软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2017,28(11):2992−3001 [doi: 10.13328/j.cnki.jos.005346]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

混杂数据的多核几何平均度量学习*

齐 忍, 朱鹏飞, 梁建青

(天津大学 计算机科学与技术学院,天津 300054)

通讯作者: 朱鹏飞, E-mail: zhupengfei@tju.edu.cn

摘 要: 在机器学习和模式识别任务中,选择一种合适的距离度量方法是至关重要的.度量学习主要利用判别性信息学习一个马氏距离或相似性度量.然而,大多数现有的度量学习方法都是针对数值型数据的,对于一些有结构的数据(比如符号型数据),用传统的距离度量来度量两个对象之间的相似性是不合理的;其次,大多数度量学习方法会受到维度的困扰,高维度使得训练时间长,模型的可扩展性差.提出了一种基于几何平均的混杂数据度量学习方法.采用不同的核函数将数值型数据和符号型数据分别映射到可再生核希尔伯特空间,从而避免了特征的高维度带来的负面影响.同时,提出了一个基于几何平均的多核度量学习模型,将混杂数据的度量学习问题转化为求黎曼流形上两个点的中心点问题.在 UCI 数据集上的实验结果表明,针对混杂数据的多核度量学习方法与现有的度量学习方法相比,在准确性方面展现出更优异的性能.

关键词: 几何平均;多核学习;度量学习;混杂数据
中图法分类号: TP181

中文引用格式: 齐忍,朱鹏飞,梁建青.混杂数据的多核几何平均度量学习.软件学报,2017,28(11):2992−3001. http://www.jos.org.cn/1000-9825/5346.htm

英文引用格式: Qi R, Zhu PF, Liang JQ. Multiple kernel geometric mean metric learning for heterogeneous data. Ruan Jian Xue Bao/Journal of Software, 2017,28(11):2992−3001 (in Chinese). http://www.jos.org.cn/1000-9825/5346.htm

Multiple Kernel Geometric Mean Metric Learning for Heterogeneous Data

QI Ren, ZHU Peng-Fei, LIANG Jian-Qing

(School of Computer Science and Technology, Tianjin University, Tianjin 300054, China)

Abstract: How to choose a proper distance metric is vital to many machine learning and pattern recognition tasks. Metric learning mainly uses discriminant information to learn a Mahalanobis distance or similarity metric. However, most existing metric learning methods are for numerical data, and it is unreasonable to calculate the similarity between two heterogeneous objects (e.g., categorical data) using traditional distance metrics. Besides, they suffer from curse of dimensionality, resulting in poor efficiency and scalability when the feature dimension is very high. In this paper, a geometric mean metric learning method is proposed for heterogeneous data. The numerical data and categorical data are mapped to a reproducing kernel Hilbert space by using different kernel functions, thus avoiding the negative influence of the high dimensionality of the feature. At the same time, a multiple kernel metric learning model based on geometric mean is introduced to transform the metric learning problem of heterogeneous data into solving the midpoint between two points on the Riemannian manifold. Experiments on benchmark UCI datasets show that the presented method shows promising performances in terms of accuracy in comparison with the state-of-the-art metric learning methods.

Key words: geometric mean; multi-kernel learning; metric learning; heterogeneous data

距离度量用来测量两个数据对象之间的距离或不相似性,在许多机器学习和模式识别任务中起着重要作

* 基金项目: 国家自然科学基金(61502332, 61732011)
Foundation item: National Natural Science Foundation of China (61502332, 61732011)
本文由复杂环境下的机器学习研究专刊特约编辑胡清华教授、张道强教授、张长水教授推荐.
收稿时间: 2017-05-13; 修改时间: 2017-06-16; 采用时间: 2017-08-23

用.例如在分类中,K 近邻分类器[1]使用距离度量来识别最近的邻居;许多聚类算法,如 K-means[2]也依赖于数据点之间的距离度量;在信息检索中,文档通常根据其与给定查询的相似性或相关性进行排名.有些距离度量被广泛使用,包括欧几里德距离和特征向量的余弦相似度等.距离度量对方法的性能有极大的影响,针对不同的任务和不同类型的数据选择合适的距离度量,是度量学习的主要任务.

2002 年,Xing 等人[3]开创性的工作意味着度量学习的真正发展,他们把度量学习定义为凸优化问题.度量学习的目标是从给定的样本对约束中学习到一个期望的度量,使得相似的样本间的距离越小越好,不相似的样本之间的距离越大越好,二元组约束一般表示为

$$S:=\{(\boldsymbol{x}_i,\boldsymbol{x}_j)\mid \boldsymbol{x}_i和\boldsymbol{x}_j是相似的\},$$

$$D:=\{(\boldsymbol{x}_i,\boldsymbol{x}_j)\mid \boldsymbol{x}_i和\boldsymbol{x}_j是不相似的\}.$$

三元组的约束一般表示为

$$R=\{(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{x}_k):\boldsymbol{x}_i 应该比\ \boldsymbol{x}_k 更接近\ \boldsymbol{x}_j\}.$$

度量学习在模式识别和计算机视觉任务中发挥着重要作用,它在 2011 年被 Shaw 等人[4]应用于网络中的链路预测,被 Taylor 等人[5]用于强化学习中的状态表示,在 2012 年被 McFee 等人[6]用于音乐推荐等.在计算机视觉任务中,存在着大量的度量学习研究工作,如人脸识别[7]、图像分类[8]、视觉跟踪[9]等.生物信息学中的许多问题涉及比较 DNA、时间序列等.这些比较基于结构化度量,例如将距离度量用于时间序列的字符串或动态时间扭曲距离的编辑,又如 Xiong 和 Chen[10]在工作中对度量学习的应用.

对于纯数值型数据集,距离计算是容易处理的,因为已经提出了大量数值型数据的度量学习方法可以直接应用.而混杂数据的出现,使得之前的算法不再适用,尽管使用数值型度量学习方法也得到了不错的实验结果,但其结果仍是不合理的.举例来说,若数字 1 代表黄色,2 代表紫色,3 代表蓝色,用“3–1”表示黄色与蓝色的距离、“3–2”表示紫色与蓝色的距离显然是不合理的.因此,提出一种对于混杂型数据的度量学习方法是很有必要的.

