

基于类间和类内密度的多视角距离度量学习

任双艳, 郭威, 范昌琪, 王喆, 吴松洋

引用本文

任双艳, 郭威, 范昌琪, 王喆, 吴松洋 [基于类间和类内密度的多视角距离度量学习](#) [J]. 计算机科学, 2022, 49(11A): 211000131-6.

REN Shuang-yan, GUO Wei, FAN Chang-qi, WANG Zhe, WU Song-yang. [Multi-view Distance Metric Learning with Inter-class and Intra-class Density](#) [J]. Computer Science, 2022, 49(11A): 211000131-6.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[结合注意力机制与几何信息的特征融合框架](#)

Feature Fusion Framework Combining Attention Mechanism and Geometric Information
计算机科学, 2022, 49(5): 129-134. <https://doi.org/10.11896/jsjx.210300180>

[面向多标签小样本学习的双流重构网络](#)

Dual-stream Reconstruction Network for Multi-label and Few-shot Learning
计算机科学, 2022, 49(1): 212-218. <https://doi.org/10.11896/jsjx.201100143>

[融合频率和通道卷积注意的脑电\(EEG\)情感识别](#)

EEG Emotion Recognition Based on Frequency and Channel Convolutional Attention
计算机科学, 2021, 48(12): 312-318. <https://doi.org/10.11896/jsjx.201000141>

[基于信任度匹配的改进PBFT共识算法](#)

Improved PBFT Consensus Algorithm Based on Trust Matching
计算机科学, 2021, 48(2): 303-310. <https://doi.org/10.11896/jsjx.200500112>

[基于帧级特征的端到端说话人识别](#)

End-to-End Speaker Recognition Based on Frame-level Features
计算机科学, 2020, 47(10): 169-173. <https://doi.org/10.11896/jsjx.190800054>

基于类间和类内密度的多视角距离度量学习

任双艳¹ 郭威¹ 范昌琪² 王喆¹ 吴松洋³

¹ 华东理工大学信息科学与工程学院 上海 200237

² 上海移动互联网产业促进中心 上海 200333

³ 公安部第三研究所 上海 201204

(shuangyan_ren99@163.com)

摘要 几何信息可以为分类方法提供先验知识和直观解释。从几何角度观察样本是一种新的样本学习方法,密度则是几何信息中非常直观的表现形式。提出了基于类间和类内密度的多视角距离度量学习方法来学习一个度量空间。在这个空间内,异类样本更加分散,同类样本更加紧密。首先,在大边际框架下引入类间密度,通过最小化类间密度来约束度量空间中的样本,从而实现类间分散,提高分类性能。其次,引入类内密度,通过最大化类内密度来达到同类样本互相靠近的效果,从而实现类内紧凑。最后,为了更好地挖掘多视角样本的互补信息,最大化度量空间中各视角之间的相关性,使各视角自适应地相互学习,探索视角之间的互补信息。在真实数据集上的大量实验结果证明了该方法的优越性。

关键词: 几何信息;类间密度;类内密度;互补信息;视角相关性

中图法分类号 TP181

Multi-view Distance Metric Learning with Inter-class and Intra-class Density

REN Shuang-yan¹, GUO Wei¹, FAN Chang-qi², WANG Zhe¹ and WU Song-yang³

¹ School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China

² Shanghai Mobile Internet Industry Promotion Center, Shanghai 200333, China

³ The Third Research Institute of Ministry of Public Security, Shanghai 201204, China

Abstract Geometric information can provide prior knowledge and intuitive explanation for classification methods. Observing samples from geometric perspective is a novel method of sample learning, and density is a very intuitive form of geometric information. This paper proposes a multi-view distance metric learning method with inter-class and intra-class density to learn a metric space. In this space, the heterogeneous samples are more scattered, and the homogeneous samples are closer. First, the inter-class density is introduced under the large margin framework, and the samples in the metric space are constrained by minimizing the inter-class density, so as to realize the inter-class dispersion and improve the classification performance. Second, maximize the intra-class density to achieve the effect of similar samples close to each other, so as to achieve intra-class compactness. Finally, to better mine the complementary information of the multi-view samples, the correlation between the views in the metric space is maximized, so that the views can learn from each other adaptively and explore the complementary information among the views. A large number of experimental results on real-world datasets demonstrate the superiority of the proposed method.

Keywords Geometric information, Inter-class density, Intra-class density, Complementary information, View correlation

1 引言

几何学是研究空间的结构和支撑的一门学科^[1]。几何带给人们更多的是一种直观的象征,这种直观的解释可以帮助我们更好地概括理解目标的抽象内容化^[2]。在机器学习领域,探索样本的几何关系可以帮助我们更有效地进行分类^[3]、回归^[4]、聚类^[5]和特征提取^[6]。几何信息包括样本间的距离和方向角、样本的局部和非局部几何结构以及样本与超平面的位置关系。距离度量学习研究样本之间的距离和位置关系来提高算法性能,通常用于改进 K 近邻分类器^[7]和 K-means

聚类^[8]的分类结果。监督度量学习、半监督度量学习^[9]和非监督度量学习^[10]是基于训练样本标签可用性的 3 种距离度量学习。本文研究监督度量学习。监督度量学习是学习一个距离度量空间,在这个距离空间中,同一类样本之间尽可能接近,而不同类样本之间距离尽可能远^[11]。

最早的方法是大边界最近邻分类度量学习(LMNN)^[12],该方法在度量空间中采用较大的裕度来区分不同类别的样本。考虑到 LMNN 算法的计算复杂性,Zadeh 等提出了通过黎曼几何正定矩阵来实现具有固有几何吸引力的几何平均度量学习(GMML)^[13],该方法采用半正定逆矩阵对不同样本

基金项目:上海市科技计划项目(20511100600);国家自然科学基金(62076094);信息网络安全公安部重点实验室开放课题项目(C20603)

This work was supported by the Shanghai Science and Technology Program(20511100600), Natural Science Foundation of China(62076094) and Key Lab of Information Network Security of Ministry of Public Security(The Third Research Institute of Ministry of Public Security)(C20603).

