

基于主动学习的人脸标注研究

孙 金 陈若煜 罗恒利

(南京航空航天大学计算机科学与技术学院 南京 211106)

摘 要 在大数据时代,图片数量非常巨大,但是具有标签的图片非常少。在学习和研究中,常常需要分类标注图片,而大部分图片都是与人脸相关的,因此人脸标注成为了一种进行图片分类标注的有效方法,但人工标注的成本较大。针对有标签图片数量较少以及人工标注成本较大的问题,提出了在主动学习算法的基础上建立计算人脸类标签后验分布的判别模型的方法。该方法基于马尔可夫随机场和高斯过程,考虑到了样本位置、特征的客观联系,在样本之间加入了匹配约束和非匹配约束,匹配约束表示样本之间具有相同的类标签,非匹配约束表示样本之间具有不同的类标签。实验结果表明,根据判别模型得到的类标签后验分布选择样本进行人工标注,大大提高了分类器的精确度。

关键词 主动学习,匹配约束,非匹配约束,人脸标注

中图法分类号 TP391.4 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.09.050

Research on Face Tagging Based on Active Learning

SUN Jin CHEN Ruo-yu LUO Heng-li

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

Abstract In the era of big data, tremendous images are available, whereas images with tags are sparse relatively. For the purpose of learning and research, it's necessary to classify and annotate images, and most images are relevant to faces, consequently face tagging is an effective tool to annotate images. However, the cost of manual annotation is high. Aiming at solving the problems of lacking tagged images and high manual annotation cost, a discriminative model based on the active learning inducing the posterior distribution over labels was proposed. The discriminative model is based on markov random field(MRF) and gaussian process(GP), and considers the objective connections between the positions and features of samples with the addition of match constraint and non-match constraint between samples. Match constraint means that samples have the same label, while non-match constraint means that samples have different labels. Experimental results indicate that choosing samples for manual annotation according to the posterior distribution over labels induced by the discriminative model can greatly improve the classification accuracy.

Keywords Active learning, Match constraint, Non-match constraint, Face tagging

1 引言

人脸标注是分类整理图片的一个非常重要且有效的方法,所谓人脸标注就是根据图片中人脸的身份来给人脸添加标签,并以此作为图片分类的依据。在日常生活中,我们经常需要对家庭数字相册进行分类整理,进行人脸标注是一种有效的方法。另一方面,在机器学习、人脸识别等学术研究中,一些模型的建立需要将大量已标注的样本作为训练数据。例如,在图片检索中^[1-2]需要用这些样本来进行实验验证,包括最近在深度学习领域取得的重大突破也依赖于大量的训练数据^[3]。

关于人脸标注,刘胜宇等^[4]曾提出基于 AP 聚类算法对人脸样本进行分类,然后根据分类结果对每一个簇进行人脸标注。郑士鹏等^[5]提出对提取的人脸特征进行主成分分析,

然后根据欧氏距离计算人脸之间的相似度,最后根据人脸之间的相似度进行人脸标注分类。这类方法的不足之处在于,分类器的正确率和分类率比较低,得到的效果不理想。

当对分类器进行训练时,由于训练样本较少(具有标签的样本较少),分类器的正确率和分类率达不到很好的效果,不能满足需求,这时需要对样本进行人工标注,然后将其加入训练集中,从而优化分类器,使分类器达到要求。然而人工标注的成本非常昂贵,如何选择样本进行人工标注非常关键。本文采用主动学习的方法,该方法通过一定的算法,选择对分类器最有用的样本进行人工标注,从而提高分类器的精确度。

在主动学习领域中提出了较多的判别模型,包括支持向量机模型(SVM)^[2,6]、高斯过程模型^[7-8]、马尔可夫随机场模型^[9-11]、贝叶斯深度学习模型^[3]等。Gu 等^[12]基于一个高斯随机模型,将主动学习和半监督学习相结合。Zhao 等^[13]则考虑