目前,针对符号型数据最直接的距离度量是 Esposito 等人[11]使用的汉明距离.更多的研究人员试图通过考虑名义属性值的分布特征来度量距离.例如,Cost 等人[12]提出了一种用于监督学习任务的修正值差分度量(MVDM).Ienco 等人[13]提出了上下文的概念,基于来自当前属性的上下文的属性值来测量属性的两个值之间的距离.数值实验和分析发现:如果给定数据集的属性之间不相互独立,这 3 种间接定义的距离度量[14−16]无法起作用.所有这些相似度度量单独地对待名义属性,并忽略变体属性关系.

核度量学习对于处理具有特殊结构的数据独具优势,此外也可以解决维度灾难问题.对于度量学习中的高维度挑战,通常在学习度量之前进行降维[17].虽然研究显示降维有助于降低过拟合风险,但缺少理论支持.目前已有许多基于核的度量学习方法,He 等人[18]提出了基于概率的距离测度核密度度量学习,虽然可以处理数值型与符号型数据,但却一概而论.核化的判别成分分析方法(KDCA)[19]、核化的大间隔成分分析方法(KLMCA)[20]和核化的基于信息理论的度量学习(KITML)[21]都直接使用内核技巧将线性算法对应扩展到核度量学习.

在本文中,我们提出了一种基于几何平均的混杂数据度量学习算法.该算法通过拉近相似样本对之间的距离和推远不相似样本对之间的距离来进行度量学习.首先,将数据集拆分为符号型与数值型,其中,符号型数据用汉明距离处理;然后,通过高斯核函数将其分别映射到可再生核希尔伯特空间;之后,将计算所得核矩阵代入目标函数,为充分利用不同类型数据性质,我们为核矩阵分配了权重,为保持相似矩阵和不相似矩阵的平衡,我们在测地的视角来分配矩阵的权重,从而将混杂数据的度量学习问题转化为求黎曼流形上的两个点的中心点的问题,最终分别计算出对应的矩阵 A,求得马氏距离.与现有的度量学习方法相比,基于几何平均的混杂数据度量学习具有高度可扩展性和高效性.UCI 数据集的实证结果验证了我们的算法在分类精度上有较大的性能提升.

本文第 1 节简要回顾度量学习的相关工作.第 2 节介绍所提出的核化几何平均度量学习和多核几何平均度量学习.第 3 节给出优化和算法.第 4 节分析在数值型数据、符号型数据和混杂数据上的 3 组实验结果.第 5 节总结我们的研究.

1 相关工作

这一节我们介绍几种经典的有监督的马氏距离度量学习算法.

MMC是Xing等人[3]开创的第1种马氏距离学习方法.它建立在一个没有正则项的凸公式上,目标是最大化不相似点之间的距离总和,同时保持相似点之间的距离总和尽可能地小.

$$\left.\begin{aligned}&\max_{\boldsymbol{M}\in\boldsymbol{S}_+^d}\sum_{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\boldsymbol{D}}d_{\boldsymbol{M}}(\boldsymbol{x}_i,\boldsymbol{x}_j)\\&\text{s.t.}\sum_{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\boldsymbol{S}}d_{\boldsymbol{M}}^2(\boldsymbol{x}_i,\boldsymbol{x}_j)\le 1\end{aligned}\right\}\tag{1}$$

为了求解公式(1),Xing等人[3]使用梯度下降算法结合投影半正定矩阵进行优化,需要在每次迭代时对M进行全特征值分解.这通常对高维问题是难处理的.

LMNN是Weinberger等人[22]提出的,它是使用很广泛的一种马氏距离学习方法.其约束以如下方式定义:

$$\left.\begin{aligned}&S=\{(\boldsymbol{x}_i,\boldsymbol{x}_j):\boldsymbol{y}_i=\boldsymbol{y}_j\text{且}\boldsymbol{x}_j\text{属于}\boldsymbol{x}_i\text{的}k\text{-近邻}\}\\&R=\{(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{x}_k):(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\boldsymbol{S},\boldsymbol{y}_i\ne\boldsymbol{y}_k\}\end{aligned}\right\}\tag{2}$$

该方法的距离度量使用如下的凸公式:

$$\left.\begin{aligned}&\min_{\boldsymbol{M}\in\boldsymbol{S}_+^d}\ (1-\mu)\sum_{(\boldsymbol{x}_i,\boldsymbol{x}_j)\in\boldsymbol{S}}d_{\boldsymbol{M}}^2(\boldsymbol{x}_i,\boldsymbol{x}_j)+\mu\sum_{i,j,k}\boldsymbol{\xi}_{ijk}\\&\text{s.t. }d_{\boldsymbol{M}}^2(\boldsymbol{x}_i,\boldsymbol{x}_k)-d_{\boldsymbol{M}}^2(\boldsymbol{x}_i,\boldsymbol{x}_j)\ge 1-\boldsymbol{\xi}_{ijk},\forall(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{x}_k)\in R\end{aligned}\right\}\tag{3}$$

其中,$\mu\in[0,1]$决定着是“拉近”还是“推远”,$\boldsymbol{\xi}_{ijk}$为松弛变量.LMNN虽然由于没有正则项有时候会过拟合,但结果一般很好,特别是在高维的情况下.

ITML是Davis等人[21]提出的,它用布雷格曼散度衡量亲密度: $D_{ld}(\boldsymbol{M},\boldsymbol{M}_0)=tr(\boldsymbol{M}\boldsymbol{M}_0^{-1})-\log\det(\boldsymbol{M}\boldsymbol{M}_0^{-1})-d$,其中,$d$是论域的维度,$\boldsymbol{M}_0$是正定矩阵.同时,正则项用来保持学习的距离尽可能地接近欧式距离.当且仅当M是正定的,LogDet主要特征才是有限的.因此,最小化$D_{ld}(M,M_0)$有一个条件是:M至少是半正定的.ITML公式如下:

$$\left.\begin{aligned}&\min_{M\in\boldsymbol{S}_+^d}\ D_{ld}(\boldsymbol{M},\boldsymbol{M}_0)+r\sum_{i,j}\boldsymbol{\xi}_{ij}\\&\text{s.t. }d_{\boldsymbol{M}}^2(\boldsymbol{x}_i,\boldsymbol{x}_j)\le u+\boldsymbol{\xi}_{ij}\forall(\boldsymbol{x}_i,\boldsymbol{x}_j)\in S\\&\quad d_{\boldsymbol{M}}^2(\boldsymbol{x}_i,\boldsymbol{x}_j)\ge v-\boldsymbol{\xi}_{ij}\forall(\boldsymbol{x}_i,\boldsymbol{x}_j)\in D\end{aligned}\right\}\tag{4}$$

其中,u和v是阈值参数,用来控制相似点间距离小和不相似的点间距离大的程度;r为惩罚因子.最小化$D_{ld}(M,M_0)$等价于最小化两个由M和M_0参量化的多变量高斯分布,最终能够达到收敛到全局最小值的结果.