通信作者:王喆(wangzhe@ecust.edu.cn)

进行学习,既保证了不同样本之间的距离较远,又提高了计算效率。与计算样本之间的距离不同,度量学习指导的最小二乘分类器(MLG-LSC)^[14]用样本最小二乘回归的误差来代替距离,通过最小化相似样本的误差,最大化不同样本的误差,也可以达到相似样本接近而不同样本远离的效果。

从单个视角学习一个对象往往会忽略一些重要信息,因此多视角学习逐渐流行。多视角几何平均度量学习(MvGMML)^[15]是GMML的多视角版本,它在每个视角上分别实现GMML算法,然后为每个视角设置不同的权值,以达到多样化视角的效果。在基于KL-divergence(MML-KL)的多视角度量学习中,KL-divergence被应用到多视角分类学习任务中,以获得视角之间的共识信息^[16]。Fisher-HSIC多视角度量学习(FISH-MML)^[17]结合了fisher判别分析(FDA)^[18]和Hilbert-Schmidt独立标准(HSIC)^[19],通过最小化变量之间的依赖来探索视角之间的互补性。

上述多视角度量学习方法虽然在一定程度上提高了分类性能,但仍存在一些不足。1)上述方法仅计算每个样本对之间的距离,导致度量空间中样本混沌的现象。本文在最大边际框架下考虑异类样本的类间密度以及同类样本的类内密度,通过最小化类间密度和最大化类内密度来约束度量空间中的样本,从而将不同类别的样本分散到不同的类中,使得同类样本之间更加紧凑,提高分类性能。2)上述多视角度量学习模型只研究了视角的一致性,而忽视了多视角的另一个重要原则:互补性。互补信息是探索每个视角缺失特性的关键。本文方法通过自适应学习,最大化度量空间中各视角之间的相关性,探索各视角之间的互补信息。根据每个视角的唯一性,赋予视角不同的权值,并对其进行自适应更新。本文方法在现实世界的多视角数据集上具有较显著的性能优势。

2 相关工作

距离度量学习一直是分类任务里非常重要的方法,受到很多学者的关注。有监督距离度量学习最基本的思想就是通过学习一个度量空间,在这个空间里让样本具有最大的可分性,来提高KNN-Classifier等分类器的分类精度。总体来说,其是要学习一个Mahalanobis度量矩阵,通过一个对称半正定矩阵 \mathbf{M} 来约束两个样本^[20]。Mahalanobis距离公式如下:

$$d_{\mathbf{M}} = \sqrt{(x-y)^T \mathbf{M} (x-y)} \quad (1)$$

其中, \mathbf{M} 是对称半正定矩阵,它可以被分解为 $\mathbf{M}=\mathbf{W}\mathbf{W}^T$ 并且 $\mathbf{W} \in \mathbb{R}^{D \times d}$, D 为样本特征维度, $d < D$ 代表降维后的维度。

最大边际准则是距离度量学习的通用框架。它的主要思想就是让属于不同类别的样本之间的距离尽可能大,让属于同一类别的样本尽可能相互靠近。它的实现框架如下:

$$\frac{1}{D} \sum_{(x_i, x_j) \in \mathcal{D}} d_{\mathbf{M}}^2(x_i - x_j) - \frac{1}{S} \sum_{(x_i, x_j) \in \mathcal{S}} d_{\mathbf{M}}^2(x_i - x_j) \quad (2)$$

假设 $x_i, x_j \in \mathbf{X}$,可以将样本组合为样本对 (x_i, x_j) 。根据样本对内两个样本的标签信息,可以将样本空间分为异类空间 \mathcal{D} 和同类空间 \mathcal{S} 。上述式子中的 $d_{\mathbf{M}}^2(x_i - x_j)$ 表示样本 x_i 和 x_j 的Mahalanobis距离的平方。考虑到Mahalanobis矩阵

$$\sum_{c_j \neq c_i} \sum_{v=1}^V \left(\frac{1}{|S_{c_i}|} \sum_{i=1, i \in c_i}^{|S_{c_i}|} \mathbf{x}_i^v - \frac{1}{|S_{c_j}|} \sum_{j=1, j \in c_j}^{|S_{c_j}|} \mathbf{x}_j^v \right)^T \mathbf{W}^v \mathbf{W}^{vT} \left(\frac{1}{|S_{c_i}|} \sum_{i=1, i \in c_i}^{|S_{c_i}|} \mathbf{x}_i^v - \frac{1}{|S_{c_j}|} \sum_{j=1, j \in c_j}^{|S_{c_j}|} \mathbf{x}_j^v \right) \quad (6)$$

可以分解为 $\mathbf{M}=\mathbf{W}\mathbf{W}^T$,那么上述式子等价于:

$$\frac{1}{D} \sum_{(x_i, x_j) \in \mathcal{D}} (x_i - x_j)^T \mathbf{W}\mathbf{W}^T (x_i - x_j) - \frac{1}{S} \sum_{(x_i, x_j) \in \mathcal{S}} (x_i - x_j)^T \mathbf{W}\mathbf{W}^T (x_i - x_j) \quad (3)$$

3 基于类间和类内密度的多视角度量学习

3.1 符号描述

本文公式中的矩阵用粗体大写字母表示,向量用粗体小写字母表示。数据矩阵用 $\mathbf{X} \in \mathbb{R}^{d \times N}$ 表示, \mathbf{X} 的每一列代表一个样本,即 $\mathbf{X}=[x_1, x_2, \dots, x_N]$,其中 N 表示样本个数。在多视角学习中,第 i 个样本定义为 $x_i = [(x_i^1); \dots; (x_i^v); \dots; (x_i^V)]$, $x_i^v \in \mathbb{R}^{d_v}$,其中 V 表示视角个数, d_v 为样本第 v 个视角的维度,并且有 $d = \sum_{v=1}^V d_v$ 。文中的样本对表示为 (x_i, x_j) , \mathcal{D} 和 \mathcal{S} 分别为异类样本对集合和同类样本对集合。 D 和 S 分别是异类样本对数量和同类样本对数量。 \mathbf{W}^v 代表第 v 个视角的权重矩阵。

