

test title

zhangle

April 20, 2016

Abstract

在这篇文章中，我们就现有的马氏度量学习方法在可扩展性和所需要的阶层的监督提出了重要的观点。在大规模计算中通常使用冗长的迭代，这使计算变得非常困难。更进一步考虑，如果考虑到数据数量的持续增长，为所有数据指定完全被监督的标记是不可行的。而利用等价约束指定标记会更加简单些。我们介绍了一个简单却非常有效的策略，通过等价约束学习距离度量，基于数学推导。相比于现有的方法，我们的方法并不依赖于需要花费大的迭代计算的复杂优化问题。因此，我们的方法会在数量级上比现有方法快。在多种特征下多种具有挑战性的基准测试下的结果证明了我们方法的力量，包括无约束环境下的人脸识别，匹配未曾见过的物体实例以及通过空间不连续的摄像头进行行人再识别等。在后两个基准测试中，我们远远超过现有的最高水平。

1 前言

距离或相似度的度量学习是机器学习中非常重要急迫的领域，在计算机视觉中有各种各样的应用。它可以明显提高跟踪，图像检索，人脸识别，聚类或者行人再识别。度量学习算法的目标是利用更加简单更加通用的相似度方法

在很多机器学习问题中展示出非常好的普遍性的一种经典距离函数是马氏度量学习。其目标是寻找一个全局的特征空间先行转换，从而使相关的维数被强调而不相关的被丢弃。因为在马氏度量和多变量高斯函数中存在双射，我们可以想到与之相关的协方差矩阵。矩阵通过任意的线性旋转和比例变换适应所要求的几何。映射之后采用浅显易懂的欧式距离进行测量。

学习马氏度量的机器学习算法最近在计算机视觉领域吸引了大量的兴趣。包括大边缘最近邻学习 (LMNN)，信息论度量学习 (ITML)，和被认为最高水平的逻辑斯蒂判别式度量学习 (LDML)。LMNN 目标在于提高 K-nn 分类。它为每个实例建立了一个局部的周界，这个周界包围着相同标记 (目标近邻) 的 K 最近邻实例，加上一个边缘。为了减少侵入周界不同标识实例 (冒充目标) 的数量，度量方法被迭代使用。通过加强目标近邻间的联系同时减弱与冒充者的联系。概念上听起来，LMNN 有时候比较容易由于对数据缺乏规则化产生过拟合的现象。David 通过明确集成了一个数据规则化的步骤来避免过拟合。他们的公式在满足距离函数所给约束同时最小化与先前优化的距离度量区域的差异熵。Guillaumin 介绍了一个马氏距离的概率视角，在其中一个后验型种类概率被看成相同 (不同) 的方法。因此，他们的目标在于迭代地适应马氏度量从而使对数似然函数的值最大。后验概率是通过 sigmoid 函数建模的，sigmoid 函数反映了实例如果距离在某个阈值之下共享标识的事实。原则上，所有的这些方法都可以在没有见过的数据上概括得很好。他们专注在强健的损失函数，规则化解决方案从而避免过拟合。

考虑到数据数量不断增长，在大规模数据集上学习马氏度量对可扩展性和监督所需要的阶数提出了更高的要求。时常对所有数据点进行具体的完全监督标识是不可行的，而通过等价约束的形式更容易进行具体标识。在一些实例中，采用这种自动弱监督的方式会更加可能。例如跟踪一个对象。所以在大规模数据集上实施将会面对额外的等价转换和等价约束挑战。

为了满足这些要求，我们研究了一个基于等价约束的有效度量方法。他们建立点与点间基本关系，被认为是距离度量学习算法中的类似相似度函数的输入。我们的方法被一个基于可能性测试的数学推断所驱动。我们展示了尺度衡量的结果不易于过拟合而且可以非常高效地获得。和其他方法相比，我们不依赖于反复的迭代优化过程。然而，我们的方法对大数据集来说是可扩展的，因为他仅仅包括了两个小型且大小固定的协方差的计算。和 KISS 原则相似，我们是我们的方法每一步设计简单且有效所以称之为简单有效的度量方法 (KISS metric)。

我们在不同的实验样本中证明我们的方法可以达到甚至优于目前最先进的尺度度量方法，同时在训练阶段有数量级倍的提高。尤其是我们提供了在两个最近的人脸识别样本上的结果。其不明显的