Doublet-SVM是王法强等人[23]提出的方法.该方法将度量学习问题变成一个样本对分类问题.它首先忽略半正定约束,并使用SVM来学习初始度量M,然后将M映射到半正定矩阵的空间上.目标函数如下所示.

$$\left.\begin{aligned}&\min_{\boldsymbol{M},b,\xi}\ \frac{1}{2}\|\boldsymbol{M}\|_F^2+C\sum_l\xi_l\\&\text{s.t. }h_l((\boldsymbol{x}_{l,1}-\boldsymbol{x}_{l,2})^T\boldsymbol{M}(\boldsymbol{x}_{l,1}-\boldsymbol{x}_{l,2})+b)\ge 1-\xi_l\\&\quad\xi_l\ge 0,\forall l\end{aligned}\right\}\tag{5}$$

目标函数的正则项为$\boldsymbol{r}_{SVM}(M)=\frac{1}{2}\|M\|_F^2$,Hinge损失惩罚项为$\rho_{SVM}(\xi)=C\sum_l\xi_l$,其中,$C$为惩罚因子,$\xi_l$为松变量.它可以利用现有的SVM工具箱来解决,比如LibSVM[24].

GMML(geometric mean metric learning)是由Pourya等人[25]提出的算法.该方法的主要创新点是在目标函数中加入了不相似点的项.与最原始的度量学习方法MMC类似,他们提出找到一个A,使得所有相似点间的距离总和尽可能地小.与之前方法不同的是,提出了用A^{-1}测量不相似点之间的距离,这样就很巧妙地可以仅通过一个目标函数而取得既满足使相似点距离小又满足使不相似点距离大的效果.他们提出的新目标函数为

$$\min\sum_{(x,y)\in\boldsymbol{S}}d_{\boldsymbol{A}}(\boldsymbol{x},\boldsymbol{y})+\sum_{(x,y)\in\boldsymbol{D}}d_{\boldsymbol{A}^{-1}}(\boldsymbol{x},\boldsymbol{y})\tag{6}$$

其中,S 为相似点样本对组成的集合,D 为不相似点样本对组成的集合.

2 混杂数据度量学习

2.1 核化几何平均度量学习

核函数是用来计算两个向量在映射过后的空间中的内积函数.它可以解决维度爆炸的问题,在处理维度高的样本时会节省大量的时间.假设学习样本为 $\boldsymbol{X}:n\times d,\boldsymbol{x}^T,\boldsymbol{y}^T\in R^d$ 分别代表着矩阵的第 x 列和第 y 列.$\phi(\boldsymbol{x})$是把 x 映射到特征空间的函数.我们选择了高斯核函数 $k(\boldsymbol{x},\boldsymbol{y})=\langle\phi(\boldsymbol{x}),\phi(\boldsymbol{y})\rangle=\exp(-\|\boldsymbol{x}-\boldsymbol{y}\|^2/(2\sigma^2))$,因此在再生核希尔伯特空间中的马氏距离被重新定义为 $d_M(\phi(\boldsymbol{x}),\phi(\boldsymbol{y}))=(\phi(\boldsymbol{x})-\phi(\boldsymbol{y}))^T\boldsymbol{M}(\phi(\boldsymbol{x})-\phi(\boldsymbol{y}))$.式中 $\boldsymbol{M}\geq 0$ 为半正定矩阵.Jain 等人[26]证明了马氏距离度量的最优解形式为 $\boldsymbol{M}=\eta\boldsymbol{I}+\phi(\boldsymbol{X})^T\boldsymbol{A}\phi(\boldsymbol{X})$,其中,$I$ 是单位矩阵,$\boldsymbol{A}$ 为半正定矩阵,η为常数.由于现有的度量学习方法η常设为 0,在这里,我们只考虑η为 0 的优化情况.我们将 $\boldsymbol{M}$ 进一步优化表示为 $\boldsymbol{M}=\phi(\boldsymbol{X})^T\boldsymbol{A}\phi(\boldsymbol{X})$,$\phi(\boldsymbol{X})$为学习样本.由上节讨论可知,原目标函数为公式(6),在核空间中重新定义相似点间距离为 $d_{\boldsymbol{A}}=(\boldsymbol{K}_x-\boldsymbol{K}_y)^T\boldsymbol{A}(\boldsymbol{K}_x-\boldsymbol{K}_y)$,其中,$K$ 为通过核函数求得的核矩阵,$\boldsymbol{K}_x$ 和 $\boldsymbol{K}_y$ 分别表示核矩阵的第 x 列和第 y 列.用 $\boldsymbol{A}^{-1}$ 矩阵计算不相似点的距离总和.因此,核化的几何平均度量学习算法的目标函数为

$$\min\sum_{(x,y)\in\boldsymbol{S}}(\boldsymbol{K}_x-\boldsymbol{K}_y)^T\boldsymbol{A}(\boldsymbol{K}_x-\boldsymbol{K}_y)+\sum_{(x,y)\in\boldsymbol{D}}(\boldsymbol{K}_x-\boldsymbol{K}_y)^T\boldsymbol{A}^{-1}(\boldsymbol{K}_x-\boldsymbol{K}_y) \tag{7}$$

我们进一步简化公式(7),用迹函数来重写马氏距离,将公式(7)变为优化问题:

$$\min\sum_{(x,y)\in\boldsymbol{S}}tr(\boldsymbol{A}(\boldsymbol{K}_x-\boldsymbol{K}_y)(\boldsymbol{K}_x-\boldsymbol{K}_y)^T)+\sum_{(x,y)\in\boldsymbol{D}}tr(\boldsymbol{A}^{-1}(\boldsymbol{K}_x-\boldsymbol{K}_y)(\boldsymbol{K}_x-\boldsymbol{K}_y)^T) \tag{8}$$

我们定义下面两个重要的矩阵 S 和 D.