3.2 方法框架

在图像分类任务中,样本之间的几何信息尤为重要。本文将样本的几何密度信息融入最大边际准则框架下,提出了一种新颖的方法来解决多视角图像分类问题。本文所提方法DMDML的整体框架如下:

$$\max L = L_{\text{pair}} + \gamma L_{\text{between-density}} - \beta L_{\text{inter-density}} + \eta L_{\text{complementarity}} \quad (4)$$

其中, L_{pair} 是在最大边际准则框架下的样本对度量,其目的是在最大化不同类样本对距离的同时最小化同类样本对之间的距离; $L_{\text{between-density}}$ 代表的是类间密度,通过最大化不同类中心的距离来最小化类间密度,从而减小样本在度量空间的杂乱度; $L_{\text{inter-density}}$ 是类内密度,通过最小化同一个类的样本到类中心的距离来提高类内的紧凑度,从而提高样本在度量空间的可行性; $L_{\text{complementarity}}$ 体现了多视角学习中的互补性原则,通过最大化各个视角的相关性来使视角间相互学习; γ, β 和 η 是正则项参数,用于平衡各个正则项的权重。

3.2.1 最大边际准则

本文采用度量学习通用的框架作为基础模型,将相关工作介绍的最大边际准则扩展到多视角学习,可以得到分式 L_{pair} :

$$\frac{1}{D} \sum_{(x_i^v, x_j^v) \in \mathcal{D}} \sum_{v=1}^V (x_i^v - x_j^v)^T \mathbf{W}^v \mathbf{W}^{vT} (x_i^v - x_j^v) - \frac{1}{S} \sum_{(x_i^v, x_j^v) \in \mathcal{S}} \sum_{v=1}^V (x_i^v - x_j^v)^T \mathbf{W}^v \mathbf{W}^{vT} (x_i^v - x_j^v) \quad (5)$$

根据上述介绍,需要最大化该分式。

3.2.2 类间密度

密度是描述几何结构的重要属性之一。通过约束样本在度量空间的密度,可以降低样本的杂乱度,从而提高可分性。首先,考虑类间密度。在本文,类间密度体现为每个不同类别的类中心之间的距离:距离越大说明密度越小,互相远离;距离越小说明密度越大,互相靠近。为了提高样本的分类准确度,需要最小化类间密度,即每个类的类中心之间的距离尽可能大,可以得到分式 $L_{\text{between-density}}$:

其中, $c_i, c_j \in [1, 2, \dots, C]$, C 是样本类别总数; S_{c_i}, S_{c_j} 分别是属于 c_i, c_j 类别的样本集合; $|S_{c_i}|, |S_{c_j}|$ 分别是属于 S_{c_i}, S_{c_j} 集合的元素个数。

3.2.3 类内密度

上一小节考虑的是不同类样本之间的关系,本节考虑同类样本之间的关系。通过最大化类内密度,即让属于某类别的所有样本都尽可能靠近该类的类中心,可以使同类样本在度量空间内更加紧凑,提高样本的可分性。根据上述分析,可以得到分式 $L_{\text{inter-density}}$:

$$\sum_{c_i \in C} \sum_{v=1}^V \left(\sum_{i=1, i \in c_i}^{|S_{c_i}|} (\mathbf{x}_i^v - \frac{1}{|S_{c_i}|} \sum_{j=1, j \in c_i}^{|S_{c_i}|} \mathbf{x}_j^v)^T \mathbf{W}^v \mathbf{W}^{vT} \sum_{i=1, i \in c_i}^{|S_{c_i}|} (\mathbf{x}_i^v - \frac{1}{|S_{c_i}|} \sum_{j=1, j \in c_i}^{|S_{c_i}|} \mathbf{x}_j^v) \right) \quad (7)$$

3.2.4 自适应权重交叉视角

互补性是多视角学习的基本原则之一。样本从不同视角提取的特征具有其独特的表现形式,这些来自不同视角的特征可以弥补其他视角特征所缺的部分,从而学习到更加丰富的样本数据。本文采用交叉视角来达到各个视角特征之间的互补效果。交叉视角的公式如下:

$$\text{trace}(\mathbf{W}^{vT} \mathbf{X}^v \mathbf{X}^{uT} \mathbf{W}^u) \quad (8)$$

其中, $v, u \in [1, 2, \dots, V]$, 并且 $v \neq u$ 。考虑到不同视角对模型的影响不同,对每个视角分配一个权重,并且让它们自适应进行学习,使每个视角之间的互补性最大化。自适应交叉视角的公式如下:

$$\sum_{u \neq v} \frac{\lambda_v^r + \lambda_u^r}{2} \text{trace}(\mathbf{W}^{vT} \mathbf{X}^v \mathbf{X}^{uT} \mathbf{W}^u) \quad (9)$$

$$\text{s. t. } \sum_{v=1}^V \lambda_v = 1, r > 1$$

其中,第一个约束条件表示每个视角的权重之和为 1,第二个约束条件确保每个视角的影响具有意义。

根据上述对方法框架每个子式进行的分析,将子式的详细表达式代入框架式(4),可以得到最终的目标函数表达式:

$$\begin{aligned} \max L = & \frac{1}{D} \sum_{(\mathbf{x}_i^v, \mathbf{x}_j^v) \in \mathcal{G}} \sum_{v=1}^V (\mathbf{x}_i^v - \mathbf{x}_j^v)^T \mathbf{W}^v \mathbf{W}^{vT} (\mathbf{x}_i^v - \mathbf{x}_j^v) - \\ & \frac{1}{S} \sum_{(\mathbf{x}_i^v, \mathbf{x}_j^v) \in \mathcal{S}} \sum_{v=1}^V (\mathbf{x}_i^v - \mathbf{x}_j^v)^T \mathbf{W}^v \mathbf{W}^{vT} (\mathbf{x}_i^v - \mathbf{x}_j^v) + \gamma \sum_{c_j \neq c_i} \sum_{v=1}^V \text{trace}[\mathbf{W}^{vT} (\frac{1}{|S_{c_i}|} \sum_{i=1, i \in c_i}^{|S_{c_i}|} \mathbf{x}_i^v - \frac{1}{|S_{c_j}|} \sum_{j=1, j \in c_j}^{|S_{c_j}|} \mathbf{x}_j^v) (\frac{1}{|S_{c_i}|} \sum_{i=1, i \in c_i}^{|S_{c_i}|} \mathbf{x}_i^v - \frac{1}{|S_{c_j}|} \sum_{j=1, j \in c_j}^{|S_{c_j}|} \mathbf{x}_j^v)^T \mathbf{W}^v \mathbf{W}^{vT} \\ & + \beta \sum_{c_i \in C} \sum_{v=1}^V \text{trace}[\mathbf{W}^{vT} (\frac{1}{|S_{c_i}|} \sum_{i=1, i \in c_i}^{|S_{c_i}|} \mathbf{x}_i^v - \frac{1}{|S_{c_i}|} \sum_{j=1, j \in c_i}^{|S_{c_i}|} \mathbf{x}_j^v) (\mathbf{x}_i^v - \frac{1}{|S_{c_i}|} \sum_{j=1, j \in c_i}^{|S_{c_i}|} \mathbf{x}_j^v)^T \mathbf{W}^v \mathbf{W}^{vT} \\ & + \eta \sum_{u \neq v} \frac{\lambda_v^r + \lambda_u^r}{2} \text{trace}(\mathbf{W}^{vT} \mathbf{X}^v \mathbf{X}^{uT} \mathbf{W}^u)] \end{aligned} \quad (10)$$

$$\text{s. t. } \mathbf{W}^{vT} \mathbf{W}^v = \mathbf{I}; v=1, \dots, V; \sum_{v=1}^V \lambda_v = 1; r > 1$$

3.3 优化过程

上一小节讨论的方法有两个变量需要优化,分别是 \mathbf{W}^v 和 λ_v 。本文采用固定一个变量的情况下更新另外一个变量的策略进行优化。观察目标函数可发现,前 4 项代表的是度量空间的距离,它们可以转化为求迹的形式,于是目标函数的前 4 项等价于如下子式:

$$\begin{aligned} & \frac{1}{D} \sum_{(\mathbf{x}_i^v, \mathbf{x}_j^v) \in \mathcal{G}} \sum_{v=1}^V \text{trace}[(\mathbf{W}^{vT} (\mathbf{x}_i^v - \mathbf{x}_j^v) (\mathbf{x}_i^v - \mathbf{x}_j^v)^T \mathbf{W}^v)] - \\ & \frac{1}{S} \sum_{(\mathbf{x}_i^v, \mathbf{x}_j^v) \in \mathcal{S}} \sum_{v=1}^V \text{trace}[(\mathbf{W}^{vT} (\mathbf{x}_i^v - \mathbf{x}_j^v) (\mathbf{x}_i^v - \mathbf{x}_j^v)^T \mathbf{W}^v)] + \gamma \sum_{c_j \neq c_i} \sum_{v=1}^V \text{trace}[\mathbf{W}^{vT} (\frac{1}{|S_{c_i}|} \sum_{i=1, i \in c_i}^{|S_{c_i}|} \mathbf{x}_i^v - \frac{1}{|S_{c_j}|} \sum_{j=1, j \in c_j}^{|S_{c_j}|} \mathbf{x}_j^v) (\frac{1}{|S_{c_i}|} \sum_{i=1, i \in c_i}^{|S_{c_i}|} \mathbf{x}_i^v - \frac{1}{|S_{c_j}|} \sum_{j=1, j \in c_j}^{|S_{c_j}|} \mathbf{x}_j^v)^T \mathbf{W}^v \mathbf{W}^{vT} \\ & + \beta \sum_{c_i \in C} \sum_{v=1}^V \text{trace}[\mathbf{W}^{vT} (\frac{1}{|S_{c_i}|} \sum_{i=1, i \in c_i}^{|S_{c_i}|} \mathbf{x}_i^v - \frac{1}{|S_{c_i}|} \sum_{j=1, j \in c_i}^{|S_{c_i}|} \mathbf{x}_j^v) (\mathbf{x}_i^v - \frac{1}{|S_{c_i}|} \sum_{j=1, j \in c_i}^{|S_{c_i}|} \mathbf{x}_j^v)^T \mathbf{W}^v \mathbf{W}^{vT}] \end{aligned} \quad (11)$$

用以上式子代替原目标函数的前 4 项,可得到目标函数的另外一种表达形式:

$$\begin{aligned} \max_{\mathbf{W}^v, \lambda_v} L = & \sum_{v=1}^V \text{trace}(\mathbf{W}^{vT} \mathbf{A}^v \mathbf{W}^v) + \\ & \eta \sum_{u \neq v} \frac{\lambda_v^r + \lambda_u^r}{2} \text{trace}(\mathbf{W}^{vT} \mathbf{X}^v \mathbf{X}^{uT} \mathbf{W}^u) \\ \text{s. t. } & \mathbf{W}^{vT} \mathbf{W}^v = \mathbf{I}; v=1, \dots, V; \sum_{v=1}^V \lambda_v = 1; r > 1 \end{aligned} \quad (12)$$

