

基于特征分组与特征值最优化的距离度量学习方法

赵永威¹ 张 蕾² 李弼程¹ 王挺进¹ 吕清秀¹

(1. 解放军信息工程大学信息系统工程学院, 郑州, 450002; 2. 郑州升达经贸管理学院, 郑州, 451191)

摘 要: 主流的距离度量学习方法都需要求解半正定规划 (Semi-definite programming, SDP) 问题, 而其中每次循环迭代中的矩阵完全特征分解运算使得现有方法计算复杂度很高, 实用性不强, 难以应用在大规模数据环境。本文提出了一种基于特征分组与特征值最优化的距离度量学习方法。引入特征分组算法, 根据特征各维数之间相关性对图像底层特征进行分组。在一定的约束条件下, 将求解 SDP 问题转化为特征值最优化问题, 在每次循环迭代中只需计算矩阵最大特征值对应的特征向量。实验结果表明该方法能有效地降低计算复杂度, 减少度量矩阵的学习时间, 并且能取得较好的分类结果。

关键词: 距离度量学习; 半正定规划; 特征分组; 特征值最优化; 度量矩阵

中图分类号: TP391 文献标志码: A

Distance Metric Learning Based on Feature Grouping and Eigenvalue Optimization

Zhao Yongwei¹, Zhang Lei², Li Bicheng¹, Wang Tingjin¹, Lü Qingxiu¹

(1. Institute of Information System Engineering, PLA Information Engineering University, Zhengzhou, 450002, China;
2. Shengda Economics Trade & Management College of Zheng Zhou, Zhengzhou, 451191, China)

Abstract: The current mainstream distance metric learning approaches that all need to solve the positive semi-definite programming problem (SDP) will lead to high computational complexity, and they are thus difficult to be applied to large scale datasets well because of fully matrix characteristics decomposition operational in each loop iteration. A distance metric learning method based on feature grouping and eigenvalue optimization is proposed considering the above problems. Firstly, a feature grouping algorithm is introduced to segment image features into several groups according to the correlations between each dimension of characteristics. Then, the SDP problem can be covered to eigenvalue optimization issue under some certain constraints. Therefore, only the maximum eigenvalues of matrix is needed in every loop iteration. Experiment results indicate that the computational complexity and the learning time of metric matrix are reduced effectively. Besides, the classification results are improved compared with the traditional methods.

Key words: distance metric learning; semi-definite programming (SDP); feature grouping; eigenvalue optimization; metric matrix

引 言

近年来,随着视觉词典模型^[1](Bag of visual words model, BoVWM)和视觉语言模型^[2,3](Visual language model, VLM)等模型的出现及其在图像检索^[4,5]、语义概念检测^[6]和近似重复图像检测^[7]等方面的广泛应用,使得计算机视觉相关领域的研究取得了较为明显的进步。然而,由于语义鸿沟问题的存在,导致目前的图像底层特征无法与相应的语义概念相对应。而语义鸿沟可从另一层面解释为度量空间的不统一,即底层特征空间和语义空间在距离度量上的不一致导致了鸿沟的出现。因此,为了更加准确地度量图像间的相似性,可以采用距离度量学习技术,也即是通过训练数据,获得一种能够反应样本空间结构信息或语义约束信息的线性或非线性的变换,得到一种区分性更好的度量空间,使图像底层特征空间和语义空间相一致。这样就能在很大程度上避免语义鸿沟造成的不利影响。