$$\left.\begin{aligned}\boldsymbol{S}&:=\sum_{(x,y)\in\boldsymbol{S}}(\boldsymbol{K}_x-\boldsymbol{K}_y)(\boldsymbol{K}_x-\boldsymbol{K}_y)^T\\ \boldsymbol{D}&:=\sum_{(x,y)\in\boldsymbol{D}}(\boldsymbol{K}_x-\boldsymbol{K}_y)(\boldsymbol{K}_x-\boldsymbol{K}_y)^T\end{aligned}\right\} \tag{9}$$

这里的 S 表示相似矩阵,D 表示不相似矩阵.通过公式(9)即可形成我们最终优化的目标函数.

$$\min_{\boldsymbol{M}\succ 0}\boldsymbol{F}_h(\boldsymbol{A})=\min tr(\boldsymbol{A}\boldsymbol{S})+tr(\boldsymbol{A}^{-1}\boldsymbol{D}) \tag{10}$$

代价函数(10)有很多重要的性质可以帮助我们去求得其最小值,首先,$F_h(\boldsymbol{A})$既是严格凸的,又是严格测地凸的.因此,当$\nabla F_h(\boldsymbol{A})=0$ 有解时,该解即为全局最小值.半正定矩阵的集合形成了一个非正曲率的黎曼流形[27].半正定矩阵 A 和 B 在流形上的中点为 $\boldsymbol{A}\#_{1/2}\boldsymbol{B}=\boldsymbol{A}^{1/2}(\boldsymbol{A}^{-1/2}\boldsymbol{B}\boldsymbol{A}^{-1/2})^t\boldsymbol{A}^{1/2}$.

在整个半正定矩阵集合上,测地凸函数的定义[28]如下.

定义 **1**. 若黎曼流形的测地凸子集上的函数 f 是测地凸的,那么该子集上的任意两点 A,B 满足:

$$f(\boldsymbol{A}\#_t\boldsymbol{B})\leq tf(\boldsymbol{A})+(1-t)f(\boldsymbol{B}),t\in(0,1) \tag{11}$$

定理 **1**. 在半正定流形上的代价函数(10)有既是严格凸的又是严格测地凸的性质,由此可得其全局最小值:

$$\nabla F_h(\boldsymbol{A})=\boldsymbol{S}-\boldsymbol{A}^{-1}\boldsymbol{D}\boldsymbol{A}^{-1} \tag{12}$$

$$\nabla F_h(\boldsymbol{A})=0\rightarrow\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}=\boldsymbol{D} \tag{13}$$

实际上,公式(13)的唯一解是 $\boldsymbol{D}^{-1}$ 和 $\boldsymbol{D}$ 测地线的中点:

$$\boldsymbol{A}=\boldsymbol{S}^{-1}\#_{1/2}\boldsymbol{D}=\boldsymbol{S}^{-1/2}(\boldsymbol{S}^{1/2}\boldsymbol{D}\boldsymbol{S}^{1/2})^{1/2}\boldsymbol{S}^{-1/2} \tag{14}$$

从上面 A 的定义很容易看出,A 是满足半正定的.

2.2 多核几何平均度量学习

为了解决混杂数据的问题,我们提出了多核几何平均度量,不同的核对应不同种类的数据.本文实验混杂数据体现为数值型和符号型的混杂,将数据拆分之后,对于符号型数据,我们通过汉明距离对其进行处理.首先,通过 Esposito 等人[11]使用的汉明距离将其转化为汉明矩阵.汉明距离定义如下:

$$D(\boldsymbol{x}_i,\boldsymbol{x}_j)=\sum_{r=1}^{d}\delta(\boldsymbol{x}_{ir},\boldsymbol{x}_{jr}),$$

$$\delta(\boldsymbol{x}_{ir},\boldsymbol{x}_{jr})=\begin{cases}1, & \text{如果}\boldsymbol{x}_{ir}\neq\boldsymbol{x}_{jr}\\ 0, & \text{如果}\boldsymbol{x}_{ir}=\boldsymbol{x}_{jr}\end{cases}.$$

然后,将得到的汉明矩阵当作新的样本特征矩阵.符号型数据和数值型数据会通过高斯核函数得到两个不同的核矩阵.假设样本集为 $\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_N]$,每个样本 $\boldsymbol{x}_i$ 有 h 种数据,则 h 即为我们多核几何平均度量学习算法中核的数目.使用不同的核求出马氏距离,同时乘以为其分配的权重系数,最终将多个核计算的加权距离相加,得到两个样本的度量距离.多核几何平均度量学习的目标函数为

$$\min\sum_{g=1}^{h}\boldsymbol{\omega}_g\left(\sum_{(\boldsymbol{x}_i^g,\boldsymbol{x}_j^g)\in\boldsymbol{S}_g}(\boldsymbol{K}_x^g-\boldsymbol{K}_y^g)^T\boldsymbol{A}_g(\boldsymbol{K}_x^g-\boldsymbol{K}_y^g)+\sum_{(\boldsymbol{x}_i^g,\boldsymbol{x}_j^g)\in\boldsymbol{D}_g}(\boldsymbol{K}_x^g-\boldsymbol{K}_y^g)^T\boldsymbol{A}_g^{-1}(\boldsymbol{K}_x^g-\boldsymbol{K}_y^g)\right)\tag{15}$$

为了充分利用不同类型数据的性质,我们设置了参数 ω_g,它决定着在目标函数中第 g 种数据所占的权重.$\boldsymbol{A}_g$ 是第 g 种数据要学得的对称正定矩阵.然后,依次将计算所得核矩阵代入目标函数.我们的实验数据都是给定标签的,由此,可以为每种数据形成样本对集合:

$$S_g:=\{(\boldsymbol{x}_i^g,\boldsymbol{x}_j^g)\mid \boldsymbol{x}_i^g\text{和}\boldsymbol{x}_j^g\text{是同类}\},$$

$$D_g:=\{(\boldsymbol{x}_i^g,\boldsymbol{x}_j^g)\mid \boldsymbol{x}_i^g\text{和}\boldsymbol{x}_j^g\text{是不同类}\}.$$

3 优化和算法

3.1 优　化

现在我们的目标函数如下所示:

$$\min\sum_{g=1}^{h}\boldsymbol{\omega}_g\left(\sum_{(\boldsymbol{x}_i^g,\boldsymbol{x}_j^g)\in\boldsymbol{S}_g}(\boldsymbol{K}_x^g-\boldsymbol{K}_y^g)^T\boldsymbol{A}_g(\boldsymbol{K}_x^g-\boldsymbol{K}_y^g)+\sum_{(\boldsymbol{x}_i^g,\boldsymbol{x}_j^g)\in\boldsymbol{D}_g}(\boldsymbol{K}_x^g-\boldsymbol{K}_y^g)^T\boldsymbol{A}_g^{-1}(\boldsymbol{K}_x^g-\boldsymbol{K}_y^g)\right).$$