其中:

$$\begin{aligned} \mathbf{A}^v = & \frac{1}{D} \sum_{(\mathbf{x}_i^v, \mathbf{x}_j^v) \in \mathcal{G}} (\mathbf{x}_i^v - \mathbf{x}_j^v) (\mathbf{x}_i^v - \mathbf{x}_j^v)^T - \frac{1}{S} \sum_{(\mathbf{x}_i^v, \mathbf{x}_j^v) \in \mathcal{S}} (\mathbf{x}_i^v - \mathbf{x}_j^v) \\ & (\mathbf{x}_i^v - \mathbf{x}_j^v)^T + \gamma \sum_{c_j \neq c_i} (\frac{1}{|S_{c_i}|} \sum_{i=1, i \in c_i}^{|S_{c_i}|} \mathbf{x}_i^v - \frac{1}{|S_{c_j}|} \sum_{j=1, j \in c_j}^{|S_{c_j}|} \mathbf{x}_j^v) \\ & (\frac{1}{|S_{c_i}|} \sum_{i=1, i \in c_i}^{|S_{c_i}|} \mathbf{x}_i^v - \frac{1}{|S_{c_j}|} \sum_{j=1, j \in c_j}^{|S_{c_j}|} \mathbf{x}_j^v)^T + \beta \sum_{c_i \in C} \sum_{v=1}^V (\mathbf{x}_i^v - \frac{1}{|S_{c_i}|} \sum_{j=1, j \in c_i}^{|S_{c_i}|} \mathbf{x}_j^v) \\ & (\mathbf{x}_i^v - \frac{1}{|S_{c_i}|} \sum_{j=1, j \in c_i}^{|S_{c_i}|} \mathbf{x}_j^v)^T \end{aligned} \quad (13)$$

3.3.1 更新 \mathbf{W}^v

当更新 \mathbf{W}^v 时,固定 λ_v 为一个常数。可以观察到更新过程是一个特征值分解问题。参考文献[21],可以通过对 $\tilde{\mathbf{A}}$ 进行特征值分解得到 \mathbf{W}^v 的结果。

$$\tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{A}^1 & \eta \frac{\lambda_1^r + \lambda_2^r}{4} \mathbf{X}^1 \mathbf{X}^{2T} & \dots & \eta \frac{\lambda_1^r + \lambda_V^r}{4} \mathbf{X}^1 \mathbf{X}^{VT} \\ \eta \frac{\lambda_2^r + \lambda_1^r}{4} \mathbf{X}^2 \mathbf{X}^{1T} & \mathbf{A}^2 & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \\ \eta \frac{\lambda_V^r + \lambda_1^r}{4} \mathbf{X}^V \mathbf{X}^{1T} & \dots & \dots & \mathbf{A}^V \end{pmatrix} \quad (14)$$

3.3.2 更新 λ_v

当更新 λ_v 时,固定 \mathbf{W}^v 为一个常数。可以观察到目标函数对 λ_v 求导时,只剩下了最后一项。将约束项考虑进去并引入拉格朗日乘子 α ,再对 λ_v 和 α 分别求导可以得到:

$$\begin{cases} \frac{\partial \mathbf{F}(\lambda_v, \alpha)}{\partial \lambda_v} = \frac{\eta}{2} r \lambda_v^{r-1} \sum_{u \neq v} \text{trace}(\mathbf{W}^{vT} \mathbf{X}^v \mathbf{X}^{uT} \mathbf{W}^u) - \alpha \\ \frac{\partial \mathbf{F}(\lambda_v, \alpha)}{\partial \alpha} = - \sum_{v=1}^V \lambda_v + 1 \end{cases} \quad (15)$$

将上述求导方程组赋予 0,可以得到 λ_v :

$$\lambda_v = \frac{(1 / \sum_{u \neq v} \text{trace}(\mathbf{W}^{vT} \mathbf{X}^v \mathbf{X}^{uT} \mathbf{W}^u))^{\frac{1}{r-1}}}{\sum_{v=1}^V (1 / \sum_{u \neq v} \text{trace}(\mathbf{W}^{vT} \mathbf{X}^v \mathbf{X}^{uT} \mathbf{W}^u))^{\frac{1}{r-1}}} \quad (16)$$

3.4 时间复杂度分析

本文方法主要可分为 4 个部分来计算时间复杂度。1) 计算最大边际框架,其时间复杂度为 $O(VN)$; 2) 计算类间密度,其时间复杂度为 $O(C^2V)$; 3) 计算类内密度,其需要花费

$O(CV)$;4)求解 W^* 时需要进行的矩阵特征值分解计算,其消耗的时间复杂度为 $O(d_0^3)$ 。综上所述,可以得到本文方法的时间复杂度为 $O(VN+C^2V+CV+d^3)$ 。

4 实验及结果分析

4.1 数据集

本文在3个多视角真实世界数据集以及一个网络安全数据集上进行了实验。

(1)Still DB。此数据集包含467张图片,这些图像分别来自6种动作:抓、跑、扔、蹲、走和踢。本文提取 Sift Bow (200-dim), Color Sift Bow (200-dim) 和 Shape context Bow (200-dim)作为3个视角。

(2)Handwritten。此数据集是一个手写数字数据集,包含10个数字类别(“0”-“9”),每个类别分别有200张图像。本文提取 Pixel Averages (240-dim), Fourier Coefficients of the Character Shapes (76-dim), Profile Correlations (216-dim), Zernike Moment (47-dim), Karhunen-Love Coefficients (64-dim) 和 Morphological (6-dim)作为6个视角。

(3)Caltech101-20。Caltech101是一个包含101个类别的物体数据集。本文选择20个被广泛应用的类别作为实验数据集,并提取了 Gabor (48-dim), Wavelet Momonets (40-dim), CENTRIST (254-dim), HOG (1984-dim), GIST (512-dim) 和 LBP (928-dim)作为6个视角。

(4)NSL-KDD。此数据集是研究网络入侵检测的基准数据集,包含了5种攻击类型:Normal, Dos, Probe, R2L, U2R, 其中每个类别的样本维度为113。本文按照一样的比例选择每个类别的部分样本作为实验所用数据集。考虑到该数据集是单视角数据集,本文在应用算法模型前对该数据集进行多核映射得到多视角数据集,其中采用3个核映射,核映射类型为RBF核^[22]。