特征和在姿势、灯光以及面部表情上的变化是学习算法中的一个挑战。更进一步地，我们研究了不连续空域摄像头下的行人再识别课题以及以玩具车为例对未见过对象实例的对比。在 VIPeR 和玩具车数据集中，我们水平提高到甚至超过了目前领域中最先进的水平。而且，在 LFW，我们采用标准 SIFT 特征得到了目前最好的结果。

文章的剩余部分是如下组织的。在第二部分中我们讨论了支持我们研究的相关尺度衡量算法。进一步，我们在第三部分中介绍了我们的 KISS 尺度学习方法。扩展的实验和性能评估以及扩展性评估在第四部分中展示。最后，我们在第五部分中得出结论并总结了这篇论文。

2 学习马氏距离

研究基于马氏距离函数的距离和相似性度量在计算机视觉领域得到了很多的关注。总体来讲，马氏距离衡量了两点 \mathbf{x}_i 和 \mathbf{x}_j 之间的平方距离

$$d_M^2(x_i, x_j) = (x_i - x_j)^T M (x_i - x_j)$$

其中是一个半正定矩阵，而是一对例子。另外，在接下来的讨论中，我们利用标识两个实例为相似的一对，他们共享相同的类别标识 $()$ 而且 $()$ 。为了解释我们的方法，我们下面给出目前在马氏尺度衡量最先进的算法简介。尤其是我们检验了 LMNN, ITML 和 LDML。

2.1 大边缘最近邻尺度

Weinberger 地方法目标在于通过利用周围数据的结构来提高 K-nn 分类器。对于每一个实例，一个包围着最近的拥有相同标识的 K 个邻居近邻区域被建立。拥有不同标识并且侵入这片区域的实例将会被处罚。下面这个目标函数解释了这个规则：

$$\epsilon(M) = \sum_{j \sim i} [d_M^2(x_i, x_j) + \mu \sum_l (1 - y_{il}) \xi_{ijl}(M)]$$

第一个变量使脚标为 $j \sim i$ 目标近邻对 $\mathbf{x}_i, \mathbf{x}_j$ 最小。

$$\xi_{ijl}(M) = 1 + d_M^2(x_i, x_j) - d_M^2(x_i, x_l)$$

为了估计M，梯度下降定义一个含有正项差异的三元组 (i, j, l) 使之按照梯度运行。

$$\frac{\partial \epsilon(M_t)}{\partial M^t} = \sum_{j \sim i} C_{ij} + \mu \sum_{(i, j, l)} (C_{ij} - C_{il})$$

其中表示成对间差异的外积。概念上来讲，对于比较活跃的三元组，这个公式加强了目标邻居间的联系，减弱了与侵入者的联系。

2.2 信息论度量学习

Davis 研究了多变量高斯分布和马氏距离之间的联系。思想在于寻找一个解决方案能够平衡约束的满足同时接近先验的距离度量 M_0 。例如欧式距离中的单位矩阵。解决方案的接近都通过相关联区域的 Kullback-Leibler 散度。先验可以认为是为了避免过拟合而进行的规则化的过程。约束强制使相同的点对在某一个距离而不同的点对超过某一个距离。优化建立与布雷格曼映射，它将当前的解决方案通过要给更新规则映射到一个单个约束：

$$M_{t+1} = M_t + \beta M_t C_{ij} M_t$$

参数 β 包含了点对的标识和步长的大笑。它在相同的点对中为正在不同的点对中为负。所以，对于相同的点对，优化向的方向进化而对于不同的点对则向相反的方向进化。每一对在梯度递减上的影响被概率所控制。

2.3 线性判别度量

未完成

如果我们概括一下以上所提到的尺度衡量方法的属性和特征，我们会发现两个共同点。第一，所有的方法依赖于迭代优化的过程，它在大规模数据集的情况下计算会变得花费很大。第二，如果我们比较一下给出的不同方法中的更新规则，我们可以发现对于相同的点对，优化总是向的方向进行而对于不同的点对则向相反方向。在接下来，我们介绍一个不用叠戴的公式，他可以在成对之间的差异上建立一个数学推导公式。这允许我们能够面对额外的可扩展性和等价约束能力的挑战。我们参数自由的方法在训练时非常的有效，可以继续研究持续增长的数据的尺度度量。