这里,我们使用迹函数来重写,公式(15)变为

$$\min_{\{\boldsymbol{A}_g\}_{g=1}^{h}\succ 0}\sum_{g=1}^{h}\boldsymbol{\omega}_g\left(\sum_{(\boldsymbol{x}_i^g,\boldsymbol{x}_j^g)\in S_g}tr(\boldsymbol{A}_g(\boldsymbol{K}_x^g-\boldsymbol{K}_y^g)(\boldsymbol{K}_x^g-\boldsymbol{K}_y^g)^T)+\sum_{(\boldsymbol{x}_i^g,\boldsymbol{x}_j^g)\in D_g}tr(\boldsymbol{A}_g^{-1}(\boldsymbol{K}_x^g-\boldsymbol{K}_y^g)(\boldsymbol{K}_x^g-\boldsymbol{K}_y^g)^T)\right)\tag{16}$$

然后,我们定义第 g 种数据的相似点矩阵 $\boldsymbol{S}_g$ 和不相似点矩阵 $\boldsymbol{D}_g$:

$$\left.\begin{aligned}\boldsymbol{S}_g&:=\sum_{(\boldsymbol{x}_i^g,\boldsymbol{x}_j^g)\in\boldsymbol{S}_g}(\boldsymbol{K}_x^g-\boldsymbol{K}_y^g)(\boldsymbol{K}_x^g-\boldsymbol{K}_y^g)^T,\\ \boldsymbol{D}_g&:=\sum_{(\boldsymbol{x}_i^g,\boldsymbol{x}_j^g)\in\boldsymbol{D}_g}(\boldsymbol{K}_x^g-\boldsymbol{K}_y^g)(\boldsymbol{K}_x^g-\boldsymbol{K}_y^g)^T\end{aligned}\right\}\tag{17}$$

因此,我们可以得到多核几何平均度量学习的优化后函数:

$$\min_{\{\boldsymbol{A}_g\}_{g=1}^{h}\succ 0}h(\{\boldsymbol{A}_g\}_{g=1}^{h}):=\sum_{g=1}^{h}\boldsymbol{\omega}_g(tr(\boldsymbol{A}_g\boldsymbol{S}_g)+tr(\boldsymbol{A}_g^{-1}\boldsymbol{D}_g))\tag{18}$$

因为 S_g 可能是不可逆的,我们对目标函数加入了一个正则项[20]:

$$\min_{\{\boldsymbol{A}_g\}_{g=1}^{h}\succ 0}\sum_{g=1}^{h}\boldsymbol{\omega}_g(tr(\boldsymbol{A}_g\boldsymbol{S}_g)+tr(\boldsymbol{A}_g^{-1}\boldsymbol{D}_g))+\lambda\sum_{g=1}^{h}\boldsymbol{\omega}_g\boldsymbol{D}_{sld}(\boldsymbol{A}_g,\boldsymbol{A}_0)\tag{19}$$

这里的 A_0 是先验矩阵,在后面实验部分会详细说明;D_{sld} 是对称的 DetLog 散度:

$$\boldsymbol{D}_{sld}(\boldsymbol{A}_g,\boldsymbol{A}_0):=tr(\boldsymbol{A}_g\boldsymbol{A}_0^{-1})+tr(\boldsymbol{A}_g^{-1}\boldsymbol{A}_0)-2d\tag{20}$$

另外,我们需要注意另一个变量是 ω_g.为了确保距离是正的,我们要求 ω_g 是非负的.然而,由于距离和偏差都是非负的,所以当每个 ω_g 等于 0 时,目标函数获得最小值.由于我们希望每种数据都能参与到目标函数中,所以我们使 ω_g 的和为一个常数.进而目标函数成为一个线性规划,由此可能导致大多数权重接近 0.为了避免过度拟合,

我们对 ω_g 也引入一个正则项,最终,正则化的多核几何平均度量学习目标函数为

$$\left.\begin{aligned}&\min_{\{\boldsymbol{A}_g\}_{g=1}^h\succ 0}\ \sum_{g=1}^h\boldsymbol{\omega}_g(tr(\boldsymbol{A}_g\boldsymbol{S}_g)+tr(\boldsymbol{A}_g^{-1}\boldsymbol{D}_g))+\lambda\sum_{g=1}^h\boldsymbol{\omega}_g\boldsymbol{D}_{sld}(\boldsymbol{A}_g,\boldsymbol{A}_0)+\gamma\sum_{g=1}^h\boldsymbol{\omega}_g^2\\&\text{s.t. }\boldsymbol{A}_g\succ 0,g=1,2,...,h\\&\quad\boldsymbol{\omega}_g\geq 0,g=1,2,...,h\\&\quad\sum_{g=1}^h\boldsymbol{\omega}_g=1\end{aligned}\right\}\tag{21}$$

$\boldsymbol{\omega}=[\omega_1,\omega_2,\ldots,\omega_h]$为一个 h 维向量,$\sum_{g=1}^h\omega_g^2$ 等于 $\|\omega\|_2^2$.

3.2 求 解

下面对公式(21)进行求解,观察可知,公式中唯一的约束 A_g 是正定的,为方便起见,用 Q 来代替公式(21),即

$$\boldsymbol{Q}=\min_{\{\boldsymbol{A}_g\}_{g=1}^h\succ 0}\sum_{g=1}^h\boldsymbol{\omega}_g(tr(\boldsymbol{A}_g\boldsymbol{S}_g)+tr(\boldsymbol{A}_g^{-1}\boldsymbol{D}_g))+\lambda\sum_{g=1}^h\boldsymbol{\omega}_g\boldsymbol{D}_{sld}(\boldsymbol{A}_g,\boldsymbol{A}_0)+\gamma\sum_{g=1}^h\boldsymbol{\omega}_g^2\tag{22}$$

Q 对 A_g 的导数为

$$\frac{\partial\boldsymbol{Q}}{\partial\boldsymbol{A}_g}=\boldsymbol{\omega}_g(\boldsymbol{S}_g-\boldsymbol{A}_g^{-1}\boldsymbol{D}_g\boldsymbol{A}_g^{-1})+\lambda\boldsymbol{\omega}_g(\boldsymbol{A}_0^{-1}-\boldsymbol{A}_g^{-1}\boldsymbol{A}_0\boldsymbol{A}_g^{-1})\tag{23}$$

