

# 基于上下文约束的人脸聚类算法

罗恒利 王文博 葛宏孔

(南京航空航天大学计算机科学与技术学院 南京 211106)

**摘要** 人脸聚类将属于同一身份的人脸图片自动分到同一组,可用于人脸标注和图像管理等领域。传统的聚类算法的准确率很高,其召回率往往较低。为了解决这个问题,文中提出了一种加入三角约束关系和上下文约束的聚类算法。该聚类算法基于条件随机场模型,结合三角约束关系,同时考虑图像中常见的上下文约束,分别在聚类迭代过程和初步聚类之后添加最大相似度约束和共同出现约束,并对簇类进行整合。实验结果显示,结合上下文约束的条件随机场模型可以快速进行人脸聚类,同时还能保证在较高准确率的基础上有较高的召回率,从而提高整体的聚类效果。

**关键词** 人脸聚类,条件随机场,三角约束关系,上下文约束

**中图分类号** TP391.4 **文献标识码** A

## Face Clustering Algorithm Based on Context Constraints

LUO Heng-li WANG Wen-bo GE Hong-kong

(School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

**Abstract** Face clustering which aims to automatically divide face images of the same identity into the same cluster, can be applied in a wide range of applications such as face annotation, image management, etc. The traditional face clustering algorithms can achieve good precision, but low recall. To handle this issue, this paper proposed a novel clustering algorithm with triangular constraints and context constraints. The proposed algorithm based on conditional random field model takes triangular constraints as well as common context constraints into account in images. During the clustering iteration and after preliminary clustering, maximum similarity and people co-occurrence constraints are considered to merge the initial clusters. Experimental results reveal that the proposed face clustering algorithm can group faces efficiently, and improve recall with the high precision, and accordingly enhance the overall clustering performance.

**Keywords** Face clustering, Conditional random field, Triangular constraints, Context constraints

## 1 引言

近年来,手机、摄像机等移动拍摄设备日益成为人们生活中不可缺少的一部分,人们越来越喜欢用图片来记录和分享生活中的喜闻乐见,图片的数量呈现出井喷式增长,其中未添加标签的人脸图片占大多数。

人脸聚类是计算机视觉领域中的热门研究话题。人脸聚类指对数据库中的人脸图片进行分组,将较高相似度的人脸图片分到同一个簇,而相似度较小的图片分到不同簇中。通过人脸聚类算法自动对人脸进行标注,可以避免大量重复性的工作,节省了大量的精力和体力。此外,人脸聚类还可以分析视频结构、创建人脸数据库、分析电影语义等,例如 Wretosta<sup>[1]</sup>利用 HSV 模型分析电影语义,文献[2]使用 HCRF 模型对视频中的人脸进行命名。

人脸聚类中的方法是无监督的,主要通过学习数据样本之间合适的距离度量<sup>[3-5]</sup>,来将图片分到对应的组别中。为了更好地对图片聚类,研究重点包括两个方面:1)设计更能挖掘深层特征的模型,增强模型处理噪音的鲁棒性,减少噪音对聚类的影响,因为人脸姿势变化、光照、遮挡等不可控环境因素在一定程度上导致人脸无法正常检测与识别;2)设计新的聚

类算法来提高聚类性能。

人脸检测和识别技术越来越成熟,已经广泛地应用于人们的日常生活。而比较常见的人工特征提取方法有 PCA<sup>[6]</sup>, LBP<sup>[7]</sup>, HOG<sup>[8]</sup>等。PCA 将人脸图片作为一个整体来处理,在满足均方误差最小的情况下找到最佳线性投影。LBP 提取人脸局部区域的纹理特征,并使用 LBP 特征谱的统计直方图来描述人脸特征。HOG 算子则是基于局部区域来统计和计算梯度方向信息。这些传统方法操作简便,易于实现,但是不易于处理表情变化丰富和姿态角度转变过大等情况,同时容易忽略图像中的局部细微特征和类别属性。近年来深度学习发展迅速,搭建和训练神经网络能够更好地提取人脸特征,因此卷积神经网络框架(如 DeepID<sup>[9]</sup>, OpenFace<sup>[10]</sup>等)越来越多地被应用于人脸识别领域中。相对于传统方法,神经网络可以提取图片中更深层次的特征,从而能够更好地表示人脸,将不同的人脸区别开来。本文使用训练好的人脸五点定位网络模型,通过该模型对人脸进行五点定位,根据五点位置信息进行特征提取。