距离度量学习方法主要可分为无监督的距离度量学习、监督的距离度量学习及半监督的距离度量学习。其中,无监督的距离度量学习主要是通过学习一种潜在的低维流形,使观测数据的几何关系(如距离)被保留下来,多指一些降维方法。如主成分分析方法(Principle component analysis, PCA)^[8]、非负矩阵分解方法^[9]及等距映射方法^[10]等。这类方法在一定程度上能够提高样本的区分性,但是仍难以反应图像语义上的相似性。监督的距离度量学习的主要思想是利用事先标注好的数据学习一个度量矩阵,对样本进行映射变换,使得在变换后的度量空间中语义相近样本之间的距离最小化,语义不相关样本之间的距离最大化。如2003年Eric等^[11]提出的基于凸规划的全局距离度量学习方法(Global distance metric learning by convex programming, GDMLCP),其目标是在相似样本点间的距离小于某一阈值时学习一个度量矩阵使得非相似样本点间的距离最大化,但是该方法并没有考虑到样本之间的局部近邻关系及拓扑结构。而随之出现的近邻成分分析方法^[12](Neighborhood components analysis, NCA)则以概率的方式定义样本数据的软近邻,然后通过最大化训练样本的分类错误率学习距离度量矩阵。该方法在训练度量矩阵的同时,保持了相邻样本数据之间关系,但却不一定能达到全局最优。2007年Jason等^[13]提出了基于信息论的距离度量学习方法(Information-theoretic metric learning, ITML),其目标是在满足约束信息的同时,使学习到的度量矩阵 M 和根据某种先验知识给出的度量矩阵 M_0 之间的KL散度最小。而2009年Kilian等^[14]提出的最大边界近邻方法(Large margin nearest neighbor, LMNN)是通过对损失函数的约束将样本的 K 个最近邻保持在同一类别中,同时使异类样本之间的边界最大。针对考虑样本数据的几何结构缺失的问题,文献[15]提出了一种距离度量学习方法,在运用相似约束对和非相似约束对的同时,考虑了数据的几何结构。此外,2010年,Wu等^[16]提出了概率相关成分分析方法,利用概率信息作为标注信息进行距离度量学习。2013年,Yang等^[17]通过集成学习的方式构建了一种局部距离度量矩阵,在进行距离度量学习的同时考虑了样本局部的近邻关系。而针对有监督的距离度量学习方法中训练样本数据不足的问题,2009年,Zheng等^[18]提出了利用少量标记数据和辅助知识来学习一种鲁棒的距离度量。2010年,Steven等^[19]提出了一种拉普拉斯正则化距离度量学习方法,该方法将样本数据点的近邻看作相似点,联合已有标注数据学习距离度量,但是没有充分考虑到各近邻样本数据点的分布信息。虽然上述方法都能在一定程度上解决相应的问题,但是研究表明^[20],它们却存在一个共同的缺点就是计算复杂度高,导致其难以适用于当前日益增长的大规模数据环境。针对传统距离度量学习方法计算复杂度高,难以适用于大规模数据库的问题,本文提出了一种基于特征分组与特征值最优化的距离度量学习方法。

1 基于特征分组与特征值最优化的距离度量学习方法

引入一种特征分组算法,根据特征各维数之间的相关性对其进行分组;在相似样本约束对的距离小于一定阈值的条件下,将传统的求解SDP问题转化为特征值最优化问题,也即是只需计算每次循环迭

代中对称矩阵最大特征值所对应的特征向量,而不必对矩阵进行全分解;最后将各分组学习得到的度量矩阵进行融合得到全局度量矩阵,形成新的距离度量空间。具体流程如图1所示。

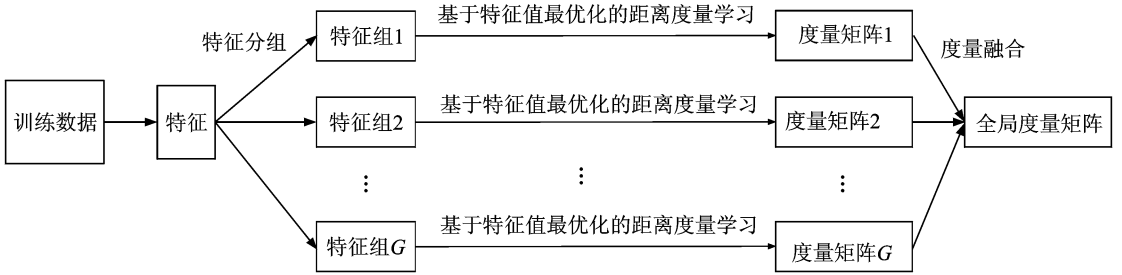


图1 基于特征分组与特征值最优化的距离度量学习方法流程

Fig. 1 Distance metric learning method based on feature grouping and eigenvalue optimization

1.1 特征分组算法

为了进一步降低计算复杂度,引入一种 Normalized Cuts (NCut)^[21] 特征分组算法将特征向量进行分解,把相关性较强的特征分为一组。设已知 l 个 K 类的训练样本: $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$, $y_l \in \{1, 2, \dots, K\}$, 则每个类别样本特征的相关矩阵为

$$C = \frac{1}{l} \sum_{k=1}^l (x_k - u)^T (x_k - u) \quad (1)$$

其中, $u = \frac{1}{l} \sum_{k=1}^l x_k$, 并将特征 x_i, x_j 的相关性定义为

$$R_{ij} = \frac{|C_{ij}|}{\sqrt{C_{ii}C_{jj}}} \quad (2)$$

式中 C_{ij} 为相关矩阵 C 的第 i 行第 j 列个元素。根据相关矩阵 R (R 的第 i 行第 j 个列元素为 R_{ij}), 用 NCut 分组算法将 d 维的初始特征 V 分成 G 组 $\{F_1, F_2, \dots, F_G\}$, F_1, F_2, \dots, F_G 均表示分组后的特征, NCut 分组算法是通过使式(3)最小化,求得 V 的一种分组方式

$$\text{NCut}_G = \frac{\text{cut}(F_1, V - F_1)}{\text{assoc}(F_1, V)} + \frac{\text{cut}(F_2, V - F_2)}{\text{assoc}(F_2, V)} + \dots + \frac{\text{cut}(F_G, V - F_G)}{\text{assoc}(F_G, V)} \quad (3)$$