令其为 0,得到ω_g=0 或者 $\boldsymbol{S}_g-\boldsymbol{A}_g^{-1}\boldsymbol{D}_g\boldsymbol{A}_g^{-1}+\lambda(\boldsymbol{A}_0^{-1}-\boldsymbol{A}_g^{-1}\boldsymbol{A}_0\boldsymbol{A}_g^{-1})=0$.又因为 $\sum_{g=1}^h\omega_g^2=1$,所以只能是

$$\boldsymbol{S}_g-\boldsymbol{A}_g^{-1}\boldsymbol{D}_g\boldsymbol{A}_g^{-1}+\lambda(\boldsymbol{A}_0^{-1}-\boldsymbol{A}_g^{-1}\boldsymbol{A}_0\boldsymbol{A}_g^{-1})=0\tag{24}$$

由此解得 $\boldsymbol{A}_g=(\boldsymbol{S}_g+\lambda\boldsymbol{A}_0^{-1})^{-1}\#_{1/2}(\boldsymbol{D}_g+\lambda\boldsymbol{A}_0)$.由几何平均的形式,我们可以知道 $\boldsymbol{A}$ 是半正定的.一旦 $\boldsymbol{A}$ 确定了,目标函数(21)就转为如下形式:

$$\left.\begin{aligned}&\min\sum_{g=1}^h c_{1g}\boldsymbol{\omega}_g+\lambda\sum_{g=1}^h c_{2g}\boldsymbol{\omega}_g+\gamma\sum_{g=1}^h\boldsymbol{\omega}_g^2\\&\text{s.t. }\boldsymbol{\omega}_g\geq 0,g=1,2,...,h\\&\quad\sum_{g=1}^h\boldsymbol{\omega}_g=1\end{aligned}\right\}\tag{25}$$

其中,c_{1g} 和 c_{2g} 均为常数,$c_{1g}=tr(\boldsymbol{A}_g\boldsymbol{S}_g)+tr(\boldsymbol{A}_g^{-1}\boldsymbol{D}_g)$,$c_{2g}=\boldsymbol{D}_{sld}(\boldsymbol{A}_g,\boldsymbol{A}_0)$.然后,我们就可以通过二次方程来解得最优的 $\boldsymbol{\omega}_g$ 组合向量.

3.3 加 权

在测地的视角来分配相似矩阵和不相似矩阵的权重,对多核几何平均度量学习方法求解同样也是很重要的.因为仅通过一个常数来放缩 $\boldsymbol{A}$ 的解来保持 $\boldsymbol{S}$ 和 $\boldsymbol{D}$ 的平衡是没有意义的.我们根据多核几何平均度量学习方法的性质,加入了对称正定矩阵的黎曼几何上的非线性的一项,等价于下面的优化问题:

$$\min_{\{\boldsymbol{A}_g\}_{g=1}^h\succ 0}h_t(\{\boldsymbol{A}_g\}_{g=1}^h):=(1-t)\sum_{g=1}^h\boldsymbol{\omega}_g\delta_R^2(\boldsymbol{A}_g,\boldsymbol{S}_g^{-1})+t\sum_{g=1}^h\boldsymbol{\omega}_g\delta_R^2(\boldsymbol{A}_g,\boldsymbol{D}_g)\tag{26}$$

其中,δ_R 是对称正定矩阵上的黎曼距离:

$$\delta_R(\boldsymbol{X},\boldsymbol{Y}):=\|\log(\boldsymbol{Y}^{-1/2}\boldsymbol{X}\boldsymbol{Y}^{-1/2})\|_F,\boldsymbol{X},\boldsymbol{Y}\succ 0\tag{27}$$

因为权重值是正的且固定的,相似和不相似矩阵也是已知的,所以,公式(26)等价于执行 h 次下面的任务:

$$\min_{\boldsymbol{A}_g\succ 0}h_t(\boldsymbol{A}_g):=(1-t)\delta_R^2(\boldsymbol{A}_g,\boldsymbol{S}_g^{-1})+t\delta_R^2(\boldsymbol{A}_g,\boldsymbol{D}_g)\tag{28}$$

每一次的唯一解为加权的几何平均 $\boldsymbol{A}_g=\boldsymbol{S}_g^{-1}\#_t\boldsymbol{D}_g$,因此正则化的解为

$$\boldsymbol{A}_g = (\boldsymbol{S}_g + \lambda \boldsymbol{A}_0^{-1})^{-1} \#_t (\boldsymbol{D}_g + \lambda \boldsymbol{A}_0), t \in [0,1] \tag{29}$$

3.4 时间复杂度分析与讨论

我们假设样本的数目为 N,样本对数目表示为 T.几何平均度量学习方法的时间消耗主要有两部分:计算 S,D,A,花费时间为 $O(TN^2)$;求矩阵的幂和乘法,花费时间为 $O(N^3)$.因此,几何平均度量学习方法花费总时间为 $O(TN^2+N^3)$.而多核几何平均度量学习方法的两部分时间消耗分别为 $O(hTN^2)$和 $O(hN^3)$,额外的二次方程带来的时间消耗为 $O(h^2)$.因为 h 是远小于 N 的,多核几何平均度量学习方法的时间为 $O(hTN^2+hN^3)$.通过以上分析不难看出,我们的方法不仅在扩展性方面优于几何平均度量学习方法,而且对于高维度数据的处理更高效.以上就是我们提出的针对于混杂数据的多核几何平均度量学习方法,总体而言,我们提出的算法框架将多种混杂的数据投射到可再生核希尔伯特空间中,然后利用加权组合来整合相应的度量.交替策略用于解决度量和权重的联合目标,通过第 4 节的实验结果证明算法是有效的.

4 实 验

在本节中,我们通过实验分析基于几何平均的混杂数据度量学习方法的性能.我们首先对数据集以及评估标准进行描述;然后,我们详细地对比较方法进行说明;最后,我们将基于几何平均的混杂数据度量学习方法与经典和最新的算法在分类精度方面进行比较.

我们选用了 UCI 中的 6 个混杂型数据集进行实验,使用软件为 Matlab.实验时,我们首先对数据集进行拆分,将上述 6 个数据集拆分为数值型数据集和符号型数据集,其基本特征描述见表 1.