聚类算法在人脸聚类过程中也有着举足轻重的作用,基于有效人脸特征的聚类算法有很多,而最为常见的是 K-means<sup>[11]</sup>算法。该算法迭代简单且迅速,算法时间复杂度

接近于线性,特别适用于挖掘规模较大的数据集。但是其也存在一定的缺点:其参数  $k$  值的大小需要预先指定,而且聚类中心的初始化对聚类效果有很大的影响。AP<sup>[12]</sup> (Affinity Propagation Clustering)算法不需要指定聚类的数量,通过吸引力和归属感传递更新节点之间的消息,相对于 K-means 算法在聚类性能和效率上都有所提升。这个方法的不足之处在于:在每次迭代的过程中需要更新每个数据点的信息,这样就导致了算法复杂度较高,运行时间较长。还有一些现阶段比较常用的方法,如基于谱图理论的聚类算法<sup>[13-14]</sup>、基于密度的聚类算法<sup>[15]</sup>等。

传统的聚类算法将人脸图片分组后,通过聚类评价方法发现:当聚类算法的参数值过高时,准确率往往很高,但是召回率极低;当模型的参数值过小时,属于不同身份的人脸照片容易被分到同一簇中。针对这种问题,本文基于相关的约束条件,结合条件随机场模型提出了一种新的聚类算法。人脸照片中的约束条件一般分为两种:1)基于人脸特征的约束条件,如年龄、性别等人脸固有属性,这类约束条件虽然可以保证聚类的准确性,但是其提取过程和步骤比较复杂和困难,并且在提取的过程中也有一定的误差;2)基于上下文的约束条件,这些约束条件考虑得更多的是人脸照片中的上下文信息,而不是仅仅限定于人脸表征特征。例如,经常一起拍照的两个人,在另一张照片中若是知道其中一人,那么照片中属于另一个人的概率就很大,这就是共同出现 (People Co-occurrence) 约束条件;在所有的图片中,若这两张人脸之间的相似度最大,那么这两张人脸属于同一身份的可能性较高,这就是最大相似度约束;若 3 个人脸中每两张人脸之间的相似度很大,那么这 3 张人脸最有可能属于同一个人,这就是三角约束条件。本文提出的算法基于条件随机场模型,结合三角约束条件对模型进行改进,在算法迭代的过程中或初步聚类后添加共同出现约束条件和最大相似度约束。

算法中的约束条件是基于一幅图像中的上下文,而不是每张人脸中的特征信息。本文中的最大相似度约束和共同出现约束比较容易提取和获得,对初始聚类添加这两种约束条件进行迭代整合可以加快聚类速度,同时使得聚类的准确率和召回率都较高,从而提高整体聚类性能。

本文第 2 节主要介绍本文算法中的相关模型,详细介绍了算法的迭代过程;第 3 节主要介绍实验中的数据、实验过程以及实验结果的对比分析;最后总结全文并展望未来。

## 2 基于上下文约束的人脸聚类算法

### 2.1 上下文约束

人脸图片中有多种上下文约束,如人们的衣服颜色和纹理、照片拍摄的地点与时间等。这些约束条件的提取过程比较复杂和困难,本文使用的三角约束和最大相似度是基于图片的相似度矩阵,共同出现约束则是基于人们拍照的习惯和活动范围,而这些约束条件都是易于挖掘和发现的。

