

# 基于用户脸部信息的推荐系统设计<sup>①</sup>

许晓飞<sup>②</sup> 常 健 杨重柱 范文超

(北京信息科技大学自动化学院 北京 100192)

**摘 要** 针对大数据时代的数据利用率不高情况,分析比较基于用户的协同过滤推荐算法和基于物品的协同过滤推荐算法;结合 2 种算法的优点,设计基于用户人脸信息的实时采集入库以及数据分析推荐系统,采用 python 语言编写程序功能模块。实验数据表明,在使用包含 106 863 张 530 名男女人脸图像的数据集训练和测试后,与传统有 76.6% 的识别率的支持向量机(SVM)分类器模型、以及有 94.8% 的识别率的融合局部二值模式(LBP)算法及 SVM 分类器算法的模型相比,在使用卷积神经网络(CNN)算法对该数据集构建模型则有 98.1% 的识别率,相较前 2 种算法分别提升了 21.5% 和 3.3%。因此,使用卷积神经网络算法训练数据集可以获得较高人脸检测及识别性别、年龄精度的模型。

**关键词** python, 协同过滤, 推荐系统, 卷积神经网络(CNN)算法, 人脸推荐

## 0 引 言

人脸识别是一项日趋成熟的生物特征识别技术,越来越广泛应用于金融、医疗、安防、行政、交通等领域。而推荐技术作为智能数据分析的一种方法,其建立在大数据的基础上,使用数学模型算法将需求者和这些数据连接起来,不仅可以使需求者更高效地发现对自己有价值的数据信息,而且还能够轻松地把数据呈现在对它感兴趣的需求者的面前,使数据和信息不浪费,需求者与生产者共利,提高了大数据时代的效率<sup>[1-3]</sup>。人脸脸部信息识别和推荐技术相结合,能够根据推荐系统的原理及其如何利用用户轨迹数据进行分析以及经典推荐算法进行应用研究,包括基于用户/基于物品的协同过滤推荐算法,并将两者进行深度研究对比,选择一种算法对推荐系统进行设计<sup>[4-6]</sup>。

人脸由于自然性,其识别具有不易察觉性,一直是模式识别和人工智能视觉领域的研究热点,但人

脸容易受到光照、噪声、表情等影响,有效准确获取人脸特征信息是研究的重点<sup>[7-9]</sup>。本文采用卷积神经网络(convolutional neural networks, CNN)算法,基于 python 和 caffe 框架实现对面脸的识别和分析,通过分析人脸面部特征,得出用户的年龄段与性别等,根据所得结果对用户进行分类。根据返回的用户分类结果,推荐系统选择相应的商品为用户进行推荐<sup>[9-12]</sup>。

## 1 推荐算法描述

面对大数据所带来的选择问题,需要某种工具或他人来帮助解决这一问题。推荐系统实质上是一种信息过滤系统,将被推荐目标的性质内容与被推荐用户的兴趣内容相互吻合,并利用特定的推荐算法计算、过滤、挑选,最后选出与用户兴趣最接近的对象作为推荐结果展现在用户面前,有助于人们做出个性化选择的软件系统。

推荐系统由 3 个最主要的部分组成:第 1 个是

① 北京市大学生科技创新(校教发[2019])和北京高等学校高水平人才交叉培养“实培计划”大学生[2018-2019 年度]资助项目。

② 女,1980 年生,博士;研究方向:机器人技术,智能系统设计与开发;联系人,E-mail: 18910782910@163.com  
(收稿日期:2018-06-07)