2.4 KISS 度量学习

我们的方法考虑到两个独立的，相同种类与否由 [1] 是否属于该种类的可信度决定。从数学推导的视角来看，最优的数据上的决策是否一对点对相同可以由一个相似度来获得。所以我们检验假说 H_0 两者不同和对立假说 H_1 两者相同：

$$\delta(x_i, y_j) = \log \left(\frac{p(x_i, x_j | H_0)}{p(x_i, x_j | H_1)} \right)$$

获得较大的值标识 H_0 是有效的。相反，较小的值意味着被假说 H_0 拒绝，点对被看做是相同的。为了使特征空间中的实际分布具有独立性，我们假设点对间的差异为 0，所以等式 9 可以被重写为

$$\delta(\mathbf{x}_{ij}) = \log \left(\frac{p(\mathbf{x}_{ij} | H_0)}{p(\mathbf{x}_{ij} | H_1)} \right) = \log \left(\frac{f(\mathbf{x}_{ij} | \theta_0)}{f(\mathbf{x}_{ij} | \theta_1)} \right)$$

其中是假说 H_1 点对相同的含有参数的概率密度函数，与假设点对不相同的假说 H_0 相对。将特征间的差异空间看作一个高斯结构我们可以简化这个问题，等式被重新写为

$$\delta(\mathbf{x}_{ij}) = \log \left(\frac{\frac{1}{\sqrt{2\pi|\Sigma_{y_{ij}=0}|}} \exp(-1/2\mathbf{x}_{ij}^T \Sigma_{y_{ij}=0}^{-1} \mathbf{x}_{ij})}{\frac{1}{\sqrt{2\pi|\Sigma_{y_{ij}=1}|}} \exp(-1/2\mathbf{x}_{ij}^T \Sigma_{y_{ij}=1}^{-1} \mathbf{x}_{ij})} \right)$$

其中

$$\Sigma_{y_{ij}=1} = \sum_{y_{ij}=1} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T$$

$$\Sigma_{y_{ij}=0} = \sum_{y_{ij}=0} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T$$

点对间的差异是对称的。所以，我们有均值为 0，且公式。高斯中最大化相似度评估等价于采用最小二乘法最小化马氏距离从。这是我们可以分别针对两个独立的集找出个字相关的方向。通过取对数，我们可以重新整理相似度的检验如下：

$$\delta(x_{ij}) = \mathbf{x}_{ij}^T \Sigma_{y_{ij}=1}^{-1} \mathbf{x}_{ij} + \log(|\Sigma_{y_{ij}=1}|) - \mathbf{x}_{ij}^T \Sigma_{y_{ij}=0}^{-1} \mathbf{x}_{ij} - \log(|\Sigma_{y_{ij}=0}|)$$

更进一步地，我们去掉常数项因为他们只是提供了一个偏移量，简化如下：

$$\delta(\mathbf{x}_{ij}) = \mathbf{x}_{ij}^T (\Sigma_{y_{ij}=1}^{-1} - \Sigma_{y_{ij}=0}^{-1}) \mathbf{x}_{ij}$$

最后，我们获得我们的可以反映了相似度检验的马氏距离度量如下

$$d_{\mathbf{M}}^2(\mathbf{x}_i - \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{M} (\mathbf{x}_i - \mathbf{x}_j)$$

通过重新映射到半正定矩阵。所以，我们通过等价裁剪的范围我们得到了。

3 实验

为了展示我们方法的应用广泛性，我们在含有多种特征的不同样本数据集上测试了我们的方法。我们实验的目标主要分为两点：第一，我们想展示我们的方法是也可以概括未知数据，甚至比目前最先进的算法效果还要好。第二，我们想要证明我们的运算速度将会产生数量级上的提高。这对大规模和在线应用都有明显的好处。在 4.1 部分，我们首先测试了无约束环境下人脸识别。接下来，在 4.2 部分，我们通过空域不连接的摄像头研究了行人再识别问题。最后，在 4.3 部分，我们在玩具车数据集上测试和评估了我们的算法，意在比较对未知物体的识别。

为了和其他尺度学习方法进行对比，我们收集了原始代码并且采用相同的输入数据。代码由各自的作者分别提供。更进一步地，我们将我们的方法和相关的区域一定的方法进行对比。并且画出了各自的 EER 曲线