其中

$$\text{cut}(F_k, V - F_k) = \sum_{i \in F_k, j \in V - F_k} R_{ij} \quad (4)$$

$$\text{assoc}(F_k, V) = \sum_{i \in F_k, j \in V} R_{ij} \quad (5)$$

式(4)表示特征组 F_k 中的特征和该组外的各维特征的相关性,式(5)表示特征组 F_k 中的特征和全部特征的相关性。NCut 分组算法具体步骤如下。

(1) 计算 $d \times d$ 对角阵 D , 对角元素 $D_{ii} = \sum_{j=1}^d R_{ij}$, R 为由 R_{ij} 组成的矩阵;

(2) 解广义特征值问题 $(D - R)x = \lambda Rx$, 解得特征值 $\{\lambda_1, \lambda_2, \dots, \lambda_d \mid \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_d\}$, 对应的特征向量为 $\{v_1, v_2, \dots, v_d\}$, 分别表示对 d 维特征的不同分组方式;

(3) 由步骤(2)知,可以利用特征向量 v_2 对特征进行 2 分组,也将 v_2 中大于 0 的元素对应维的特征分为一组,小于 0 的元素对应维的特征分为一组;

(4) 根据需要可以用后续特征向量 v_3, v_4, \dots 对分组后的特征按照步骤(3)进一步分组。

1.2 基于特征值最优化的距离度量学习

令 S^d 为 $d \times d$ 维对称矩阵构成的空间, S_+^d 为该空间内的半正定矩阵。特征值最优化的距离度量学习过程也即是使求解的度量矩阵 M 满足如下最优化问题

$$\begin{aligned} & \max_{M \in S_+^d} \min_{(i,j) \in D} d_M^2(\mathbf{x}_i, \mathbf{x}_j) \\ & \text{s. t. } \sum_{(i,j) \in S} d_M^2(\mathbf{x}_i, \mathbf{x}_j) \leq 1 \end{aligned} \quad (6)$$

用 $\text{Tr}(\cdot)$ 表示矩阵的迹, $\langle \cdot, \cdot \rangle$ 表示内积, 则对于任意 $X, Y \in \mathbf{R}^{d \times n}$ 的矩阵, 有 $\langle X, Y \rangle := \text{Tr}(X^T Y)$ 。令 $X_{ij} = (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T$, 那么对于任意矩阵 $M \in S_+^d$, $\mathbf{x}_i, \mathbf{x}_j$ 之间的距离可表示为 $d_M^2(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T M (\mathbf{x}_i - \mathbf{x}_j) = \langle X_{ij}, M \rangle$ 。将 X_{ij} 记为 X_τ , 并令 $X_S = \sum_{(i,j) \in S} X_{ij}$, 则式(6)可转化为

$$\begin{aligned} & \max_{M \in S_+^d} \min_{\tau \in D} \langle X_\tau, M \rangle \\ & \text{s. t. } \langle X_S, M \rangle \leq 1 \end{aligned} \quad (7)$$

和式(6)相比, 式(7)中的目标函数及约束条件与 M 成线性关系, 可进一步转化为特征值最优化问题, 亦即最小化一个对称矩阵的最大特征值。

对于任意 $X \in S^d$, 令 $\lambda_{\max}(X)$ 为矩阵 $X \in S^d$ 的最大特征值。设 D 为非相似约束对的个数, 定义 $\Delta = \{u \in \mathbf{R}^D : u_\tau \geq 0, \sum_{\tau \in D} u_\tau = 1\}$, 超球面 $P = \{M \in S_+^d : \text{Tr}(M) = 1\}$ 。假设 X_S 可逆, 则对于任意 $\tau \in D$, D 为样本集合, 令 $\tilde{X}_\tau = X_S^{-1/2} X_\tau X_S^{-1/2}$, 不难证明式(7)等价于下面的问题

$$\max_{S \in P} \min_{u \in \Delta} \sum_{\tau \in D} u_\tau \langle \tilde{X}_\tau, S \rangle \quad (8)$$

式中: $S = X_S^{1/2} \tilde{M}^* X_S^{1/2}$, M^* 为式(7)最优解, 那么式(8)可进一步表示为一个特征值最优化问题

$$\max_{S \in P} \min_{u \in \Delta} \langle \sum_{\tau \in D} u_\tau \tilde{X}_\tau, S \rangle = \min_{u \in \Delta} \lambda_{\max} \left(\sum_{\tau \in D} u_\tau \tilde{X}_\tau \right) \quad (9)$$