4.2 对比方法

本文选择了5个度量学习的方法作为对比方法,来验证本文所提方法的优越性。

(1)FISH-MML。该方法将FDA和HSIC结合起来用于多视角度量学习。前者探讨了单个视角内样本的可分性,后者将单视角扩展到多视角。

(2)MvGMML。该方法是GMML的多视角版本,为每个单视角分别应用GMML方法,再为其分配不同的权重,从而达到多视角的多样性。

(3)Supervised Distance Metric Learning Through Maximization of the Jeffrey Divergence(DMLMJ)^[23]。该方法通过最大化对称的Kullback-Leibler散度来约束多元高斯分布,是度量学习的一种新思路。

(4)LMNN。该方法是最典型的度量学习,用KNN分类器进行分类,是本文方法的基模型。

(5)Euclidean。经典欧氏度量学习方法。

4.3 实验设置

首先,讨论本文的实验参数设置。本文方法的模型,即式(4)中有4个参数变量 γ, β, η, r 。根据对比方法FISH-MML的参数选择,设置 γ, β 来自集合 $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$ 。对于 η 来说,首先也设置和 γ, β 一样的参数集合,发现当 $\eta < 10^0$ 和 $\eta > 10$ 时分类效果很差,所以对区间

$[1, 10]$ 进行细分,设置 η 的参数集合为 $\{1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0\}$ 。对于 r 来说,因为 λ_0 的优化过程出现了 $\frac{1}{r-1}$ 次方,所以要使优化过程有意义,本文设置参数集合 $\{1.1, 1.2, 1.5, 2.0, 2.5, 3.0\}$ 。对于参数初始化,设置对角矩阵 Λ_0 的每个元素为0.01, W^* 的每个元素为1并且它的第二维度是150。设定的终止阈值 ξ 为0.001并且最大迭代次数设为100。对比实验的参数设置都以原始论文为标准。所有的实验要进行数据集划分:80%训练集,20%测试集。将采用5折交叉验证得到的平均值作为最终的结果。本实验在KNN分类器上验证分类结果,其中 $K=5$ 。

其次,本实验采用分类准确度(ACC)和macro F1-score作为结果的衡量指标。

(1)ACC描述分类器正确判断所有样本的比例,用公式表示如下:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (17)$$

其中, TP, FP, TN, FN 分别代表正类为真、正类为假、负类为真和负类为假。

(2)macro F1-score为多分类F1分数,综合考虑精确度和召回率,用公式表示如下:

$$macro\ F1\ score = \frac{1}{C} \sum_{i=1}^C \left(2 \times \frac{precision \times recall}{precision + recall} \right) \quad (18)$$

其中, C 为类别数, $precision = \frac{TP}{TP + FP}$, $recall = \frac{TP}{TP + FN}$ 。

4.4 分类结果分析

为了验证本文所提方法的有效性,选择了两个多视角度量学习方法以及3个单视角学习方法作为对比方法。其中单视角方法采取合并多视角数据集每个视角的特征的方法,来学习更加全面的样本。本文所提方法的分类结果以及5种对比方法的结果展示在表1和表2中。从表1可以得到如下结论:本文所提方法在4个数据集上都能达到最好的分类准确度。在数据集Still DB上,本文方法和FISH-MML具有同样的分类准确度。但是,本文方法的标准差比较大,可能其对该数据集的敏感性比较高,不同的训练集和测试集具有较大差异的分类结果。对数据集Still DB进行分析可知,它由很多不同种类的渠道获得^[24],样本来源面广,可能分布比较复杂。对本文方法模型进行分析,可能对 \tilde{A} 进行特征值分解得到 W^* 的过程中对于该数据集会产生不同程度的偏差。分析实验过程可知,五轮交叉验证的训练集和测试集的划分是随机的,不同的训练集训练出来的模型在不同的测试集上的结果容易受样本复杂度的影响,从而导致实验结果具有较大的标准差。在数据集Handwritten上,本文所提方法具有最高的分类准确度和最小的标准差,说明其非常适合该数据集,并且不同的训练测试样本对结果产生的差异不大,具有稳定性。在Caltech101-20上,本文所提方法具有较大的优势,比次优的FISH-MML高了2.01%。而且可以发现DMLMJ在该数据集上仅仅11.55%的准确率,说明该方法不适合用于分类Caltech101-20。在NSL-KDD上,本文所提方法和其他5种对比方法的分类准确度和标准差都相差不大,说明这些方法都比较适合于该数据集的分类研究。但是,本文方法具有最好的效果。根据表2,也可以得到类似的F1-score结果,这里不过多阐述。综上所述,与其他5个对比方法相比,本文

所提方法具有比较明显的优势。

表1 不同数据集下的分类准确度

Table 1 Classification accuracy on different datasets (单位:%)

| Method | accuracy(mean±std.) | | | |
|-----------|---------------------|-------------|---------------|------------|
| | Still DB | Handwritten | Caltech101-20 | NSL-KDD |
| Euclidean | 37.83±3.33 | 94.25±1.13 | 74.90±0.92 | 97.19±0.63 |
| LMNN | 37.66±7.95 | 98.05±0.57 | 80.33±0.91 | 97.63±0.64 |
| DMLMJ | 38.51±3.56 | 98.65±0.52 | 11.55±9.62 | 96.92±0.41 |
| MvGMMML | 35.32±9.19 | 98.25±0.64 | 71.05±1.84 | 97.08±0.55 |
| FISH-MML | 43.83±0.55 | 97.35±0.96 | 88.24±0.17 | 96.92±0.52 |
| Ours | 43.83±4.35 | 98.90±0.34 | 90.25±0.43 | 97.83±0.50 |