推荐的对象,如视频、音乐、图书等;第2个是被推荐的用户;第3个是推荐算法,是最重要的部分,直接反映系统的推荐性能高低。图1所示为典型推荐系统模型原理图。

基于用户的协同过滤算法在推荐算法中是相当经典的古老算法<sup>[13-16]</sup>。它的主要思想是给定一个评分矩阵和一个准备为之推荐物品的用户,根据评分矩阵结果找到与这个用户相似度最接近的多个用户,然后从多个用户中选择该用户不曾购买过的商品推荐给他。除了这种基于记忆的协同过滤技术外,还有一种就是基于模型的协同过滤技术能够直接处理离线的原始数据<sup>[17]</sup>,较容易出现只推荐销量比较高的商品。还有基于因果知识的计算方法得以利用。推荐系统通常会获取额外的信息,这种方法的应用较多,基于约束条件的推荐系统便是其中之一,某些数据与用户高度相关,比如用户喜欢玩游戏,那么要求图形处理器(GPU)有强大的显示能力、中央处理器(CPU)处理能力及GPU的功耗。因此,基于知识的推荐系统也需要实时更新用户数据,在某些特定的场景还会由系统调查用户需求<sup>[18]</sup>。

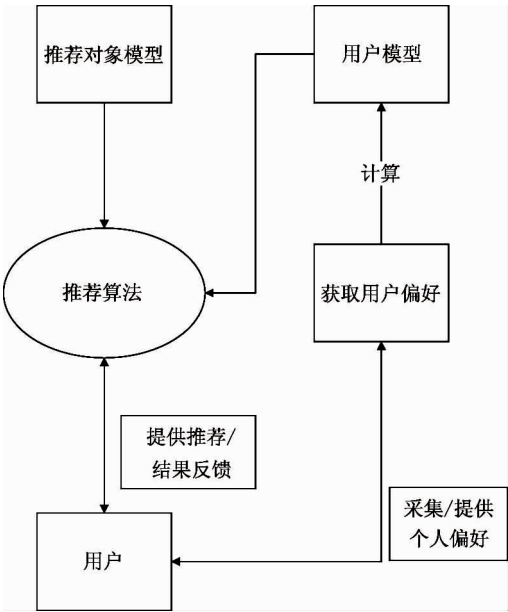


图1 推荐系统模型图

本文设计推荐算法兼顾基于用户协同推荐和基于物品协同推荐的算法,以基于用户的协同过滤算法为核心,如在为用户 $U$ 进行推荐时,首先需要找

到与该用户的兴趣相近的多个用户,然后把这些与用户 $U$ 相似的用户喜欢的东西并且这些东西用户 $U$ 不曾接触到过推荐给用户 $U$ 。本文设计也是基于这个思想,以用户为中心,兼顾用户及其社交关系,以用户ID、物品ID及对应的评分数据为输入,为用户准确推荐可能喜欢的物品为输出,使用余弦相似性公式实现用户之间相似性计算,通过预测评分公式实现预测评分,最终实现推荐目的。

## 2 基于用户脸部信息的推荐系统软件算法设计

### 2.1 用户脸部信息特征识别属性

本文使用CNN以及faster RCNN卷积神经网络算法进行测试,基于python和caffe平台实现对人脸的识别和分析。通过分析人脸面部特征,得出用户的年龄段与性别等,并根据所得结果对用户做分类处理。根据返回的用户分类结果,推荐系统选择相应的商品为用户进行推荐。

完成人脸识别有4个核心步骤,首先采集图像,其次检测图像中人脸定位位置,接着对图像进行预处理,最后是提取、匹配图中人脸特征并识别。

在匹配与识别人脸的过程中常用的4种特征,分别是图像的直接视觉特征、像素特征、变换参数及其数学特征。提取人脸图像特征的过程,是基于提取出的人脸图像的特征完成对人脸的建模。

本文人脸关键特征点距离检测分3步。第1步先在整个人脸框内预测各个关键点,关键点为眼睛、鼻子、嘴角。第2步将关键点区域画出一个小一些的框,再进行预测,对第1步的预测结果进行矫正。第3步同第2步。人脸关键点位置坐标的检测与计算,首先在人脸检测后把人脸位置标出一个框,然后预测人脸关键点与该人脸框的相对位置进行比较,训练时输入带有人脸框与关键点标签的图片,先计算人脸关键点与人脸框的相对坐标,例如左眼关键点计算时用 $x$ 坐标-人脸框左边框的 $x$ 坐标, $y$ 坐标-人脸框上边框的 $y$ 坐标,然后用计算出的 $x$ /人脸框宽度 $w$ ,用 $y$ /人脸框高度 $h$ ,计算出关键点与人脸框相对位置。