人工标注可能会出现错误,提出了重复标注的想法。Huang等^[14]基于SVM的模型将两种选择标准(信息量、不确定性)相结合。

上述模型是机器学习研究领域的一些经典模型,其中有一些判别模型在这些传统模型的基础上进行了改进,但是这些模型没有考虑样本之间的约束。例如,在同一张照片上出现的人脸肯定来自不同的人,即具有不同的类标签,而视频中来自一个图片通道的人脸肯定来自同一个人。本文提出了一个基于马尔可夫随机场和高斯过程的主动学习模型,并在这两个模型的基础上加入了约束。所加入的约束分为两大类,即匹配约束和非匹配约束。匹配约束表示样本之间具有相同的类标签,匹配约束又可以分为先验匹配约束和聚类约束。我们认为视频中来自同一个图片通道的人脸具有先验匹配约束,而聚类约束是先对人脸样本进行聚类,来自同一个簇的人脸具有相同的类标签,即具有聚类约束。非匹配约束表示样本之间具有不同的标签,我们认为同一张图片上出现的多个人脸之间具有非匹配约束。

不同于以往基于传统模型进行改进的模型,本文提出的模型考虑到了样本之间的关系,并且将加入的聚类约束运用到了无监督学习,这大大降低了人力成本的投入。

本文提出的基于主动学习的推测类标签后验分布的判别模型将在第2部分进行具体的阐述。2.1节阐述了约束具体如何生成,2.2节阐述了在该模型中样本之间的信息如何传递,2.3节阐述了模型得到的类标签的后验分布在主动学习标准中的使用。第3部分主要是对实验数据、实验过程以及实验结果进行阐述。最后总结全文。

2 具有约束的主动学习

给定一组人脸图片的集合 $C = \{c_i\}$ 。将该集合分为两个子集合:1)具有类标签的图片集合 C_L , 标签为 $x_L = \{x_i | i \in L\}$; 2)无类标签的图片集合 C_U 。一开始训练集的数量为0,随着样本的标注,训练集的数量将不断扩大。本文的目标是用尽可能少的标签标记这些图片,并且使整个人脸图片集 C 上的图片分类率和正确率尽可能的大。

本文提出了具有约束的判别模型,这些约束分为匹配约束(MC)、非匹配约束(NMC)、聚类约束、先验匹配约束(PMC)。匹配约束为两个人脸图片具有相同的标签;非匹配约束为两个人脸图片一定具有不同的标签;聚类约束表示依据聚类算法将人脸数据集分为不同的簇,来自同一个簇的人脸图片具有相同的类标签;先验匹配约束表示在一定约束条件下不同的人脸具有相同的类标签,例如,在一定的时间内穿着同样衣服的人脸来自同一个人。聚类约束与先验匹配约束包含于匹配约束。

$$NMC = \{(x_i, x_j) | x_i \neq x_j\}$$

$$MC = \{(x_i, x_j) | x_i = x_j\}$$

$$CMC = \{(x_i, x_j) | x_i = x_j\}$$

$$PMC = \{(x_i, x_j) | x_i = x_j\}$$

其中, $PMC, CMC \in MC$ 。

2.1 约束生成

先验匹配约束和非匹配约束更容易得到,这两类约束与人脸在图片中的位置有关,因此在图片预处理(人脸区域检

测)时就可以得到这两类约束,而聚类约束需要进行进一步的处理才能得到,因此这里对先验匹配约束和非匹配约束不做详细说明,具体的产生过程将在实验部分说明。

聚类约束的生成涉及到聚类算法的选择,对后面预测类标签的后验分布有很大的影响。对于聚类的结果,本文主要有两个目标,即正确率尽可能高以及簇的大小尽可能小。

本文采用的是AP(Affinity Propagation)聚类算法,该算法不指定聚类的数目,所有的数据点都是潜在的聚类中心,称为Exemplar。 $S(k, k)$ 表示 k 数据点作为聚类中心的评判标准,值越大,则该数据点作为聚类中心的可能性越大。该值又称为参考度 p ,代表了簇之间样本的聚合程度, p 越大,样本越聚合,聚类产生的簇数量越多,反之亦然。