表2 不同数据集下的分类 F1-score

Table 2 Macro F1-score on different datasets (单位:%)

| Method | F1-score(mean±std.) | | | |
|-----------|---------------------|-------------|---------------|------------|
| | Still DB | Handwritten | Caltech101-20 | NSL-KDD |
| Euclidean | 34.43±3.67 | 94.24±1.13 | NAN | NAN |
| LMNN | 36.55±9.10 | 98.08±0.68 | 62.63±2.28 | 86.41±6.91 |
| DMLMJ | 35.61±3.40 | 98.65±0.51 | NAN | NAN |
| MvGMMML | 32.82±8.54 | NAN | NAN | 83.55±5.71 |
| FISH-MML | 41.53±0.60 | 97.36±0.54 | 74.88±2.78 | NAN |
| Ours | 41.85±4.65 | 98.90±0.34 | 79.61±1.15 | 88.79±5.11 |

4.5 运行时间分析

3.4 节讨论了算法的时间复杂度,这里对本文方法以及对比方法做了时间上的对比。因为讨论的是算法模型的训练时间,而 Euclidean 方法只是计算欧氏距离,这里忽略不计。通过表 3 可以得出:GMMML 方法具有最少的训练时间;本文方法在 NSL_KDD 和 Handwritten 上的训练时间是 5 种方法中最长的,但是在 Still DB 上比 LMNN 快,在 Caltech101-20 上比 DMLMJ 快。总体来说,本文方法在时间上并不占优势。占据模型训练时间的主要代码是求解最大边际准则的子式,需要循环每个样本进行计算,这部分可以适当进行优化来降低运算时间;对 W^o 进行更新时的特征值分解也占据了一部分运算时间。以上两个部分的具体优化操作将会在后续工作中进行完善。

表3 不同数据集下的算法训练时间

Table 3 Training time on different datasets (单位:s)

| Method | time | | | |
|--------|----------|-------------|---------------|----------|
| | Still DB | Handwritten | Caltech101-20 | NSL-KDD |
| LMNN | 2.71 | 21.36 | 23.86 | 1847.39 |
| DMLMJ | 3.28 | 6.66 | 41.92 | 15386.92 |
| GMMML | 0.34 | 0.13 | 0.35 | 50.73 |
| FISH | 39.84 | 3.51 | 137.76 | 312.34 |
| Ours | 225.76 | 19.26 | 188.96 | 3843.37 |

4.6 参数分析

参数分析实验中,对 4 个不同的参数构造不同的参数组合,发现一个现象:对于不同的 γ, β 参数组合,都能得到当 $\eta = 3.0, r = 1.5$ 时可以取得最高的准确率,其他 η, r 组合的结果均很差的结论。所以,这一小节着重讨论本文方法对 γ 和 β 的敏感性,即呈现固定了 $\eta = 3.0, r = 1.5$ 之后对 γ 和 β 的数值进行组合得到的实验结果。首先,设置 $\eta = 3.0, r = 1.5$ 。接着, γ 和 β 从集合 $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$ 中依次选取,组成不同的参数组合,即 γ 先固定为 $10^{-4}, \beta$ 取遍 $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$ 中的所有值,并

记录分类准确度结果; γ 再固定为 $10^{-3}, \beta$ 取遍 $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$ 中的所有值,并记录分类准确度结果;类似地,直到 γ 取遍 $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$ 中的所有值。每个数据集在不同的参数组合下具有不同的分类准确度,这里以 Still DB 数据集为例,画出了在不同参数组合下的分类结果图。从图 1 可以看出,大体上分类准确度随着 β 的增大而增加,但对 γ 敏感度较大,分类结果比较参差不齐。从而可以得知,本文所提方法需要不断进行参数的调整才能得到最终较高的分类性能。

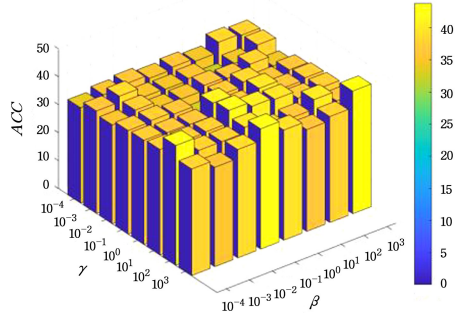


图1 参数敏感分析

Fig. 1 Parameter sensitivity analysis

4.7 消融实验

通过消融实验来进一步讨论本文所提方法的优越性。首先移除 $L_{\text{between-density}}$ 和 $L_{\text{inter-density}}$ 两项得到式(19)作为基准实验,接着在基准模型上加上 $L_{\text{between-density}}$ 项得到式(20)来探究类间密度的效果,最后在基准模型上加上 $L_{\text{inter-density}}$ 项得到式(21)来探究类内密度的效果。消融实验所选取的参数与 4.4 节分类结果分析中本文方法的 4 个结果对应的实验所选取的参数保持一致。

$$L_1 = L_{\text{pair}} + \eta L_{\text{complementarity}} \quad (19)$$

$$L_2 = L_{\text{pair}} + \gamma L_{\text{between-density}} + \eta L_{\text{complementarity}} \quad (20)$$

$$L_3 = L_{\text{pair}} - \beta L_{\text{inter-density}} + \eta L_{\text{complementarity}} \quad (21)$$

图 2 是上述 3 个变式以及本文方法在 4 个数据集上的分类效果对比图。从图中可以得出以下结论:1)除 NSL-KDD 之外,变式 L_3 的分类准确度最低,变式 L_1 和 L_2 的分类准确度介于本文所提方法和 L_3 之间;2)总体来讲,类间密度项所起的作用要大于类内密度项;3)对于 NSL-KDD 来说,只加上类内密度项会导致分类性能降低,但是它与类间密度项组合会提高分类性能。综上所述,本文所提出的两个正则项都能提高分类性能,并且两者会互相促进而进一步提高分类准确度。