证明 假设 M^* 为式(7)的最优解, 令 $\tilde{M}^* = \frac{M^*}{\langle X_S, M^* \rangle}$, 则 $\langle X_S, \tilde{M}^* \rangle = 1$ 。由约束条件 $\langle X_S, M \rangle \leq 1$ 可知, 下列不等式成立

$$\min_{\tau \in D} \langle X_\tau, \tilde{M}^* \rangle = \frac{\min_{\tau \in D} \langle X_\tau, M^* \rangle}{\langle X_S, M^* \rangle} \geq \min_{\tau \in D} \langle X_\tau, M^* \rangle \quad (10)$$

即 \tilde{M}^* 同样是最优解, 故式(7)等价于

$$\arg \max_{M \in S_+^d, \langle X_S, M \rangle = 1} \{ \min_{\tau \in D} \langle X_\tau, M \rangle \} \quad (11)$$

由 $S = X_S^{1/2} \tilde{M}^* X_S^{1/2}$, $\langle X_S, \tilde{M}^* \rangle = 1$ 可知 $S \in P$, 通过变换式 $\tilde{X}_\tau = X_S^{-1/2} X_\tau X_S^{-1/2}$, 可得 $\langle X_\tau, \tilde{M}^* \rangle = \langle \tilde{X}_\tau, S \rangle$, 又由 $\min_{\tau \in D} \langle X_\tau, X \rangle = \min_{u \in \Delta} \sum_{\tau \in D} u_\tau \langle X_\tau, X \rangle$, 进一步可知式(11)和式(8)等价。

由文献[22]可知, $\max_{S \in P} \langle X, S \rangle = \lambda_{\max}(X)$, 进一步得

$$\max_{S \in P} \min_{u \in \Delta} \langle \sum_{\tau \in D} u_\tau \tilde{X}_\tau, S \rangle = \min_{u \in \Delta} \lambda_{\max} \left(\sum_{\tau \in D} u_\tau \tilde{X}_\tau \right) \quad (12)$$

在上述论述中, 假设 X_S 可逆。实际中, 通过给 X_S 对角线元素加上一个很小的数使其可逆, 即 $X_S \leftarrow X_S + \delta I_d$, I_d 是单位阵, $\delta > 0$ 是一个非常小的常量。因为当输入空间的维数 d 很高时, 计算 $X_S^{-1/2}$ 的算法复杂度比较高, 因此和直接对矩阵求逆相比, 用 Cholesky 分解更快而且更稳定。由 Cholesky 分解得 $X_S = LL^T$, L 是一个下三角矩阵, 且对角线元素都为正。因此, 在式(11)中, 令 $S = L^T \tilde{M}^* L$ 可得最优解度量矩阵 $\tilde{M}^* = (L^{-1})^T S L^{-1}$ 。由上述过程不难看出, 对于任意 $\tau = (i, j) \in D$, $\tilde{X}_\tau = (L^{-1}(\mathbf{x}_i - \mathbf{x}_j))^T (L^{-1}(\mathbf{x}_i - \mathbf{x}_j))^T$, 式(11)和式(12)仍然成立。

1.3 特征值最优化问题的求解

由上述证明可知, 特征值最优化距离度量学习模型等价于下面的问题

$$\max_{\mathbf{S} \in P} f(\mathbf{S}) = \max_{\mathbf{S} \in P} \min_{u \in \Delta} \sum_{\tau \in D} u_{\tau} \langle \tilde{\mathbf{X}}_{\tau}, \mathbf{S} \rangle \quad (13)$$

为了求解式(13),引入一个平滑因子 $\mu > 0$, 定义

$$f_{\mu}(\mathbf{S}) = \min_{u \in \Delta} \sum_{\tau \in D} u_{\tau} \langle \tilde{\mathbf{X}}_{\tau}, \mathbf{S} \rangle + \mu \sum_{\tau \in D} u_{\tau} \ln u_{\tau} \quad (14)$$

由平滑理论可知^[23]

$$f_{\mu}(\mathbf{S}) = -\mu \ln \left(\sum_{\tau \in D} e^{-\langle \tilde{\mathbf{X}}_{\tau}, \mathbf{S} \rangle / \mu} \right) \quad (15)$$

采用平滑后的最优化问题 $\max_{\mathbf{S} \in P} f_{\mu}(\mathbf{S})$ 近似式(13), 然后就可利用 Frank-Wolfe 算法^[24] 求解特征值之最优化问题, 也即求解 \mathbf{S} 。具体步骤可描述如下:

- (1) 输入平滑参数 $\mu (\mu > 0)$, 误差门限 ϵ , 定义步长 α_t , $\{\alpha_t \in (0, 1) : t \in N\}$;
- (2) 初始化参数 $\mathbf{S}_1^{\mu}, \mathbf{S}_1^{\mu} \in \mathbf{S}_+^d$ 且有 $\text{Tr}(\mathbf{S}_1^{\mu}) = 1$;
- (3) 求使 $f_{\mu}(\mathbf{S}_i^{\mu}) + \langle \mathbf{Z}, \nabla f_{\mu}(\mathbf{S}_i^{\mu}) \rangle$ 取最大值的 $\mathbf{Z}_i^{\mu}, \mathbf{Z}_i^{\mu} \in \mathbf{S}_+^d, \text{Tr}(\mathbf{Z}_i^{\mu}) = 1$, 其中 $\mathbf{Z}_i^{\mu} = \mathbf{v}\mathbf{v}^T$, \mathbf{v} 为矩阵 $\nabla f_{\mu}(\mathbf{S}_i^{\mu})$ 的最大特征值对应的特征向量;
- (4) 更新矩阵 $\mathbf{S}_{i+1}^{\mu} = (1 - \alpha_i)\mathbf{S}_i^{\mu} + \alpha_i\mathbf{Z}_i^{\mu}$;
- (5) 判断 $|f_{\mu}(\mathbf{S}_{i+1}^{\mu}) - f_{\mu}(\mathbf{S}_i^{\mu})|$ 是否小于误差门限 ϵ , 若小于则结束, 否则跳到步骤(3);
- (6) 输出 $d \times d$ 维矩阵 $\mathbf{S}_i^{\mu}, \mathbf{S}_i^{\mu} \in \mathbf{S}_+^d$ 。

在上述步骤中, 因为 $f_{\mu}(\mathbf{S})$ 是一个平滑的函数, 所以它的梯度 $\nabla f_{\mu}(\mathbf{S}) = \frac{\sum_{\tau \in D} e^{-\langle \tilde{\mathbf{X}}_{\tau}, \mathbf{S} \rangle / \mu} \tilde{\mathbf{X}}_{\tau}}{\sum_{\tau \in D} e^{-\langle \tilde{\mathbf{X}}_{\tau}, \mathbf{S} \rangle / \mu}}$ 是 Lipschitz 连续的, 所以上述算法收敛, 且算法复杂度为 $O(d^2/\epsilon^2)$ 。最后可得相应的度量矩阵 $\mathbf{M} = (\mathbf{L}^{-1})^T \mathbf{S} \mathbf{L}^{-1}$ 。

1.4 度量融合

假设将特征分为 G 组, 对第 g 组特征, 令利用特征值最优化距离度量学习方法得到的度量矩阵为 \mathbf{M}_g , 利用该度量矩阵可以得到两个样本在该组特征下的距离

$$d_g^2(\mathbf{x}_i^g, \mathbf{x}_j^g) = (\mathbf{x}_i^g - \mathbf{x}_j^g)^T \mathbf{M}_g (\mathbf{x}_i^g - \mathbf{x}_j^g) \quad (16)$$

式中: $\mathbf{x}_i^g, \mathbf{x}_j^g$ 表示对初始特征 $\mathbf{x}_i, \mathbf{x}_j$ 分组后对应的第 g 组特征。进一步, 对 G 个特征组的距离平方和进行线性合成

$$d^2(\mathbf{x}_i, \mathbf{x}_j) = \sum_{g=1}^G \omega_g (\mathbf{x}_i^g - \mathbf{x}_j^g)^T \mathbf{M}_g (\mathbf{x}_i^g - \mathbf{x}_j^g) \quad (17)$$

整个度量矩阵为

$$\mathbf{M} = \text{diag}(\omega_1 \mathbf{M}_1, \omega_2 \mathbf{M}_2, \dots, \omega_G \mathbf{M}_G) \quad (18)$$

设样本有 K 类, 合适的度量方式求得的类间方差比较大, 类内方差比较小。因此, 本文将各子度量矩阵的权值定义为类间方差与类内方差之比, 并归一化

$$\omega_g = \frac{\sum_{k=1}^K (\mathbf{u}_k - \mathbf{u})^T \mathbf{M}_g (\mathbf{u}_k - \mathbf{u})}{\sum_{k=1}^K \sum_{(x_i, y_i=k)} (\mathbf{x}_i - \mathbf{u}_k)^T \mathbf{M}_g (\mathbf{x}_i - \mathbf{u}_k)} \quad (19)$$

$$\sum_{g=1}^G \frac{\sum_{k=1}^K (\mathbf{u}_k - \mathbf{u})^T \mathbf{M}_g (\mathbf{u}_k - \mathbf{u})}{\sum_{k=1}^K \sum_{(x_i, y_i=k)} (\mathbf{x}_i - \mathbf{u}_k)^T \mathbf{M}_g (\mathbf{x}_i - \mathbf{u}_k)}$$