基于已有知识的特征提取方法主要包括分析人脸结构的几何特征和将人脸与模板作对比获取特征。采用 FaceScrub 数据集,这些图像是从互联网上检索出来的,并且是在现实世界的情况下(不受控制的情况下)拍摄的。包括面孔的名称和性别注释。它共包含 106 863 张 530 名男女人脸图像,每人约 200 张图片。因此,它是最大的公众人脸数据库之一。

使用 caffe 框架与上述数据集训练 CNN 算法模型并保存为 caffe 模型,在识别过程中调用训练得到的 caffe 模型和网络结构,caffe 训练模型时的网络结构是 vgg-16。vgg-16 一种深度卷积神经网络模型,16 表示其深度,该结构有 13 个卷积层和 3 个全连接层,卷积神经网络的话训练时图片输入进模型后经过卷积最终变为一个向量,该向量即为图像特征,当待识别的图片输入后,经过卷积获取特征向量后与训练出的特征向量作比来完成识别。可以加快程序运行速度,避免每一次运行程序都要对神经网络进行训练。

匹配图像中的人脸并识别,将前一步从图像中获取出的人脸特征数据与预先制作完成的样例进行对比,通过匹配结果判断图中是否为人脸、该人脸是否是数据库中登录过的某人。从图像中获取的需要识别的面部特征与已有样例的特征进行匹配,并依据这些特征的相似程度来判断人脸的身份信息的过程就是人脸识别的过程。设置一个阈值,当提取出的特征与数据库中样例的特征相似度超过该阈值时输出匹配结果。这个步骤可分为是确认该人脸(将一张图像与另一张进行比较),或是识别该人脸(将一张图像与多张图像匹配和对比)。

## 2.2 用户脸部信息识别算法描述分析

本文中构建人脸识别系统使用的是识别效率高的 CNN 深度学习卷积神经网络算法。在处理图像时通常选择用一个个像素构成的矢量表示整张图像;假设构建神经网络时隐藏层的数目与输入层相同,则隐藏层输入层的参数数据将是图片本身像素值数的平方,数据量过于庞大易导致训练难度过高。因此,为了使用神经网络处理图像,首要任务是减少每个神经元对应的参数数量和加快数据处理速度。

### (1) 局部感知

不需要每一个神经元都感知图像整体,只需要感知图像的一部分,然后通过组合较高级别的局部信息来获得全局信息。当一个神经元仅与少量的像素相连时,参数数据将是神经元连接的像素数值乘以原图像的像素矢量,数据量会大幅度减少,相当于做了卷积运算。

### (2) 参数共享

在局部感知的过程中,所有神经元都与其所连像素的参数相对应,在所有的神经元中,如果有部分神经元的参数都相等,则那一部分参数的数目可以看作单个神经元的参数,即参数相同的神经元可以作合并处理减少参数数量,与神经元及其所连像素的位置没有关联。

### (3) 卷积层计算流程

假设  $L$  层是卷积层, $L+1$  层是降采样层。则计算在第  $L$  层的第  $J$  个特征映射的残差时可以使用式(1)。

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

公式里的  $*$  代表卷积核  $k$  与在  $L-1$  层上所有相关的特征映射做卷积运算,然后进行求和运算并加上偏置参数后取  $S$  函数得到激励值。