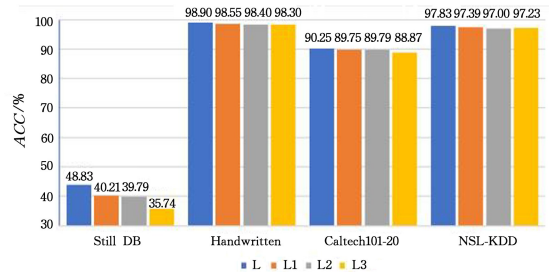


图2 消融实验

Fig. 2 Ablation experiments

结束语 本文考虑到样本的几何信息可以更直观地描述目标,并将几何信息集成到多视角学习中,提出了一种新颖的

基于密度的多视角度量学习方法。它通过最小化类间密度和最大化类内密度使样本在度量空间中更有规则性,并且让视角之间可以相互学习,最大限度地提高视角之间的相关性。基于个真实世界的多视角数据集的实验结果,可以评价所提出方法的有效性。在进行实验过程中发现本文方法的5轮结果相差比较大,下一步将探讨如何提高方法的鲁棒性。

参 考 文 献

- [1] WILDER B L, CASSIO G L. Geometric-Algebra Adaptive Filters [J]. IEEE Transactions on Signal Processing, 2019, 67(14):3649-3662.
- [2] MENG M, LAN M C, YU J, et al. Constrained Discriminative Projection Learning for Image Classification [J]. IEEE Transactions on Image Processing, 2020, 29:186-198.
- [3] DONG M Z, WANG Y J, YANG X C, et al. Learning Local Metrics and Influential Regions for Classification [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(6):1522-1529.
- [4] TSAKIRIS M C, PENG L Z, CONCA A, et al. An Algebraic-Geometric Approach for Linear Regression Without Correspondences [J]. IEEE Transactions on Information Theory, 2020, 66(8):5130-5144.
- [5] POURBAHRAMI S, KHANLI L M, AZIMPOUR S. Improving Neighborhood Construction with Apollonius Region Algorithm Based on Density for Clustering [J]. Information Sciences, 2020, 522:227-240.
- [6] CHEN H, SUN D G, LIU W Q, et al. An Automatic Registration Approach to Laser Point Sets Based on Multidiscriminant Parameter Extraction [J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(12):9449-9464.
- [7] MA J J, ZHOU S S. Metric Learning-Guided K Nearest Neighbor Multilabel Classifier [J]. Neural Computing and Applications, 2021, 33:2411-2425.
- [8] SUSAN S, KUMAR A. DST-ML-EkNN: Data Space Transformation with Metric Learning and Elite k-Nearest Neighbor Cluster Formation for Classification of Imbalanced Datasets [C]//Advances in Artificial Intelligence and Data Engineering. Select Proceedings of AIDE 2019. Advances in Intelligent Systems and Computing, 2021:319-328.
- [9] LIANG J Q, ZHU P F, DANG C Y, et al. Semisupervised Laplace-Regularized Multimodality Metric Learning [J]. IEEE Transactions on Cybernetics, 2022, 52(5):2955-2967.
- [10] DUTTA U K, HARANDI M, SEKCHAR C C. Unsupervised Deep Metric Learning via Orthogonality based Probabilistic Loss [J]. IEEE Transactions on Artificial Intelligence, 2020, 1(1):74-84.
- [11] LI Y Q, FAN X H, GAUSSIER E. Supervised Categorical Metric Learning With Schatten p-Norms [J]. IEEE Transactions on Cybernetics, 2022, 52(4):2059-2069.
- [12] WEINBERGER K Q, SAUL L K. Distance Metric Learning for Large Margin Nearest Neighbor Classification [J]. Journal of Machine Learning Research, 2009, 10:207-244.
- [13] ZADEH P H, HOSSEINI R, SRA S. Geometric Mean Metric

Learning [C]//Proceedings of the 33rd International Conference on Machine Learning(ICML). 2016:3660-3667.

- [14] GENG C X, CHEN S C. Metric Learning-Guided Least Squares Classifier Learning [J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(12):6409-6414.
- [15] HU J L, LU J W, LIU L, et al. Multi-View Geometric Mean Metric Learning for Kinship Verification [C]//2019 IEEE International Conference on Image Processing(ICIP). 2019:1178-1182.
- [16] WANG H B, FENG L, MENG X Z, et al. Multi-View Metric Learning Based on KL-Divergence for Similarity Measurement [J]. Neurocomputing, 2017, 238:269-276.
- [17] ZHANG C Q, LIU Y Q, LIU Y, et al. FISH-MML: Fisher-HSIC Multi-View Metric Learning [C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAD). 2018:3054-3060.
- [18] SUGIYAMA M. Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis [J]. Journal of Machine Learning Research, 2007, 8:1027-1061.
- [19] WANG T H, LU J, ZHANG G Q. Two-Stage Fuzzy Multiple Kernel Learning Based on Hilbert-Schmidt Independence Criterion [J]. IEEE Transactions on Fuzzy Systems, 2018, 26(6):3703-3714.
- [20] BERRENDERO J R, BUENO-LARRAZ B, CUEVAS A. On Mahalanobis Distance in Functional Settings [J]. Journal of Machine Learning Research, 2020, 21:1-33.
- [21] ABHISHEK S, ABHISHEK K, HAL D, et al. Generalized multiview analysis: A discriminative latent space [C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012, 2160-2167.
- [22] WANG Z, CHEN S C, SUN T K. MultiK-MHKS: A Novel Multiple Kernel Learning Algorithm [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(2):348-353.
- [23] BAC N, CARLOS M, BERNARD D B. Supervised Distance Metric Learning Through Maximization of the Jeffrey Divergence [J]. Pattern Recognition, 2017, 64:215-225.
- [24] IKIZLER N, CINBIS R G, PEHLIVAN S, et al. Recognizing Actions from Still Images [C]//2008 19th International Conference on Pattern Recognition. 2008:1-4.



REN Shuang-yan, born in 1997, post-graduate. Her main research interests include pattern recognition and machine learning.



WANG Zhe, born in 1981, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include pattern recognition and image processing.