式中: \mathbf{u}_k 为第 k 类样本的均值, 若将式(19)代入式(17)就得到了两个样本之间的距离。

不难看出, 通过上述分组与融合的过程, 能够较大地降低传统距离度量学习方法的计算复杂度。假

设分解前特征的维数是 d , 基于特征值最优化的距离度量学习方法的算法复杂度为 $O(d^2/\epsilon^2)$, 将特征分组为维数相同的 G 组特征后, 经过特征分组后算法的复杂度就变为 $O(G(d/G)^2/\epsilon^2) = O(d^2/(\epsilon^2 G))$, 降低为原来的 $1/G$ 。需要注意的是, 实际中按相关性进行分组时, 各组特征维数未必相同, 所以算法复杂度会降低, 但至多能降低到原来的 $1/G$ 。

2 实验结果与性能分析

2.1 实验设置与性能评价

为了验证本文方法的有效性, 在内存为 2 GB, 主频为 2.4 GHz 的 PC 机上进行了 3NN 图像分类的实验。实验数据集为包括海滩、建筑、恐龙等 10 类共 1 000 幅图像的 Corel 图像集。图 2 给出了部分图像示例。首先, 从每类图像中选取 30 幅图像, 共 300 图像作为训练数据, 其余作为测试数据。然后, 采用文献[25]中的方法来提取图像特征, 也即通过计算 YCbCr 空间下图像的 DCT 块内像素的均值 u 和标准方差 σ 得到 58 维直方图作为表征图像内容的底层特征。图 3 给出了 Corel 图像库中一幅公交车图像及其相应的直方图特征, 具体过程这里不再赘述。



图 2 Corel 数据库的各类图像示例

Fig. 2 Image samples of Corel database

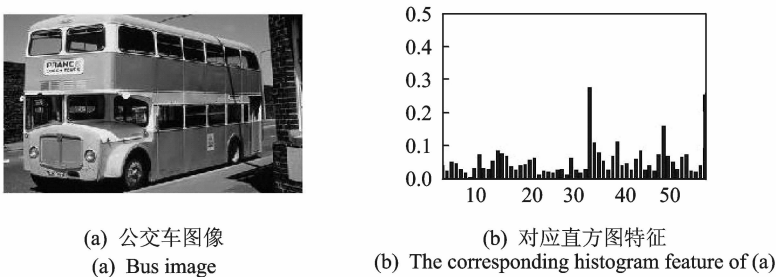


图 3 公交车图像和对应直方图特征

Fig. 3 Bus image and its corresponding histogram feature

实验性能评价指标为距离度量学习方法的学习时间和 3NN 分类的错误率, 分类错误率的定义为

$$\text{错误率} = \frac{\text{错误分类图像数}}{\text{测试图像总数}} \times 100\% \quad (20)$$

2.2 实验结果与分析

为了更好地表达特征的相关性, 将特征各维数之间的相关性转化为灰度图像。如图 4 所示, 其中亮度值越大表示相关性越强, 亮度值越小表示相关性越弱。亮度值为 255 表示相关度为 1, 亮度值为 0 表示相关度为 0。

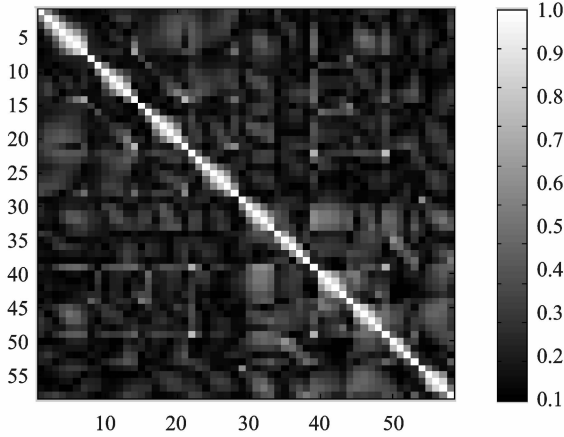


图4 特征各维相关性示意图

Fig. 4 Schematic diagram of each dimension correlation

实验首先分析了特征分组数对学习时间和分类错误率的影响,结果图5所示。由图5结果不难看出,算法的学习时间随着分组数的增加而降低。而分类错误率却随着分组数的增加呈现出增长的趋势。这是因为尽管特征分组算法依据各维数之间相关性进行分组,但是随着分组数的增加,难免会降低各组之间的相关性,进而减弱原有特征对图像表达的准确性,因而会在一定程度上增大分类错误率。因此,在实际应用中需要选择合适的分组数。