卷积层之后是采样层,由于在采样时使用一对一且非重叠的方式,所以在计算特征映射的残差时可以使用式(2),其中  $L$  代表层数, $J$  代表个数。

$$\delta_j^l = \beta_j^{l+1}(f'(u_j^l) \circ up(\delta_j^{l+1})) \quad (2)$$

$L$  层为卷积层, $L+1$  为取样层。采样层与卷积层是一对一的。 $Up(x)$  将  $L+1$  层的大小扩展到与  $L$  级别相同的大小。由于它是一对一采样, $L$  层的每个节点仅对应于扩展后的  $L+1$  层中的唯一节点。所以在计算残差时使用误差反向传播算法(BP 算法)中计算残差的公式可得出第  $L$  层中每个节点的残差值等于  $L+1$  层中与该节点相对应的节点的残差值乘以权值  $W$  然后乘以  $F(z)$  的导数。

第 1 次做卷积时,每个神经元只与所设定的卷积核的大小相等的像素连接,故得出的结果很片面。而当第 2 次做卷积时,每个神经元会与卷积核大小增加两行两列的像素连接,比如卷积核大小为  $3 \times 3$ ,则第 2 次做卷积时每个神经元会与  $5 \times 5$  个像素

连接。每一次做卷积,神经元都会扩大在原图上的感知范围,因此经过多次卷积后即可通过较少的神经元感知整个图片并获得图片内物体的轮廓。

2.3 推荐算法运行分析

本文设计的推荐算法,兼顾基于用户的协同过滤算法和基于物品的协同过滤算法设计的个性化推荐系统,首先对用户、商品以及评分数据进行训练,在相似度计算过程中同样先建立用户到商品的倒排序,在推荐数据的过程前需要先确定用户对于被推荐商品的兴趣度作为偏好值,最终推送给用户排分在前的商品。

在推荐物品相似度计算中,第 1 步列出商品到用户的倒排列列表,然后计算用户间的相似度。对于每一个商品都已经在列表中存储了评价过这个商品的用户,建立代表用户-物品倒排表和物品的相似度矩阵,如图 2 所示。计算出物品的相似度后,进入第 2 步,给用户生成物品的推荐清单,通过式(3)计算:

$$P(u, i) = \sum_{v \in S(u, K) \cap N(i)} W_{u, v} r_{v, i} \tag{3}$$

其中,  $P(u, i)$  代表用户  $u$  对物品  $i$  的兴趣值,  $N(i)$  代表用户  $u$  喜欢的物品群,  $S(u, k)$  表示和物品  $i$  相像的数量为  $K$  的物品群,  $W_{u, v}$  代表用户  $u$  的感兴趣的物品和  $v$  的感兴趣的物品间相似度,  $r_{v, i}$  是用户  $v$  对商品  $i$  的喜好水平值。然后根据求得的感兴趣水平的数值,选择该用户感兴趣水平较高的商品并为他进行推荐。

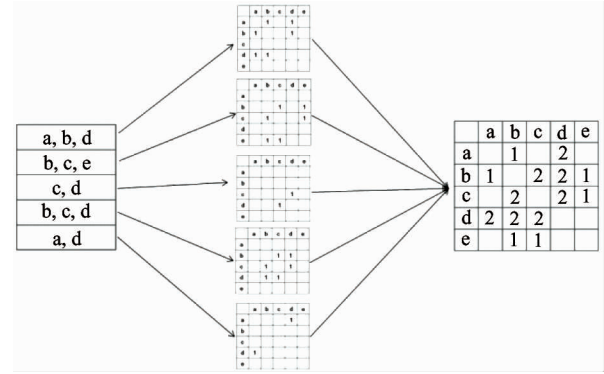


图 2 物品的相似度矩阵

基于用户的协同过滤算法、基于物品的协同过滤算法、以及本文提出的算法,这 3 种算法运行时间分析,本文所提出的算法在计算机运行同样推荐任务完成运行时间最短,而基于用户的协同过滤算法