AP聚类算法基于样本点之间的信息进行传递,这些信息主要分为3类:1)样本之间的相似度,记为 S ,这里将样本之间的欧氏距离作为样本之间的相似度;2)吸引度信息,记为 $R, r(i, k)$ 表示样本 k 作为样本 i 的聚类中心的适合程度;3)归属度信息,记为 $B, b(i, k)$ 表示样本 i 选择样本 k 作为其聚类中心的适合程度。初始时, R 和 B 为0矩阵,通过样本之间的信息传递,不断更新 R 和 B 两个矩阵。

R 矩阵的更新过程如下:

$$r_{t+1}(i, k) = s(i, k) - \max_{k' \neq k} \{b_t(i, k') + s(i, k')\} \quad (1)$$

B 矩阵的更新过程如下:

$i \neq k$ 时,

$$b_{t+1}(i, k) = \min(0, r_t(k, k) + \sum_{i' \neq (i, k)} \max\{0, r_t(i', k)\}) \quad (2)$$

$i = k$ 时,

$$b_{t+1}(k, k) = \sum_{i' \neq k} \max\{0, r_t(i', k)\} \quad (3)$$

对每个样本点的吸引度和归属度信息求和,根据得到的信息选择聚类中心。

2.2 判别模型

首先提出一个马尔可夫随机场模型网络,在该网络中每一个预测都会受到相邻点预测的影响,所提模型如图1所示。图1中每个矩形节点表示一个类标签 $x_i (i \in L \cup U)$, 阴影节点为已经标注的标签,其余节点为还未标注的标签,相邻的两个节点所连接的边为得到的约束边(匹配约束和非匹配约束),其中实线表示匹配约束,虚线表示非匹配约束,将匹配约束和非匹配约束定义为两个集合 δ^+ 和 δ^- 。圆形节点 y_i 表示第 i 个样本点的本地势能,这是由先验高斯过程得到的,其值应该与相似的点 y_i 的值相近。

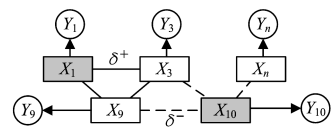


图1 马尔可夫随机场

Fig.1 Markov random field

在马尔可夫随机场概率图模型的基础上,采用LBP(Loopy Belief Propagation)算法,利用节点之间的信息传递更新MRF的标记状态。经过多次迭代之后,节点的信度收敛,不再发生变化,此时的节点即为最优标记。

$$b_i(x_i) = \frac{1}{z_i} \mathcal{O}_i(x_i, y_i) \prod_{j \in N(i)} m_{ji}(x_i) \times \prod_{(i, j) \in \delta^+} \mu^+(x_i, x_j) \prod_{(i, j) \in \delta^-} \mu^-(x_i, x_j) \quad (4)$$

其中, $b_i(x_i)$ 表示节点 i 的置信度; $m_{ji}(x_i)$ 表示节点 j 传递给节点 i 的消息, 说明节点 j 对节点 i 的影响; $O_i(x_i, y_i)$ 表示节点 i 的本地势能; $\mu^+(x_i, x_j)$ 和 $\mu^-(x_i, x_j)$ 分别表示匹配约束和非匹配约束; z_i 为归一化常量。

更新过程需要计算每一次迭代的本地势能 $O(M^k, x_u)$, M^k 为高斯过程均值向量。

$$M^{k+1} = K(K + \sigma^2 I)^{-1} \left[\begin{array}{c} \overline{x_L} \\ q^{k+1}(x_u) \end{array} \right] \quad (5)$$