相似度矩阵中的元素表示数据库中每两张图片之间的相似度,在矩阵中找到较大的值,这样就可找到满足最大相似度约束条件的图片对。接下来介绍三角约束关系,用  $X_i, X_j, X_k (i \neq j \neq k)$  表示从数据库中随机选取出的 3 张人脸图片,  $S_{ij}$  表示  $X_i$  和  $X_j$  是否分到同一簇中,1 表示属于同一簇,而 0 表示不属于同一簇。三角约束关系可以表示为:

$$W(S_{ij}, S_{ik}, S_{jk}) = (1 - S_{ij})S_{ik}S_{jk} + S_{ij}(1 - S_{ik})S_{jk} + S_{ij}S_{ik}(1 - S_{jk}) \quad (1)$$

其中,  $W$  表示 3 张图片是否能分到同一簇,若值为 1 则表示这 3 张照片不能分配到同一簇中,若值为 0 则表示 3 张照片能分配到同一簇中,这就是三角约束关系。

这里的共同出现约束条件是基于一群人之间的联系程度,若两个人经常照相,且在一张照片中知道其中一个人的身份,那么照片中是另外一个人的概率就很大,这样就可以很好地将簇类进行整合,同时还不会降低准确率。假设有两个单簇,分别记为  $C_i$  和  $C_j$ ,这两个簇类之间的约束条件为:

$$Link(C_i, C_j) = \sum \min(N_{ik}, N_{jk}) \quad (2)$$

其中,  $N_{ik}$  表示  $C_i$  和  $C_k$  相互连接的次数。  $C_k$  表示簇类集,它可以是一个簇,也可以是多个簇类的集合。

对于  $C_k$ ,我们可以这样表示:  $I_m, I_n$  表示图片数据集中的两张图片,而  $C_{I_m}$  表示图片  $I_m$  中包含的簇类的类别。

$$\exists I_m, I_n, \text{ s. t. } C_{I_m} = (C_i \cup C_k) \wedge C_{I_n} = (C_j \cup C_k) \quad (3)$$

当一张照片  $I_m$  中的所有簇类可以由  $C_i$  和  $C_k$  来表示时,  $C_k$  还要满足:  $C_i \cap C_k = \emptyset$ 。同理,对于照片  $I_n$  也是如此。当式(3)成立时,式(2)中选择的  $C_k$  是合理的。

根据式(2),我们得到了每两个簇类之间的联系值,联系值越大,这两个簇属于同一个人的可能性就越大,因此根据簇类之间的联系值就可以对簇类进行合并。

### 2.2 算法模型

本文提出采用三角约束改进条件随机场模型,同时在迭代过程中或初步聚类后添加最大相似度约束条件和共同出现约束条件。虽然 Shi 等提出的 ConPaC<sup>[16]</sup>算法使用条件随机场模型结合三角约束,但是经过聚类后发现:其准确率极高,但是召回率却不是很好。针对这种问题,该论文从整体聚类效果出发,在聚类的过程中添加常见的上下文约束。

首先介绍结合三角约束关系后的条件随机场模型。如图 1 所示,图中每个圆形节点表示每张人脸图片,每个矩形节点表示函数关系。提取每张人脸中的特征并使用向量表示,使用  $\varphi_r$  函数度量两张人脸之间的相似度,函数  $\varphi_r$  表示三角约束,这样就构成了整体的模型框架。

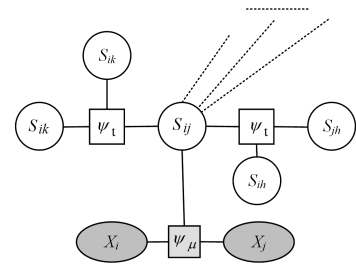


图 1 模型框架

$$p(S|X) = \frac{1}{Z} \prod_{i < j} \phi_\mu(S_{ij}) \prod_{i < j < k} \phi_r(S_{ij}, S_{ik}, S_{jk}) \quad (4)$$