略长于基于物品的协同过滤算法。这说明,本文提出的算法事先对用户、物品进行了分类,新用户进入后对其分类并推荐相应的物品类别,根据用户选择的物品相应调整分类权重,因而小范围测试运行速度比 2 种算法的运行速度要快,且可以根据用户反馈实时调整,其缺点是无法判别反馈的正确性;在运行计算时迭代次数上目前都是单次计算的。

2.4 评估标准

推荐质量的评价标准有 3 个量化指标:准确度、信任度和实时性。准确度用于衡量一个推荐引擎的预测能力,在电子商务行业,用户的信任度是非常重要的。透明度的增加有利于推荐系统受用户信任的程度的增加。最后,推荐系统具有快速实时性,才能更好为用户做出推荐。

人脸识别的评估标准主要有误识率、拒识率和识别正取率。误识率是错误地将别人识别作指定人的概率,拒识率是错误地将指定人识别作别人的概率,识别正取率是识别正确的人次与总参与人次的比值。本文采用识别正取率作为人脸识别系统的评估标准。

推荐系统的评估标准主要为用户点击推荐系统所推荐的物品的比率,即点击率,假设为  $N$  个用户推荐了  $n$  次物品,而其中有  $k$  次用户选择了推荐的物品,则点击率  $c$  为  $c = \frac{k}{n}$ 。

3 实验与结果分析

3.1 实验环境

在 python2.7 环境下编写推荐系统程序,在 python2.7、Anaconda2、opencv2.4.13.4、caffe 环境下编写人脸识别程序。

3.2 数据集

大型人脸数据集对于推进人脸识别研究很重要,本文设计开发了一种摄像头实时采集周围空间环境的人脸信息,直接后台入库建立人脸数据集硬件支持模块。数据集加入检测因特网上的公众人物搜索返回的图像中的人脸,自动舍弃那些不属于被查询的人的图像,综合处理实现数据集且能快速有效地构建。

使用 CNN 网络训练模型时,首先在训练通过提取人脸脸型、器官结构及其间距等特征获取人脸信息,根据这些信息分别建立人脸检测与识别、年龄识别、性别识别、表情识别与种族识别模型,并在识别过程中使用这些模型分工完成各项识别,相较单个模型完成多项特征的识别更容易准确获取图像信息。

3.3 用户脸部信息实时采集推荐系统

3.3.1 用户脸部信息实时追踪采集模块

用户人脸实时追踪采集模块中,通过系统摄像头的全方位扫描式采集周围空间环境中的人脸动态图像集,抓取人脸静态图像后,自动转入系统上位机软件部分的人脸检测模块,经过处理后可以获取图像中检测出人脸的区域,并将该区域在图像中的位置标出。使用舵机与摄像头作为主要零件构成人脸追踪系统,其硬件为一个 2 自由度的机械臂以及夹在机械臂前端的摄像头,机械臂可以通过上下 2 个在不同平面上旋转的舵机完成转动,扩大摄像头的探测范围,并实现人脸追踪的目的。人脸追踪模块的硬件部分组成如图 3 所示。

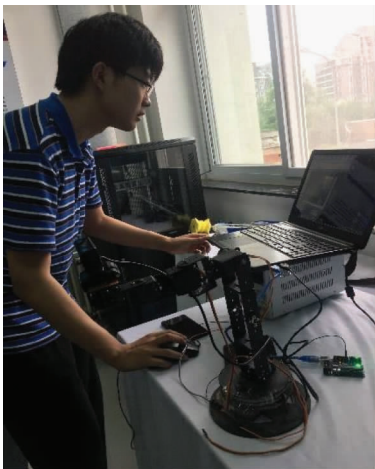


图 3 人脸追踪实时采集功能模块硬件部分

在摄像头中未出现人脸的时候,程序控制机械臂转动,摄像头可以调试的频率转动进行人脸搜索,一旦图像中检测到人脸,立即识别该人脸的信息,并将该人脸作为人脸追踪的目标,根据该区域在图像中所处的像素区域中心点与图像中心点之间的位置关系,相应调整摄像头拍摄角度,从而将人脸位置转