其中, $\overline{x_L}$ 表示一个指示矩阵, 在该指示矩阵中, 第 (i, j) 个元素的值为: 当 c_i 标记为标签 j 时为 1, 否则为 0。 K 为核矩阵, $K_{ij} = k(c_i, c_j)$, 表示 c_i 与 c_j 之间的相似度。 $q^{BP}(x_i)$ ($i \in U$) 表示边际概率, $q(x_u)$ 表示未标注标签的联合分布概率, 边际概率可以在置信度传播中得到, 由此可以得到未标注标签的联合分布概率, 即:

$$q^{k+1}(x_u) = \prod_{i \in U} q^{BP}(x_i) \quad (6)$$

马尔可夫随机场中节点状态的更新, 以及类标签后验分布的计算流程如算法 1 所示。

算法 1 推测未标注的标签

step1 计算核矩阵 K , 指示矩阵 $\overline{x_L}$, 匹配约束 δ^+ 和非匹配约束 δ^- , 并计算常数项 $A = K(K + \sigma^2 I)^{-1}$

step2 初始化 $q^0(x_u) = 0$, $M^0 = K(K + \sigma^2 I)^{-1} [\overline{x_L}; q^0(x_u)] = A[\overline{x_L}; q^0(x_u)]$

step3 for $i=0$:MAXIteration

更新 $q(x_u)$:

$$q^{k+1}(x_u) = \prod_{i \in U} q^{BP}(x_i)$$

更新 M :

$$M^{k+1} = A[\overline{x_L}; q^{k+1}(x_u)]$$

end

最终得到 $q(x_u)$

2.3 主动学习标准的选择

本文选择的主动学习标准具有不确定性, 即无法确定某些样本分类器属于何类。基于计算出的未观测标签的后验分布, 根据香农熵公式, 选择不确定性最大的样本进行人工标注。

$$c^* = \arg \max_{i \in U} f(x_i) \quad (7)$$

其中, $f(\cdot) = - \sum_{c \in \text{classes}} h_c \log(h_c)$, $h_i = q^{BP}(x_i = c)$ 。

3 实验

3.1 实验数据

实验数据来自两个数据集: 1) 家庭电子照片集; 2) 电视剧集的视频。

1) 家庭照片集: 从家庭电子照片集中提取一组照片, 在这组照片中一共有 32 人, 即具有 32 个类标签, 从中随机选取了 700 张照片进行实验。

2) 电视剧集视频: 来源于一段 40 分钟的情景喜剧《爱情公寓》的视频。把每 5 帧图片作为一个通道, 来自同一通道的人脸图片具有先验匹配约束。这组数据集中包含 30 人, 即具有 30 个类标签, 从中随机选取了 700 张人脸图片进行实验。

3.2 样本处理

3.2.1 人脸区域的检测和特征提取

本文采用中科院山世光老师开发的 SeetaFace 人脸识别

引擎来进行人脸区域的检测和人脸特征的提取。人脸检测为五点检测法, 根据检测到的 5 点的坐标位置提取 2048 维的人脸特征。

3.2.2 约束生成

先验匹配约束: 产生先验匹配约束的主要是视频中的人脸图片, 这部分的约束生成主要是针对视频进行处理。首先提取一段视频中的所有图片, 然后随机选取 700 张。对于这 700 张照片, 将每 5 帧作为一个图片通道, 来自同一通道的人脸图片产生先验匹配约束。

非匹配约束: 非匹配约束表示如果两张人脸图片来自同一张图片, 那么这两张人脸图片产生非匹配约束。本文采用 SeetaFace 人脸检测引擎对人脸区域进行检测, 对每张图片进行人脸区域检测时, 如果一张图片中有多个脸, 那么在每两张人脸图片之间产生一个非匹配约束。