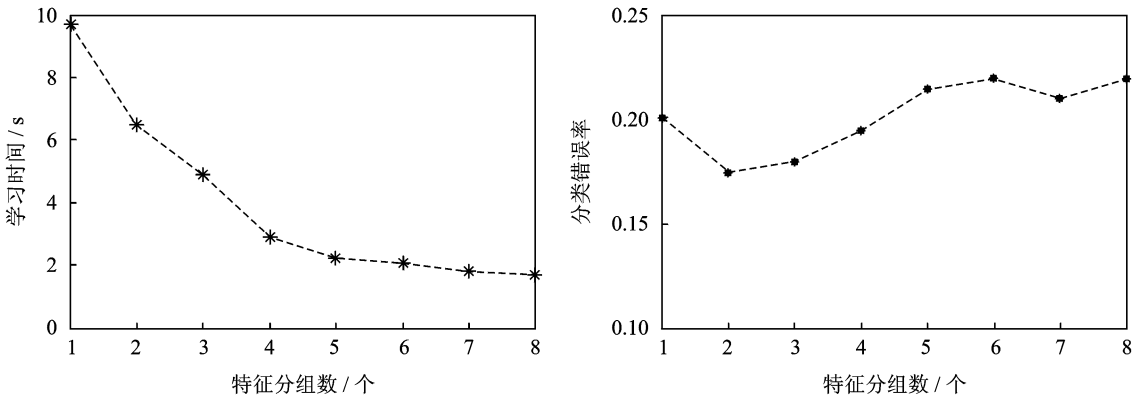


图5 特征分组数对学习时间和分类错误率的影响

Fig. 5 The effects of feature grouping number on learning time and classification error rate

将本文方法(Eigenvalue optimization distance metric learning + NCut, EODML + Ncut)与经典的近邻成分分析方法^[10](Neighborhood components analysis, NCA)、基于信息论的距离度量学习方法^[11](Information theoretic metric learning, ITML)距离度量学习方法、最大边界近邻方法^[12](Large margin nearest neighbor, LMNN)、未使用特征分组的特征值最优化度量学习方法(Eigenvalue optimization distance metric learning, EODML)及传统欧氏距离在Corel数据库上的分类结果及学习时间做了比较,如表1所示,其中的分类错误率和学习时间值都是经过10次实验后的平均值。

由表1结果可知,与传统的欧氏距离度量相比,采用的距离度量学习方法之后均能在一定程度上降

低分类错误率,同样地也会消耗一定的学习时间。然而,在取得相当分类结果的情况下,ITML方法和LMNN方法的学习时间要远低于NCA方法。本文提出的EOML方法则能以更小的学习时间取得最优的分类结果,且在引入特征分组算法以后能进一步降低学习时间,因而本文方法具有较强的实用性。

表1 不同距离度量学习方法的分类错误率及学习时间

Table 1 The classification error rates and learning time of different distance metric learning methods

方法	分类错误率	学习时间/s
Euclidean	27.9	—
NCA	20.6	95.5
ITML	24.3	37.4
LMNN	21.0	11.3
EOML	20.1	9.7
EOML+Ncut(分2组)	17.5	6.5

3 结束语

本文提出了一种基于特征分组与特征值最优化的距离度量学习方法。它将传统的距离度量学习方法中的求解半正定规划问题转化为特征值最优化问题,进而使得在每次循环迭代中只需求解对称矩阵的最大特征值所对应的特征向量,而不必对矩阵进行完全分解。引入了一种Ncut特征分组算法,进一步降低了距离度量学习方法的计算复杂度。在Corel数据库上的图像分类实验有效地验证了该方法能够缩短度量矩阵的学习时间且能在一定程度上改善图像分类结果,有效地增强了距离度量学习方法的实用性。需要指出的是一个合适的分组数目是取得较好分类结果的前提。

参考文献:

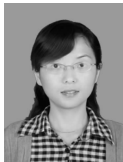
- [1] Sivic J, Zisserman A. Video Google: A text retrieval approach to object matching in videos[C]//Proc of 9th IEEE International Conference on Computer Vision. Nice, France: [s. n.], 2003:1470-1477.
- [2] Wu Lei, Li Mingjing, Li Zhiwei, et al. Visual language modeling for image classification[C]//Proc of International Workshop on Multimedia Information Retrieval (MIR'07). Augsburg, Germany: [s. n.], 2007:115-124.
- [3] Wu Lei, Hu Yang, Li Mingjing, et al. Scale-invariant visual language modeling for object categorization[J]. IEEE Transactions on Multimedia, 2009,11(2):286-294.
- [4] Jurie F, Triggs B. Creating efficient codebooks for visual recognition[C]//Proc of International Conference on Computer Vision. Beijing, China: [s. n.], 2005:604-610.
- [5] Jegou H, Douze M, Schmid C. Packing bag-of features[C]//Proc of IEEE 12th International Conference on Computer Vision (ICCV). Kyoto, Japan: [s. n.], 2009:2357-2364.
- [6] 万建平,彭天强,李弼程.基于证据理论的视频语义概念检测[J].数据采集与处理,2011,26(5):536-541.
Wan Jianping, Peng Tianqiang, Li Bicheng. Video semantic concept detection based on evidence theory[J]. Journal of Data Acquisition and Processing, 2011,26(5),536-541.
- [7] 蒯博宇,郭志刚,李弼程,等.基于随机化映射和模式熵的近似重复图像检测[J].数据采集与处理,2012,27(3):363-367.
Lin Boyu, Guo Zhigang, Li Bicheng, et al. Near-duplicate image detection based on random mapping and pattern entropy [J]. Journal of Data Acquisition and Processing, 2012,27(3):363-367.
- [8] Hotelling H. Analysis of a complex of statistical variables into principal components[J]. Journal of Educational Psychology, 1933,24(6):417-441.
- [9] Lee D, Seung H S. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999,401(6755):788-791.
- [10] Tenenbaum J, Silva V D, Langford J. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000,290(5500):2319-2323.
- [11] Eric P X, Andrew Y N, Michael I J, et al. Distance metric learning with application to clustering with side information[J]. Advances in Neural Information Processing Systems, 2002,15:505-512.

- [12] Jacob G, Sam R, Geoff H, et al. Neighborhood components analysis[C]//Advances in Neural Information Processing Systems. Washington, MIT Press, 2004;13-18.
- [13] Jason V D, Brian K, Prateek J, et al. Information theoretic metric learning[C]//Proceedings of the International Conference on Machine Learning. Florida, USA; [s. n.], 2007;209-216.
- [14] Kilian Q W, John B, Lawrence K S. Distance metric learning for large margin nearest neighbor classification[J]. *Journal of Machine Learning Research*, 2009(10):207-244.
- [15] Mahdih S, Shouraki S B. Semi-supervised metric learning using pairwise constraints[C]//Proceedings of the 21st international joint conference on Artificial intelligence. Pasadena, California, US; [s. n.], 2009;1217-1222.
- [16] 吴磊. 视觉语言分析:从底层视觉特征表达到语义距离学习[D]. 合肥:中国科学技术大学,2011.
Wu Lei. The analysis of visual language expression; From low-level visual features to semantic distance learning[D]. Hefei: University of Science and Technology of China, 2011.
- [17] Yang Mu, Wei Bing. Local discriminative distance metrics ensemble learning[J]. *Pattern Recognition*, 2013,46(8):2337-2349.
- [18] Zheng Junzha, Tao Mei, Wang Meng, et al. Robust distance metric learning with auxiliary knowledge[C]//Proceedings of the 21st international joint conference on Artificial intelligence. Pasadena, California, US; [s. n.], 2009;1327-1332.
- [19] Steven H, Liu W, Chang S F. Semi-supervised distance metric learning for collaborative image retrieval and clustering[J]. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2010,6(3):1-18.
- [20] Ying Y, Li P. Distance metric learning with eigenvalue optimization[J]. *Journal of Machine Learning Research*, 2012,13:1-26.
- [21] 刘博. 距离测度学习理论与应用研究[D]. 合肥:中国科学技术大学,2010.
Liu Bo. Research on the theory and application of the distance metric learning[D]. Hefei: University of Science and Technology of China, 2010.
- [22] Overton L. On minimizing the maximum eigenvalue of a symmetric matrix[J]. *SIAM Journal on Matrix Analysis and Application*, 1988,9(2):256-268.
- [23] Nesterov Y. Smoothing technique and its applications in semi-definite optimization[J]. *Mathematical Programming*, 2007, 110(2):245-259.
- [24] Hazan E. Sparse approximation solutions to semi-definite programs[C]//Proceedings of the 8th Latin American Conference on Theoretical informatics. Berlin Heidelberg; Springer, 2008;306-316.
- [25] Guo Canfeng, Jiang Jianmin. JPEG compressed image retrieval via statistical features[J]. *Pattern Recognition*, 2003,36(4): 977-985.

作者简介:



赵永威(1988-),男,博士研究生,研究方向:图像检索及分类,E-mail: zhaoyongwei369@163.com。



张蕾(1975-),女,讲师,研究方向:计算机应用及电子通信,E-mail:zhanglei_hnsd@163.com。



李弼程(1970-),男,博士,教授,博士生导师,研究方向:智能信息处理,E-mail: lbclm@163.com。



王挺进(1988-),男,硕士研究生,研究方向:图像分类,E-mail: wtj881021@126.com。



吕清秀(1986-),男,硕士研究生,研究方向:图像处理,E-mail: lvqxhz87@163.com。