移至图像中央,实现人脸追踪。若不需要进行人脸追踪,也可以关闭人脸追踪模块,在人脸搜索的过程中保持人脸检测,收集广域图像内多个人脸的信息以进行身份识别。

3.3.2 用户脸部信息的年龄、性别特征识别推荐情况分析

如图 4 所示,程序使用摄像头进行图像实时采集并进行图像处理一系列操作后反馈出的识别结果,可以看到该程序能够准确识别出人脸位置及性别。



图 4 系统实时采集人脸脸部信息进行性别识别

图 4 中右上角 3 个按钮分别为开始人脸检测、暂停人脸检测和退出,下方 4 个复选框中 Detection 代表是否检测人脸位置、Landmark 代表是否标示出人脸检测的关键点、Gender 代表是否显示性别。

如图 5 所示,该程序也能够准确识别出人脸位置及该人脸所处年龄段。

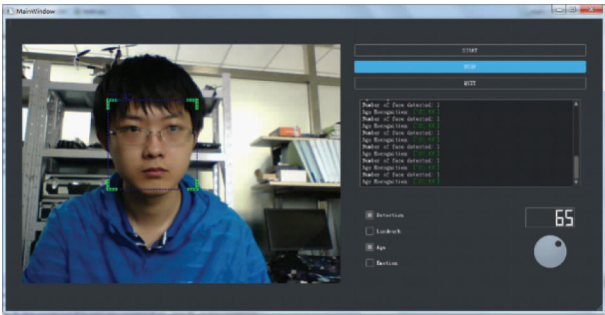


图 5 系统实时采集人脸脸部信息进行年龄段识别

图 5 下方 4 个复选框中 Detection 代表是否检测人脸位置、Landmark 代表是否标示出人脸检测的关键点、Age 代表是否显示年龄。

通过对图 6 进行分析,在使用相同的数据集进

行训练后,传统的支持向量机(support vector machine, SVM)分类器对该人脸数据集的识别率不高,仅有约 76.6% 的识别率。融合局部二值模式(local binary patterns, LBP)算法及 SVM 分类器算法后约有 94.8% 的识别率,而本文使用的 CNN 算法对该数据集约有 98.1% 的识别率,相较前两种算法分别提升了 21.5% 和 3.3%。由此可知,使用卷积神经网络算法可以有效地提升人脸检测及识别性别精确程度。

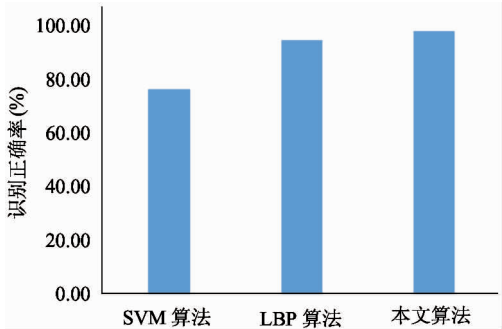


图 6 3 种算法人脸识别率统计图

2 种传统的协同过滤算法从性能角度上看,基于用户的协同过滤算法更适合计算人数较少的场景,如果用户的数量多了,求得用户间的相似度矩阵的过程的复杂程度就会很大。而基于物品的协同过滤算法则倾向于物品数量明显小于用户数量的场景,如果物品的数量巨大,求得物品间的相似度矩阵的过程的复杂程度就会很大。鉴于基于用户的算法以及基于物品的算法各有优劣,本文使用以上这 2 种算法实现本推荐系统的设计。首先为每一类用户手动初始化物品评分矩阵,根据人脸识别系统的反馈结果为新的用户进行分类,根据该分类及物品评分矩阵为用户进行推荐;而后根据用户是否点选该物品以及用户对该物品的评分,对物品评分矩阵进行优化,选择用户评分高的物品及与其相似的物品进行推荐。

