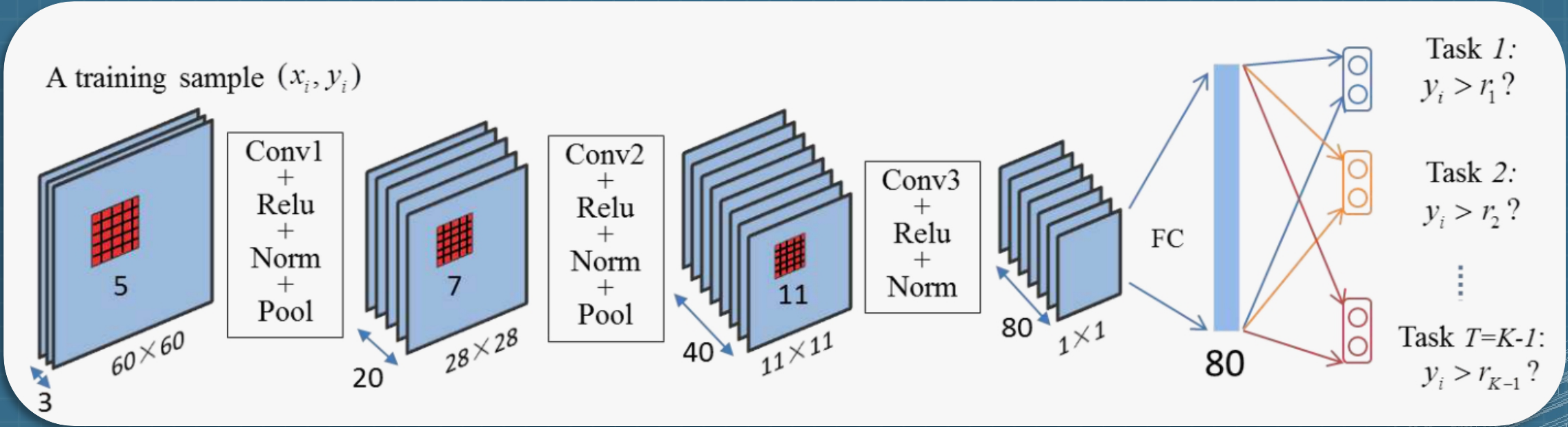
● Benefits

（1）使用端到端的学习方式，将图像中特征的提取与偏好模型的训练结合成一个整体，避免了手动设计特征的同时，可以更好地利用特征信息。

（2）K-1个二分类器的训练集仅仅是类别标签不同，它们可以共享中间卷积的信息，所以CNN能够学的不同任务之间的联系，从而提升最终的预测表现。

05 论文工作

● Architecture of the Multiple Output CNN



05 论文工作

- **Learning the Multiple Output CNN**

- 交叉熵损失函数：

$$E_m = -\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda^t 1\{o_i^t = y_i^t\} w_i^t log(p(o_i^t|\boldsymbol{x}_i, \boldsymbol{W}^t))$$

05 论文工作

● Experiment

Performance Metric:

- Mean Absolute Error (MAE)
- Cumulative Score (CS)

05 论文工作

● Experiment

Metric Regression VS Ordinal Regression

Dataset	Metric Regression			Ordinal Regression		
	BIFs + LSVR [15]	BIFs + CCA [14]	CNN + LSVR [31]	BIFS + OR-SVM [5]	BIFS + OHRank [6]	Ours (OR-CNN)
MORPH II	4.31	4.73	5.13 (4.77 in [31])	4.21	3.82	**3.27**
AFAD	4.13	4.40	5.56	4.36	3.84	**3.34**

05

论文工作

● Discussion

Task importance analysis

$$E_m = -\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda^t 1\{o_i^t = y_i^t\} w_i^t log(p(o_i^t|\boldsymbol{x}_i, \boldsymbol{W}^t))$$

$$\lambda_t = \frac{\sqrt{N_k}}{\sum_{k=1}^{K}\sqrt{N_k}}$$

Datasets	Uniform	Data-specific Scheme
MORPH II	3.30	**3.27**
AFAD	3.41	**3.34**

05 论文工作

Discussion

The color information

Datasets	Gray Image	Color Image
MORPH II	3.42	**3.27**
AFAD	3.44	**3.34**

05 论文工作

● Trick

为了防止过拟合，在图像预处理时进行了随机剪裁

06 论文主要贡献

- 使用端到端的多任务深度学习方法代替之前的SVM解决有序回归问题
- 将提出的模型应用到年龄预测任务中，并取得了目前最好的结果
- 公开发行了目前最大的以亚洲人为主的人脸年龄数据集AFAD

07 启示与展望

- 将一个排序问题转换为若干个二分类子问题处理
- 提升二分类器的准确度以提升年龄预测的精度