Table 1 Basic description of the datasets

表 1 数据集基本描述

数据集	#标签类别数	#符号类别	#混杂型维度	#数值型维度	#符号型维度	#样本数
german	2	1~10	20	17	3	1 000
heart	2	0~4	13	8	5	270
hepatitis	2	1,2	19	13	6	155
horse	2	1~6,9	22	15	7	368
icu	3	0~2	20	16	4	200
veteranLungCancer	2	0~4	7	3	4	137

为了保证实验的公平性,我们统一规范了评价标准,对比算法的参数设置均使用了其最优参量.实验采用 10 折交叉验证,其中,90%用来训练,10%用来测试,记录格式为“平均值±均方差”.最后,结合对比算法对数据集分别进行 3 组实验,实验结果中,字体加粗并加下划线部分为全部实验的最优结果,仅加粗部分为仅次于最优的结果.

- 线性对比算法
  - ITML:一种提出了使用 LogDet 散度正则化的信息理论度量学习方法[20].
  - LMNN:一种基于马氏距离并使用 k 近邻定义约束方式的监督度量学习方法[22].
  - GMML:一种基于几何平均并具有封闭形式解决方案的监督度量学习方法[25].
  - DOUBLESVM:一种将度量学习问题变成一个样本对分类问题的监督度量学习方法[23].
- 核化对比算法
  - kITML:上述 ITML 的核化版本.使用高斯核函数来处理训练集与测试集.
  - kLMNN:上述 LMNN 的核化版本.在 LMNN 的基础上对数据集进行了核化处理.
  - kGMML:上述 GMML 的核化版本.将线性的几何平均度量学习方法转为非线性的核化度量学习方法.

第 1 组使用经典的线性度量学习算法对数值型数据集进行实验.分别采用 ITML[20],LMNN[22],GMML[25],DOUBLESVM[23]这 4 种线性度量学习方法对如下 6 个数据集进行实验.对于 DOUBLESVM[23]方法,我们依照算法描述,设置 k=1,惩罚因子 C<10.对于 GMML,我们设置参数λ为 0.1,t 保持在[0,1]区间内,设置步长为 0.1.这些对比方法均使用其算法最佳设置来得到算法最优的实验结果.具体实验结果见表 2.

Table 2 Linear algorithm results of numerical datasets

表 2 数值型数据集的线性算法结果

	ITML	LMNN	GMML	DOUBLESVM
german	0.5980±0.1891	0.6620±0.1311	0.6200±0.0548	0.6080±0.0476
heart	0.6963±0.2202	0.6074±0.1225	0.6667±0.1132	0.6037±0.0814
hepatitis	0.8516±0.2693	0.8133±0.0281	0.8633±0.0808	0.8750±0.0988
horse	0.6957±0.2200	0.6304±0.0126	0.6847±0.1142	0.6818±0.0571
icu	0.9050±0.2862	0.9260±0.0225	0.8406±0.0918	0.8511±0.0362
veteranLungCancer	0.6715±0.2124	0.7002±0.0637	0.6781±0.1200	0.6496±0.0456

第 2 组使用经典单核的度量学习算法对数值型数据集进行实验,对上述的 3 种线性方法均使用高斯核函数进行核化.在 kGMML 方法中,设置参数λ=0.45,t=0.5.kITML 反复迭代,直至其结果收敛.最终实验结果为 10 折交叉验证所得平均值与均方差,见表 3.

Table 3 Nonlinear algorithm results of numerical datasets

表 3 数值型数据集的非线性算法结果

	kITML	kLMNN	kGMML
german	0.6060±0.2534	0.6620±0.1273	0.6410±0.0507
heart	0.6296±0.1100	0.5963±0.0708	0.6630±0.0591
hepatitis	0.8323±0.2440	0.8133±0.0281	0.8733±0.0584
horse	0.6576±0.2305	0.6304±0.0126	0.7125±0.0602
icu	0.8900±0.3011	**0.9261±0.0225**	0.8373±0.0733
veteranLungCancer	0.7080±0.1953	0.6430±0.1295	0.6864±0.0808

第 3 组使用基于几何平均的混杂数据度量学习算法对混杂数据集和符号型数据集进行实验.符号型数据度量使用处理所得的符号型数据集,然后将其经汉明距离函数处理得到汉明矩阵进行实验.本文提出的算法对每种数据类型均设置了权重,若某种数据有较好的性能,则其在总体目标函数中所占的份量会较大.针对不同的数据集,其权重分配显然也会有所不同.实验结果见表 4.

Table 4 Multiple kernel geometric mean metric learning algorithm results of heterogeneous datasets

表 4 基于几何平均的混杂数据度量学习算法结果

	符号型数据度量	混杂数据度量	加权混杂数据度量
german	0.6940±0.0386	0.6780±0.0329	**0.7579±0.1146**
heart	0.7704±0.0834	0.7741±0.0881	**0.8037±0.0677**
hepatitis	0.8133±0.0422	0.8650±0.0631	**0.9217±0.0533**
horse	**0.9188±0.0473**	0.9079±0.0334	**0.9108±0.0351**
icu	0.8682±0.1285	0.8950±0.0797	**0.9108±0.0381**
veteranLungCancer	0.6466±0.1201	0.6662±0.1073	**0.6722±0.1105**

从实验结果可以看出,基于几何平均的度量学习算法在大部分数据集上都表现最优,说明对于不同数据集,使用不同的度量学习方法进行处理是很有必要的.

5 结 论

在本文中,我们提出一种基于几何平均的多核度量学习算法来解决混杂数据的度量学习问题.传统的度量学习旨在学习一个全局的线性度量,这种方法处理混杂数据是不合理的.我们利用了混杂数据之间的相同和互补的性质,所提出的方法与现有的大多数方法相比,学习指标与技术具有更优异的性能,我们的算法的时间复杂度仅与样本数有关,这在高维的数据集实验中可体现其优势.实验结果表明,基于几何平均的混杂数据度量学习算法是成功的.

References:

[1] Cover T, Hart P. Nearest neighbor pattern classification. IEEE Trans. on Information Theory, 1967,13(1):21–27. [doi: 10.1109/TIT.1967.1053964]

[2] Lloyd S. Least squares quantization in PCM. IEEE Trans. on Information Theory, 1982,28(2):129−137. [doi: 10.1109/TIT.1982.1056489]

[3] Xing EP, Ng AY, Jordan MI, Russell S. Distance metric learning with application to clustering with side-information. In: Proc. of the Int'l Conf. on Neural Information Processing Systems. MIT Press, 2002. 521−528.