3.1 人脸识别

以下部分，我们在两个具有挑战性的人脸识别数据集上证明了我们方法的性能，分别是 Labeled Faces in the Wild (LFW) 和 Public Figures Face Database (PubFig)。据此，对人脸识别的研究被分为两个不同的目标：人脸识别（识别一张人脸）和人脸验证（判断两张人脸图片是否是相同的个体）。人脸识别自身要求有大量的带有注解的个人人脸，真实世界中的数据库往往不能满足。相反，人脸验证需要较少的注释而且可以在大规模上也能实现。所以，根据以上分析，我们更加关注人脸验证课题。

3.1.1 Wild 数据集中的被标识人脸

Wild 数据集中被标识人脸共包含了 5749 个人的 13233 张无约束人脸，可以认为是当前最高水平的人脸识别样本库。这个数据库被认为非常具有挑战性，因为它在姿势、亮度、面部表情、年龄、性别、种族和总体摄影以及环境状况等方面的有多种多样的变化。在图表 2 (a) 中我们给出了一些说明性的例子。LFW 的一个重要特征是数据库中的主体在数据库的任何一个部分中都是和其他主体互相排斥的。所以，对于人脸验证课题来讲，可以根据在训练中未出现过的个体进行测试。

数据被分为十个部分以便进行交叉验证。每个部分含有 300 对相同和 300 对不同的数据。结论的分数在 10 个部分上被平均。在受约束的协议下只有对相同的组或不同的组进行等价约束变化，不能根据主体的身份进行推算。例如，取样更多的训练数据是允许的。

在我们的实验中，我们采取了 Guillaumin 的人脸表示法。这种方法主要三个不同尺度上提取了 9 个利用 SIFT 算法自动检测到的面部标志。最终的描述结果是一个 3456 维的向量。为了使它在距离尺度学习算法中可跟踪，我们利用 PCA 将之降维为 100 维的子空间。为了评估不同的尺度衡量算法，有一个公正的比较评估，我们利用相同的特征和 PCA 得出的维数训练分类器。PCA 的影响不是非常重要。利用不同的降维算法对所有测试的算法得出的结论表示在性能上没有很大的差异。但是在线性支持向量机上，我们直接对人脸描述子进行训练得出了更好的结果。

在图表 2 (b) 我们展示了 LDML, ITML, LMNN, SVM 和我们的方法 KISSME 的受试者工作特征曲线以及以欧氏距离为基线的相同对的马氏距离。需要特别说明的是在 LMNN 中我们基于实际分类信息提供了更多的有监督训练，因为它需要三元组。

相同分类样本对的马氏距离比欧式距离的性能好。马氏距离性能提高了大约 7%。有趣的是，LMNN 不能概括其他尺度上的额外信息。KISSME 比其他尺度度量方法稍微好一些。它达到了 80.5% 的等错误率，是目前这种样本类型中效果最好的。当然，最近在 LFW 上最先进 de 方法提供了更好的结果但是同时它需要了更多领域的知识，因为这些方法关注仅仅在人脸上。

当分析表格 1 中给出的训练次数是，我们方法最主要的优势也很明显。事实上，与 LMNN, ITML 和 LDML 相比，我们的方法在计算上更加高效，尽管如此，仍然可以输出非常具有竞争力的结果。

3.1.2 Public Figures 人脸图片数据库

PubFig 数据集和 LFW 有很多的共同点。他也是一个非常具有挑战性的大规模，真实的数据库，包含了 200 个人的 58797 张图片，这些图片从谷歌图片和 FlickrR 上手机。人脸验证的样本包括了 10 个交叉验证文件夹，其其中包括了 1000 对内部人员和 1000 对外部人员的信息。每个文件夹中的信息从 14 个个人中获得。和 LFW 相同，出现在测试中的人脸在训练过程中并未出现过。

在数据库中比较有趣的是一种被称为“高水平”特征被用来描述人脸视觉特征的出现和消失。容貌被用一些可以命名的属性例如性别、种族、年龄、头发等或者和面部区域相似度有关联的“微笑”等信息来编码对应确切代表的一个人。这种间接的描述方式输出的很好的属性，例如和低水平的特征相比有一定程度的鲁棒性提升。而且，它给我们在评估距离尺度学习算法的评估上也提供了补足的属性特征类型。