在改进的条件随机场模型基础上,采用 BP(Belief Propagation)算法进行迭代。在每一次迭代的过程中,结合三角约束关系传播每个节点的信息值,从而更新每个节点的状态,直到整体的信息不再发生改变为止。

$$V_{ij}(S_{ij}) = \sum_{S_{ik}, S_{jk}} \min(V_{ik}(S_{ik}) + V_{jk}(S_{jk})) + \gamma W(S_{ij}, S_{ik}, S_{jk}) \quad (5)$$

式(5)表示每一次迭代中信息的传播过程。参数  $r$  表示

超参数,其值一般设置得很大,因为当三角关系值为1时,不能满足最小值条件,这样就可以利用三角约束条件删掉一些节点。式(5)中  $V_{ij}(S_{ij})$  表示节点  $i$  和节点  $j$  (也就是图片  $i$  和图片  $j$ ) 之间的信息值,通过两个图片之间的相似度进行初始化,而每一次迭代后的节点状态用作下一次迭代的初始化。

首先介绍本文的 IConPaC 算法,该算法选择使用式(5)迭代后添加最大相似度约束条件和共同出现约束条件。IConPaC 算法如算法 1 所示。

#### 算法 1 IConPaC 算法

```

Step1 计算相似度矩阵 S, 初始化  $V^0$ 
Step2 更新  $V_{ij}(S_{ij}) = \sum_{S_{ik}, S_{jk}} \min(V_{ik}(S_{ik}) + V_{jk}(S_{jk})) + \gamma W(S_{ij}, S_{ik}, S_{jk})$ 
Step3 初始聚类  $C_1, C_2, \dots, C_N$ 
Step4 for  $i=1:T$ 
    Link( $C_i, C_j$ ) =  $\sum \min(N_{ik}, N_{jk})$ 
    if  $\text{Link}_{\text{Min}} S_{ij} > \tau_i$ 
         $M_{C_i, C_j} = 1$ 
        if  $S_{ab} = \max(S)$ 
             $M_{C_i, C_j} = 1$ 
        update  $C_1, C_2, \dots, C_N$ 
    end

```

IConPaC 算法中  $V^0 = -\log \varphi_{\mu}(S_{ij})$ , Step2 中根据三角约束关系传递信息值,这样就可以得到每一次迭代后的簇  $C_1, C_2, \dots, C_N$ 。设置合适的阈值  $\tau_i$ , 根据共同出现约束选择相关的簇进行合并,同时根据最大相似度约束整合簇。其中,  $M_{C_i, C_j}$  值为 1 时表示  $C_i, C_j$  可以合并。

IConPaC 算法是在迭代的过程添加相关的约束条件,但是也可以在整个迭代过程结束,也就是重复执行步骤 Step2 直到聚类结束后,然后添加相关约束条件,这样就得到了本文提出的另一种算法 OConPaC。该算法与 IConPaC 算法的不同之处是添加约束的位置不同,但是使用约束条件的步骤和方法相同,因此如何使用约束条件的过程就不再详细介绍。

## 3 实验

### 3.1 实验数据

本文采用的数据库有两个:1)家庭照片集<sup>[17]</sup>,该数据集记录家庭成员之间的合照,总共包含 32 个人,随机选择 600 张图片作为测试集,一共有 1200 张人脸;2)从电视剧《武林外传》视频中得到的数据集,每 5 min 截取一张图片,共截取 300 张图片,包含 16 个人,一共有 700 张人脸。

### 3.2 人脸检测与识别

本文采用 SeetaFace<sup>[18]</sup> 人脸识别引擎对图片进行处理,使用五点定位法检测人脸,提取检测之后的人脸并对其进行矫正,使用 2048 维的向量表示每一张人脸图片的特征,这样就可以形成相应的向量矩阵。

使用距离度量方法处理特征向量矩阵,这样就可以得到数据库中每两张图片之间的相似度,本文采用夹角余弦公式  $\cos$  函数量化图片之间的相似性。

### 3.3 聚类效果评价

本文用来评价聚类效果的标准是 BCubed F-measure<sup>[19]</sup> 方法。该方法的评价标准主要有 3 个方面,分别是准确率(记为 Precision)、召回率(记为 Recall)以及将准确率和召回率相结合的 F-measure。对于任意两张人脸图片  $i$  和  $j$ ,  $C(i)$  表示