本文构建的推荐系统尚未经过大型数据集测试,只在小范围的测试内对 50 人进行了测试,推荐点击率为 16%。从图 7 中可以看出本文算法优于 2 种传统算法,但是由于测试人数过少且训练不充分可能引起结果的不确定性,下一步将在扩大训练集

的基础上对如何有效利用用户的评分以及对用户评分稀疏时如何减少其负面影响进行研究。

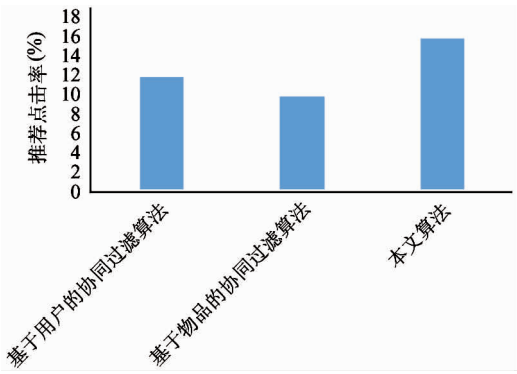


图 7 3 种推荐算法的推荐点击率对比图

本文对系列用户脸部信息数据挖掘出的年龄、性别、表情、种族提取其中的年龄和性别信息用于给用户分类进行个性化推荐。用户的性别与年龄一定程度上代表了用户可能购买的商品类型以及用户的购买能力,因此获取这些信息有助于推荐系统为用户做出合理的推荐。将人脸识别应用到推荐系统中使得推荐系统不必获取用户的购物历史即可为用户作出个性化推荐,实现通过采集用户人脸信息、数据挖掘出用户年龄、性别用于推荐的基于用户的推荐系统。

目前此推荐系统尚未使用网络上的大型数据集,仅采用少量的用户进行初步的测试,初始化物品评分矩阵时采用 20 位用户对 30 个物品的共 600 条评价的数据。

4 结 论

本文设计开发的一种摄像头实时采集周围空间环境的人脸信息,直接后台入库等综合处理实现数据集的快速有效地构建;配置了最新的 python 环境,编写了人脸推荐系统的功能软件,其中采用了卷积神经网络算法,采集 106 863 张 530 名男、女人脸图像的数据集训练,测试实现了 98.1% 的识别率;对比应用 SVM 分类器模型,以及融合 LBP 算法及 SVM 分类器算法的模型证明了本文所建模型是较高人脸检测及识别性别、年龄精度的模型;将人脸识别应用到推荐系统中使得推荐系统不必获取用户的

购物历史即可为用户做出个性化推荐,与此同时也能更快地获取用户信息为用户进行分类,提高推荐速度,实现通过采集用户人脸信息、数据挖掘出用户年龄、性别用于推荐的基于用户的推荐系统。

下一步投入使用 Faster R-CNN 图像处理算法,实现更快地训练和识别速度以及更高的精度,克服目前本文构建的人脸推荐系统存在人脸信息提取及反馈较慢的问题;另一方面需要提高在动态、快速、模糊情况下的人脸识别精度,拟在图像处理步骤中加入图像分割以及图像增强步骤,将背景以及前景中模糊且不可恢复的区域剔除,将剩下的模糊区域进行图像增强,提高识别精度。

## 参考文献

- [ 1 ] 祝晓斌,蔡强,白璐,等. 一种基于标签和协同过滤的并行推荐算法[J]. 高技术通讯, 2015, 25(3): 307-312
- [ 2 ] Zou J, Ji Q, Nagy G. A comparative study of local matching approach for face recognition[J]. *IEEE Transactions on Image Processing*, 2007, 16(10): 2617-2628
- [ 3 ] 殷复莲,柴剑平,高雅. 数字电视节目推荐系统结构及推荐算法研究[J]. 高技术通讯, 2014, 24(7): 677-683
- [ 4 ] Gong J B, Wang L L, Sun S T, et al. iBole: a hybrid multi-layer architecture for doctor recommendation in medical social networks[J]. *Journal of computer science and technology*, 2015, 30(5): 1073-1081
- [ 5 ] Herlocker J L, Konstan J A, Riedl J. Explaining collaborative filtering recommendations[C]. In: Proceedings of 2000 ACM Conference on Computer Supported Cooperative Work, Philadelphia, USA, 2000. 241-250
- [ 6 ] 叶剑锋,王化明. AdaBoost 检测结合 SOM 的自动人脸识别方法[J]. 哈尔滨工程大学学报, 2018, 39(1): 129-134
- [ 7 ] Ahonen T, Hadid A, Pietikainen M. Face description