在图表 4 中我们画出了 LDML、ITML、LMNN、SVM 和 KISSME 的 ROC 曲线以及两条极限。可以看出我们的方法比 LDML、ITML、LMNN 表现出色，达到了目前最先进的性能——基于 Kumar 的方法的 SVM。LDML 我们的算法达到了相同的结果然而在训练时间上慢了数量级倍。这使该方法在在线计算和大规模计算中变得不现实。有趣的是，ITML 的性能下降到甚至普通欧氏距离水平以下。在图表 2 中我们也阐述了 LFW 的属性特征的性能。

3.2 行人再识别

VIPeR 数据集包括 632 对两个不同摄像头视角下在户外拍摄的个人图片。这些低分辨率的图片在姿势，视角有明显的变化，并且在亮度上也有很大的不同，例如有高亮和阴影的部分。示例中大多数的图片对包含九十度以上的视角转化，使行人再识别很具挑战性。在图 3 中给出了一些示例。为了将我们的方法和其他方法作比较，我们采用了文献【5，7】中所定义的评价协议。作者将 632 张图片集随机分成两组 316 张图片集，一组用于训练，另外一组用于测试，在多次实验之后计算平均值。由于没有预先确定图片集和如何使两组得到不同的图片对的规则。因此，我们根据随机组合不同人的图片生成不同的图片对。

为了表示图片，我们编译了一个极其简单的描述子。首先，我们将图片按照 8×16 分为 8 行 8 列的有重叠部分的块状。然后，为了描述图片的颜色信息，我们用了 HSV 和 Lab 直方图，每个通道含有 24 个颜色小区间。第三，我们利用 LBPs 捕捉纹理信息。最后，对于距离尺度学习方法，我们将连接起来的描述子通过 PCA 映射到一个 34 维的子空间。

为了展示各种算法的性能，我们展示了累计匹配特征曲线 (CMC)。它们表达了在前 n 级项正确匹配的期望。为了得到一个合理的数据，我们将 100 次运行进行平均。在图 3 中，我们展示了不同尺度学习算法的累计匹配特征曲线，并且在表 2 中展示了我们方法与目前最先进的方法【4，5，10，23】在前 50 级命中的性能对比结果。如我们可见，我们在所有级别均得到了有力的结果。我们的方法比其他方法性能更好，尽管和他们相比我们并没有使用前景-背景分离处理。与此同时，我们在计算上更加高效。

3.3 ToyCars

LEAR ToyCars 数据集包含了 14 种玩具小汽车和玩具卡车，每种 256 张图片截图。数据集在姿势，光照和凌乱的背景上有很大的不同。实验目的是根据已知种类的汽车来对之前未见过的物体实例进行对比（如图 5 所示）。所以，此实验是为了分类是否一组图片展示的是同一物体。训练数据集包含了 7 种物体的相关的 1185 张相同和 7330 张不同的图片组。剩余的 7 个物体的图片示例用作测试。由于图片在水平方向上的像素不同，我们将其补零使之具有一个规范的图像大小。

我们采用行人再识别中所使用的方法进行图像信息提取和标识。所以，图像被分为 30×30 的不重叠的块状。我们利用 HSV 和 Lab 提取颜色信息，利用 LBP 记录纹理信息。整体图片的描述子是由块状描述子组合而成。利用 PCA 将描述子映射到一个 50 维的子空间。

我们将在这个数据集上的实验结果与 Nowak 和 Jurie 最新的方法进行了对比，他们的方法基于一个极其随机的树结构的集合体。集合体通过使匹配成对图片的相关补丁输出相似的结果来量化相关补丁对的不同。相关补丁通过 NCC 定位到一个局部的邻近区域。在测试中，两张图片对的相似度是在同一叶子节点结束的相关补丁的加权和。

在图 6 中，我们对比了我们的方法、Nowark 和 Jurie 的方法和其他尺度学习方法的 ROC 曲线。然后，我们提供了一个标准线性支持向量机下的基准线。利用支持向量机输出的等错误率为 82%，是一个合乎情理的结果。有趣的是，一些尺度学习算法在欧氏距离上进行提高。只有 LMNN 和 SVM 性能相近。利用通过正确的图片学习到的马氏距离，我们达到了 89.8% 的等错误率，比 Nowak 和 Jurie 的方法性能好。KISSME 大大地将算法性能增强，提高至 93.5%。如果有人考虑 17 小时的计算时间，我们的方法再一次展示了他有效性和高效性的优势。

参考文献

- [1] Bingpeng Ma, Yu Su, and Frédéric Jurie. Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image and Vision Computing*, 32(6):379 - 390, 2014.