聚类约束: 首先利用 AP 聚类算法对人脸样本进行分类, 根据人脸之间的相似度(这里将人脸特征之间的欧氏聚类作为相似度的依据)将人脸分为不同的簇。对于聚类生成的每个簇中的所有人脸样本, 在每两个人脸样本之间添加聚类约束。

3.3 实验结果

对于聚类的实验结果, 我们观察的标准有分类的正确率和产生的聚类数量大小。

图 2 给出了 AP 聚类算法在不同的 p 值下在两个数据集上产生的聚类数量的大小。图 3 给出了 AP 聚类算法在不同的 p 值下在两个数据集上聚类的正确率的比较。

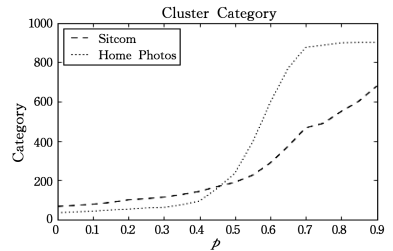


图 2 聚类的数量

Fig. 2 Number of clusters

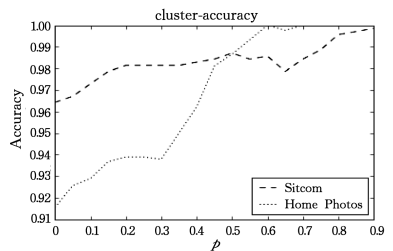


图 3 聚类的正确率

Fig. 3 Accuracy of clustering

由图 2 可知, p 值越大, 正确率越高, 但是聚类的数量随之增多。由图 3 看出, p 值为 0.5 左右及 0.8~0.9 之间时的效果较好, 但考虑到当 p 值为 0.8~0.9 之间时产生的聚类数量接近于数据集的大小, 这使得聚类结果在判别模型中的作用大大减小, 因此最终取 p 值为 0.5。

产生约束之后, 将这些约束运用到判别模型的信息传递中。图 4、图 5 分别给出了两个数据集上的实验结果。

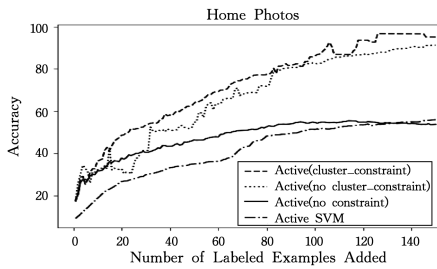


图4 家庭照片集

Fig. 4 Family photo album

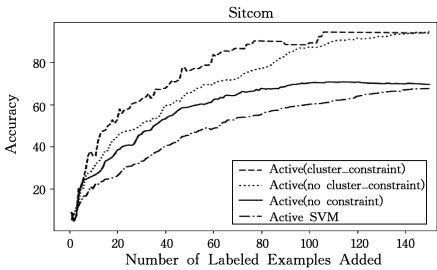


图5 电视剧集视频

Fig. 5 TV series video

对每个数据集分别进行了4次对比实验:加入聚类约束、不加聚类约束、无约束的主动学习以及经典算法——支持向量机(SVM)模型。

由图4和图5可知,本文采取的主动学习算法对人脸进行标注比经典的算法(如支持向量机SVM模型)分类器的正确率更高,在部分数据点处不加聚类约束的正确率略高于加入聚类约束的正确率,但是总体来看,加入聚类约束的效果好于不加聚类约束(只加入先验匹配约束和非匹配约束)的效果。而不加任何约束且在加入的标签较少时,与加入约束的情况相差不多,但是随着加入样本的增多,不加任何约束的效果明显下降。由此得到,在判别模型中加入样本之间的约束可以大大提高分类器分类的正确率。

结束语 人脸标注是图片分类的重要方法,基于主动学习的人脸标注是一个有效的解决训练样本不足的方法。而在主动学习领域有两大方向值得我们更深入地研究:1)在主动学习中加入无监督学习,降低专家成本;2)当数据的维度较高时,进行适当的降维操作,减少查询的复杂度^[15]。本文适当加入了聚类算法,在以后的主动学习的算法研究中,可以适当加入降维的操作,特别是当图片的特征维数较大时,才能充分发挥人脸标注在图片识别领域的作用。