- with local binary patterns: application to face recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(12): 2037-2041
- [ 8 ] 何林巍,黄福珍. 基于 POEM\_SLPP 的人脸识别算法[J]. 计算机应用研究, 2017, 34(6): 1896-1899
- [ 9 ] 陈雯柏 许晓飞. 智能机器人原理与实践 [M]. 北京: 清华大学出版社, 2016: 47-66
- [ 10 ] 孙亭,贾元江,张永胜,等. 基于 Hadoop 的推荐算法研究[J]. 软件工程, 2015(4): 59-62
- [ 11 ] Guo Q L. The similarity computing of documents based on VSM[C]. In: Proceedings of the 32nd Annual IEEE International Computer Software and Applications, Turku, Finland, 2008. 585-586
- [ 12 ] 马腾,陈庶樵,张校辉,等. 基于规则集划分的多决策树报文分类算法[J]. 计算机应用, 2013, 33(9): 2450-2454
- [ 13 ] Burke R. Hybrid recommender systems: survey and experiments[J]. *Artificial Intelligence Magazine*, 2002, 12(4): 331-370
- [ 14 ] Ricci F, Rokach L, Shapira B, et al. Recommender Systems Handbook[M]. Boston: Springer, 2011. 1-35
- [ 15 ] 郑美茹. 遗传算法在求解函数优化问题中的应用研究[J]. 装备制造技术, 2011(6): 20-21
- [ 16 ] Yeung K F, Yang Y, Ndzi D. A proactive personalised mobile recommendation system using analytic hierarchy process and Bayesian network [J]. *Journal of Internet Services and Applications*, 2012, 3(2): 195-214
- [ 17 ] Ochi P, Rao S, Takayama L, et al. Predictors of user perceptions of web recommender systems: how the basis for generating experience and search product recommendations affects user responses[J]. *International Journal of Human-Computer Studies*, 2010, 68(8): 472-482
- [ 18 ] 王小林,王义. 改进的基于知网的词语相似度算法[J]. 计算机应用, 2011, 36(11): 3075-3078

# Design of recommendation system based on user's face information

Xu Xiaofei, Chang Jian, Chen Wenbai, Fan Wenchao

( School of Automation, Beijing Information Science and Technology University, Beijing 100192)

## Abstract

In view of the low data-using efficiency situation in the era of big data, item-based collaborative filtering algorithm and user-based collaborative filtering algorithm are analyzed and compared, in which the face image could be gathered real-time; Python program language is used to conduct functional module program design in a data mining recommendation system which is the advantages of these two algorithms. The experiment results indicate that after using convolutional neural network (CNN) algorithm to training and testing by a dataset which contains 106 863 images of 530 different males and females, there is about 98.1% recognition rate. Compared with the traditional support vector machine (SVM) classifier model with a recognition rate of about 76.6%, and a fusion algorithm of LBP and SVM classifier algorithm with a recognition rate of about 94.8%, CNN algorithm model improves the face recognition rate by 21.5% and 3.3%. Therefore, the high accuracy of face detection and recognition model for two algorithms variables including gender and age, which could be accomplished by using convolution neural network algorithm to train data sets.

**Key words:** python, collaborative filtering algorithm, recommendation system, convolutional neural network (CNN) algorithm, face recommendation