图片  $i$  经过聚类之后所在的簇别,  $L(i)$  则表示图片  $i$  应该被分到的簇别。

$$\text{Correc}(i, j) = \begin{cases} 1, & \text{if } L(i) = L(j) \wedge C(i) = C(j) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

根据式(6),准确率、召回率以及 F-measure 的定义如下:

$$\text{Precision} = \text{Avg}_i [\text{Avg}_{j \in C(i)} [\text{Correc}(i, j)]] \quad (7)$$

$$\text{Recall} = \text{Avg}_i [\text{Avg}_{j \in L(i)} [\text{Correc}(i, j)]] \quad (8)$$

$$F = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

### 3.4 实验结果与分析

根据相似度矩阵提取最大相似度约束条件,找出并记录相似度比较大的图片对。同时利用相似度中的三角约束改进条件随机场,搭建整体聚类模型框架。在聚类迭代过程中,提取人脸图片之间的共同出现约束条件,并结合最大相似度约束对簇进行整合。同时,还可以在初步聚类后添加相应的约束条件,以此提高整体的聚类效果。

图 2 和图 3 分别给出了针对家庭合照数据集和视频数据集的聚类效果。图 2 和图 3 用参数的阈值表示坐标系的横轴,纵轴表示经过 F-Measure 方法计算得到的 F 值。图 4 给出了 IConPaC 算法与 ConPaC 算法对家庭合照数据集聚类性能的比较结果,这里用 P-R 图表示。其中,横轴表示召回率,纵轴表示准确率。

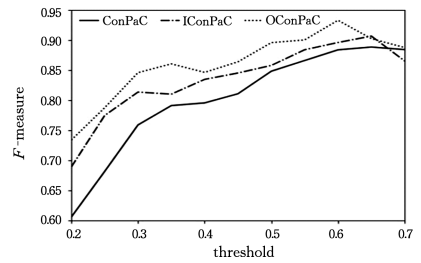


图 2 聚类算法的 F 值(家庭合照)

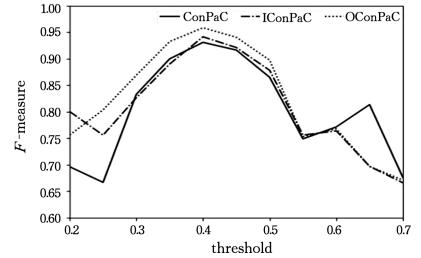


图 3 聚类算法的 F 值(视频)

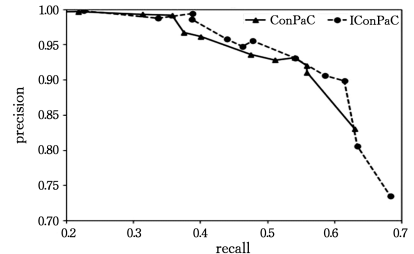


图 4 聚类算法的 P-R 图(家庭合照)

如图 2 和图 3 所示,结合上下文约束的模型比之前的模型的整体聚类效果更好。同时可以发现, OConPaC 算法中的聚类效果相对于 ConPaC 算法有显著提高。

在图 2 所示的家庭照片集中,对于 IConPaC 算法来说,相对于 ConPaC 算法没有提升很多。但是通过图 4 所示的 P-R

图可以发现, IConPaC 算法与坐标轴包围的面积大于 ConPaC 算法, 从而证明了 IConPaC 算法的性能优于 ConPaC 算法。

从表 1 中可以发现, OConPaC 算法的聚类效果明显高于 ConPaC 算法, 在家庭合照和视频数据库上分别提升了 4.5% 和 5.4%。IConPaC 算法的效果虽然没有 OConPaC 算法的效果明显, 但是相对于 ConPaC 算法也提高了不少。由此可以看出, 考虑上下文中的约束条件对聚类效果有一定的提升。