参考文献

- [1] HOI S C H, LYU M R. A Semi-Supervised Active Learning Framework for Image Retrieval[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2005:302-309.
- [2] WANG L, CHAN K L, ZHANG Z. Bootstrapping SVM active learning by incorporating unlabeled images for image retrieval [C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2003: 629-634.
- [3] GAL Y, ISLAM R, GHAHRAMANI Z. Deep Bayesian Active

Learning with Image Data [J]. arXiv preprint arXiv: 1703.02910, 2017.

- [4] LIU S Y, LIU J F, HUANG Q C, et al. Research on Annotation Technology of Face based on improved AP Clustering Algorithm [J]. Intelligent Computer and Application, 2011, 1(3): 35-38. (in Chinese)
刘胜宇, 刘家锋, 黄庆成, 等. 基于改进 AP 聚类算法的人脸标注技术研究[J]. 智能计算机与应用, 2011, 1(3): 35-38.
- [5] ZHENG S P, LIU H Y, SUN F M, et al. Face Detection and Annotation Based on a Family Digital Photo Album [J]. Journal of Liaoning University of Technology (Natural Science Edition), 2016, 36(3): 160-162. (in Chinese)
郑士鹏, 刘海云, 孙福明, 等. 基于家庭数字相册的人脸检测与标注[J]. 辽宁工业大学学报(自然科学版), 2016, 36(3): 160-162.
- [6] TONG S, CHANG E. Support vector machine active learning for image retrieval [C]// ACM International Conference on Multimedia. ACM, 2001: 107-118.
- [7] RODRIGUES F, PEREIRA F C, RIBEIRO B. Gaussian process classification and active learning with multiple annotators [C]// International Conference on International Conference on Machine Learning. JMLR. org, 2014: II-433.
- [8] KAPOOR A, GRAUMAN K, URTASUN R, et al. Active Learning with Gaussian Processes for Object Categorization [C]// IEEE International Conference on Computer Vision. IEEE, 2015: 1-8.
- [9] KAPOOR A, HUA G, AKBARZADEH A, et al. Which faces to tag: Adding prior constraints into active learning [C]// IEEE International Conference on Computer Vision. IEEE, 2009: 1058-1065.
- [10] SUN S, ZHONG P, XIAO H, et al. An MRF Model-Based Active Learning Framework for the Spectral-Spatial Classification of Hyperspectral Imagery [J]. IEEE Journal of Selected Topics in Signal Processing, 2015, 9(6): 1074-1088.
- [11] HU B, MOSER G, SERPICO S B, et al. An active learning heuristic using spectral and spatial information for MRF-based classification [C]// Geoscience and Remote Sensing Symposium. IEEE, 2015.
- [12] GU Y, JIN Z, CHIU S C. Combining Active Learning and Semi-supervised Learning Using Local and Global Consistency [C]// International Conference on Neural Information Processing. Springer International Publishing, 2014: 215-222.
- [13] ZHAO L, SUKTHANKAR G, SUKTHANKAR R. Incremental Relabeling for Active Learning with Noisy Crowdsourced Annotations [C]// IEEE Third International Conference on Privacy, Security, Risk and Trust. IEEE, 2011: 728-733.
- [14] HUANG S J, JIN R, ZHOU Z H. Active learning by querying informative and representative examples [C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2010: 892-900.
- [15] LIU K, QIAN X, WANG Z Q. The summarization of active learning algorithm [J]. Computer Engineering and Application, 2012, 48(34): 1-4. (in Chinese)
刘康, 钱旭, 王自强. 主动学习算法综述 [J]. 计算机工程与应用, 2012, 48(34): 1-4.