[4] Shaw B, Huang B, Jebara T. Learning a distance metric from a network. In: Proc. of the Advances in Neural Information Processing Systems. 2011. 1899−1907.

[5] Taylor ME, Kulis B, Sha F. Metric learning for reinforcement learning agents. In: Proc. of the 10th Int'l Conf. on Autonomous Agents and Multiagent Systems. Int'l Foundation for Autonomous Agents and Multiagent Systems, 2011. 777−784.

[6] McFee B, Barrington L, Lanckriet G. Learning content similarity for music recommendation. IEEE Trans. on Audio, Speech, and Language Processing, 2012,20(6):2207−2218. [doi: 10.1109/TASL.2012.2199109]

[7] Yu Q, Gao Y, Huo J, Zhuang YK. Discriminative joint multi-manifold analysis for video-based face recognition. Ruan Jian Xue Bao/Journal of Software, 2015,26(8):2897−2911 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4894.htm [doi: 10.13328/j.cnki.jos.004894]

[8] Mensink T, Verbeek J, Perronnin F, Csurka G. Metric learning for large scale image classification: Generalizing to new classes at near-zero cost. In: Proc. of the Computer Vision (ECCV 2012). 2012. 488−501. [doi: 10.1007/978-3-642-33709-3_35]

[9] Li X, Shen C, Shi Q, Dick A, Hengel AVD. Non-Sparse linear representations for visual tracking with online reservoir metric learning. In: Proc. of the 2012 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). IEEE, 2012. 1760−1767. [doi: 10.1109/CVPR.2012.6247872]

[10] Xiong H, Chen X. Kernel-Based distance metric learning for microarray data classification. BMC Bioinformatics, 2006,7(1):1−11. [doi: 10.1186/1471-2105-7-299]

[11] Esposito F, Malerba D, Tamma V, Bock HH. Classical resemblance measures. In: Bock HH, Diday E, eds. Proc. of the Analysis of Symbolic Data: Exploratory Methods for Extracting Statistical Information from Complex Data. Berlin: Springer-Verlag, 2002. 139−152.

[12] Cost S, Salzberg S. A weighted nearest neighbor algorithm for learning with symbolic features. Machine Learning, 1993,10(1):57−78. [doi: 10.1023/A:1022664626993]

[13] Ienco D, Pensa RG, Meo R. Context-Based distance learning for categorical data clustering. In: Proc. of the Int'l Symp. on Intelligent Data Analysis. Berlin, Heidelberg: Springer-Verlag, 2009. 83−94. [doi: 10.1007/978-3-642-03915-7_8]

[14] Ienco D, Pensa RG, Meo R. From context to distance: Learning dissimilarity for categorical data clustering. ACM Trans. on Knowledge Discovery from Data (TKDD), 2012,6(1):1−25. [doi: 10.1145/2133360.2133361]

[15] Le SQ, Ho TB. An association-based dissimilarity measure for categorical data. Pattern Recognition Letters, 2005,26(12):2549−2557. [doi: 10.1016/j.patrec.2005.06.002]

[16] Ahmad A, Dey L. A method to compute distance between two categorical values of same attribute in unsupervised learning for categorical data set. Pattern Recognition Letters, 2007,28(1):110−118. [doi: 10.1016/j.patrec.2006.06.006]

[17] Chen J, Zhao Z, Ye J, Liu H. Nonlinear adaptive distance metric learning for clustering. In: Proc. of the 13th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. ACM Press, 2007. 123−132. [doi: 10.1145/1281192.1281209]

[18] He Y, Chen W, Chen Y, Mao Y. Kernel density metric learning. In: Proc. of the 2013 IEEE 13th Int'l Conf. on Data Mining (ICDM). IEEE, 2013. 271−280. [doi: 10.1109/ICDM.2013.153]

[19] Hoi SCH, Liu W, Lyu MR, Ma WY. Learning distance metrics with contextual constraints for image retrieval. In: Proc. of the 2006 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, Vol.2. IEEE, 2006. 2072−2078. [doi: 10.1109/CVPR.2006.167]

[20] Torresani L, Lee K. Large margin component analysis. In: Proc. of the Advances in Neural Information Processing Systems, Vol.19. 2007. 1385−1392.

[21] Davis JV, Kulis B, Jain P, Sra S, Dhillon IS. Information-Theoretic metric learning. In: Proc. of the 24th Int'l Conf. on Machine Learning. ACM Press, 2007. 209−216. [doi: 10.1145/1273496.1273523]

[22] Weinberger KQ, Saul LK. Distance metric learning for large margin nearest neighbor classification. Journal of Machine Learning Research, 2009,10(1):207−244.

[23] Wang F, Zuo W, Zhang L, Meng D. A kernel classification framework for metric learning. IEEE Trans. on Neural Networks and Learning Systems, 2015,26(9):1950−1962. [doi: 10.1109/TNNLS.2014.2361142]

[24] Chang CC, Lin CJ. LIBSVM: A library for support vector machines. ACM Trans. on Intelligent Systems and Technology (TIST), 2011,2(3):1−27. [doi: 10.1145/1961189.1961199]

[25] Zadeh P, Hosseini R, Sra S. Geometric mean metric learning. In: Proc. of the Int'l Conf. on Machine Learning. 2016. 2464−2471.

[26] Jain P, Kulis B, Dhillon IS. Inductive regularized learning of kernel functions. In: Proc. of the Advances in Neural Information Processing Systems. 2010. 946−954.

[27] Papadopoulos A. Metric Spaces, Convexity and Nonpositive Curvature. European Mathematical Society, 2014.

[28] Sra S, Hosseini R. Conic geometric optimization on the manifold of positive definite matrices. SIAM Journal on Optimization, 2015,25(1):713−739. [doi: 10.1137/140978168]

附中文参考文献:

[7] 于谦,高阳,霍静,庄韫恺.视频人脸识别中判别性联合多流形分析.软件学报,2015,26(8):2897−2911. http://www.jos.org.cn/1000-9825/4894.htm [doi: 10.13328/j.cnki.jos.004894]

齐忍(1993－),女,河北晋州人,硕士生,主要研究领域为度量学习,集成学习.

梁建青(1990－),女,博士生,主要研究领域为半监督学习,距离学习.

朱鹏飞(1986－),男,博士,副教授,主要研究领域为机器学习,计算机视觉.