表 1 不同数据集的 F 值

算法	家庭电子照片集	视频数据库
ConPaC	0.888	0.932
IConPaC	0.906	0.944
OConPaC	0.933	0.986

**结束语** 人脸聚类在科研和生活中应用广泛, 相关的聚类算法研究已经获得了越来越多的关注。常见的人脸聚类算法普遍存在一个问题, 即聚类的准确率很高, 但是召回率往往不能满足需求。本文采用条件随机场模型, 结合三角约束关系和上下文约束(最大相似度约束以及共同出现约束)对模型进行改进。根据添加约束条件的位置不同, 本文分别提出了 IConPaC 算法和 OConPaC 算法, 实验表明使用这两种算法的聚类效果在整体上有明显提升。

### 参 考 文 献

- [1] VRETOS N, SOLACHIDIS V, PITAS I. A mutual information based face clustering algorithm for movie content analysis[J]. *Image & Vision Computing*, 2011, 29(10): 693-705.
- [2] ZHANG Y, TANG Z, WU B, et al. A Coupled Hidden Conditional Random Field Model for Simultaneous Face Clustering and Naming in Videos[J]. *IEEE Transactions on Image Processing* A Publication of the IEEE Signal Processing Society, 2016, 25(12): 5780-5792.
- [3] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015: 815-823.
- [4] ZHANG Z P, LUO P, CHEN C L, et al. Joint face representation adaptation and clustering in videos[C]// *European Conference on Computer Vision*. Springer, 2016: 236-251.
- [5] SUN Y, CHEN Y H, WANG X G, et al. Deep learning face representation by joint identification/verification[C]// *Advances in Neural Information Processing Systems*. 2014: 1988-1996.
- [6] TURK M, PENTLAND A. Eigenfaces for recognition[J]. *J Cogn Neurosci*, 1991, 3(1): 71-86.
- [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] TAN H, YANG B, MA Z. Face recognition based on the fusion of global and local HOG features of face images[J]. *IET Computer Vision*, 2014, 8(3): 224-234.
- [9] SUN Y, WMG X, TANG X. Deep learning face representation from predicting 10,000 classes[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014: 1891-1898.
- [10] <http://cmusatyalab.github.io/openface>.
- [11] MACQUEEN J. Some Methods for Classification and Analysis of Multi-Variate Observations[C]// *Proc of Berkeley Symposium on Mathematical Statistics & Probability*. 1965.
- [12] FREY B J, DUECK D. Clustering by Passing Messages Between Data Points[J]. *Science*, 2007, 315(5814): 972-976.
- [13] SHI J, MALIK J. Normalized cuts and image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 888-905.
- [14] NG A Y, JORDAN M I, WEISS Y. Y. On spectral clustering: analysis and an algorithm[J]. *Proc Nips*, 2001, 14: 849-856.
- [15] RODRIGUEZ A, LAIO A. Machine learning, Clustering by fast search and find of density peaks[J]. *Science*, 2014, 344(6191): 1492.
- [16] SHI Y, OTTO C, JAIN A K. Face Clustering: Representation and Pairwise Constraints[J]. *IEEE Transactions on Information Forensics & Security*, 2017, PP(99): 1-1.
- [17] GALLAGHER A C, CHEN T. Clothing cosegmentation for recognizing people[C]// *IEEE Conference on Computer Vision and Pattern Recognition*, 2008(CVPR 2008). IEEE, 2008.
- [18] <https://github.com/seetaface/SeetaFaceEngine>.
- [19] AMIGÓ E, GONZALO J, ARTILES J, et al. A comparison of extrinsic clustering evaluation metrics based on formal constraints[J]. *Information Retrieval*, 2009, 12(4): 461-486.
- [20] ZHANG L, KALASHNIKOV D V, MEHROTRA S. A unified framework for context assisted face clustering[C]// *Acm Conference on International Conference on Multimedia Retrieval*. 2013.