

基于 CUR 矩阵分解的多核学习正则化路径近似算法

王梅, 李董, 薛成龙

(东北石油大学计算机与信息技术学院, 大庆, 163318)

摘要: 多核学习在解决不规则、大规模数据问题时表现出良好的优越性。正则化路径是一种多次求解多核学习, 选择最优模型的措施。针对多核学习正则化路径算法处理大规模数据时, 核矩阵规模较大, 计算代价高, 影响优化模型效率的问题, 提出一种基于 CUR 矩阵分解的多核学习正则化路径近似算法 (Multiple kernel learning regularization path approximation algorithm with CUR, MKLRPCUR)。该算法首先采用 CUR 算法获得核矩阵的低秩近似矩阵的多个分解矩阵, 然后在求解过程中利用低维的分解矩阵相乘替代核矩阵, 调整相关矩阵计算的顺序, 从而简化算法中核矩阵和拉格朗日乘子向量乘积的计算。MKLRPCUR 算法降低了矩阵的计算规模, 优化了矩阵计算, 提高了精确算法的计算效率。从理论上分析低秩近似矩阵的相对误差和算法的时间复杂度, 验证了近似算法的合理性。同时, 在 UCI 数据集、ORL 和 COIL 图像数据库上的实验结果表明, 本文提出的近似算法不仅保证了学习的准确率, 并且降低了算法的运行时间, 提高了模型的效率。

关键词: 多核学习正则化路径; 核矩阵; 矩阵近似; 抽样方法; CUR 算法

中图分类号: TP391 **文献标志码:** A

Multiple Kernel Learning Regularization Path Approximation Algorithm Based on CUR Matrix Decomposition

WANG Mei, LI Dong, XUE Chenglong

(School of Computer and Information Technology, Northeast Petroleum University, Daqing, 163318, China)

Abstract: Multiple kernel learning shows good superiority in solving irregular and large-scale data problems. Regularization path is a method to select the optimal model by solving the multiple kernel learning multiple times. Aiming at the problems that the kernel matrix size is large, the computational cost is high and the efficiency of the optimization model is affected when multiple kernel learning regularization path processes large-scale data, a multiple kernel learning regularization path approximation algorithm based on CUR matrix decomposition is proposed, which is named MKLRPCUR. This algorithm firstly adopts CUR algorithm to obtain multiple decomposition matrices of low-rank approximation matrix of kernel matrix. Then, in the solution process, the low-dimensional decomposition matrices are used to replace the kernel matrix, and the order of the correlation matrix calculation is adjusted, thereby simplifying the calculation of the kernel matrix and the Lagrange multiplier vector product. MKLRPCUR algorithm reduces the calculation scale of matrix, optimizes matrix calculation, and improves the

calculation efficiency of exact algorithm. The relative error of the low-rank approximation matrix and the time complexity of the algorithm are theoretically analyzed to verify the rationality of the approximation algorithm. At the same time, the experimental results on the UCI dataset, ORL and COIL image databases show that the proposed approximate algorithm not only ensures the accuracy of learning, but also reduces the running time of the algorithm and improves the efficiency of the model.

Key words: multiple kernel learning regularization path; kernel matrix; matrix approximation; sampling method; CUR algorithm

引 言

核方法是学习理论中处理非线性问题的办法。核方法首先将低维空间的非线性数据映射到高维空间或无限维空间,然后在映射的空间中实现线性可分。然而针对目前异构、不规则的非线性大规模多维数据,不同的核函数映射会产生较大的波动,单个核函数难以合理映射,同时也没有强大的理论基础辅助选择合适的核函数,鉴于单核函数学习能力不足的问题提出了多核学习(Multiple kernel learning, MKL)方法^[1-3]。多核学习的思想是将多个基本核函数以各种形式组合起来,处理异构、不规则数据和多数据源问题,避免单个核函数处理数据精度不高的问题。目前多核学习提高分类的优越性已经得到广泛认可,同时多核学习的求解需要考虑正则化参数的影响,计算正则化参数所有值的解的集合也被广泛讨论。

正则化路径算法是在路径跟踪理论上发展起来的一种求解参数规划问题的算法^[4-7],可以求解正则化参数取所有可能值时对应的优化问题的解,从而找到优化模型的最优解。Hastie等^[8]最早将正则化路径引入支持向量机(Support vector machine, SVM),提出了SVMpath算法,划分训练样本集,检测活动集随正则化参数变化的情况。SVMpath算法突破了SVM模型求解时需要提前设定正则化参数的限制,在相当于一次SVM求解的复杂度内求解并记录从 $+\infty$ 到0的所有正则化参数和对应的解路径,由此构建完整的正则化路径。根据SVMpath的优点,Bach提出了多核学习正则化路径(Multiple kernel learning regularization path, MKLRP)算法^[4],该路径利用数值延拓技术得到,其时间复杂度是求解正则化参数为一个值时问题复杂度的常数倍。多核学习正则化路径的求解涉及核矩阵的相关计算。核方法中的核矩阵包含着全部数据信息,是储存和计算的核心。多核学习中核矩阵是多个基本核矩阵的组合,数据规模大时,核矩阵的储存空间和计算复杂度都达到令人难以接受的地步,制约着多核学习的发展。如何简化矩阵的计算,提高正则化路径的效率,是目前亟待解决的问题。

国内外很多针对矩阵近似或分解的研究,如矩阵低秩分解近似、矩阵抽样降维等^[9]。矩阵抽样降维常用的方法有Nystrom算法、Monte Carlo算法和CUR算法等。CUR算法基于抽样技术,是一种简单但使用广泛的矩阵近似方法,适用于任何矩阵,典型的算法是标准CUR算法和快速CUR算法^[10]。CUR算法有很多应用:CUR算法应用于大规模推荐算法中,能降低算法的运行时间,极大减少内存的占用^[11-12];利用CUR算法对肿瘤基因数据矩阵进行行列双向聚类,运行时间要比奇异值矩阵分解的速度快数倍^[11]。矩阵低秩分解常用的方法有QR分解、Cholesky分解、奇异值分解(Singular value decomposition, SVD)分解以及非负矩阵因子分解(Non-negative matrix factorization, NMF)分解等^[9,13-14]。很多研究将矩阵分解方法应用于矩阵计算中,例如,基于NMF的低秩矩阵恢复模型^[14],把NMF分解运用到低秩矩阵恢复中,避免求解大规模矩阵核范数的SVD问题;基于Cholesky分解的K2DPCA算法^[15],将Cholesky算法用于分解近似矩阵,避免大规模计算矩阵的特征值和特征向量问题等。这些研究使用矩阵抽样方法或者矩阵分解方法将高维矩阵转变为低维矩阵,将高维矩阵计算转化为计算小规模矩阵的特征值、奇异值等问题。

考虑到CUR矩阵分解算法具有广泛适用性,低维的特点和多核学习正则化路径较大的计算规模,

本文提出了一种基于CUR矩阵分解的多核学习正则化路径近似算法(Multiple kernel learning regularization path with CUR, MKLRPCUR)。该算法在有关核矩阵的计算中,引入CUR算法获得近似矩阵的低维分解矩阵,降低了核矩阵相关计算复杂度,提高了学习模型的效率。

1 背景介绍

1.1 多核学习正则化路径

1.1.1 多核学习正则化

考虑一般的二分类正则化学习优化问题。假设训练样本集

$$T = \{(\mathbf{x}_i, y_i)\} \quad i = 1, \dots, n \quad (1)$$

式中: $\mathbf{x}_i \in X \subseteq \mathbf{R}^p$ 为输入空间 X 的输入向量; $y_i \in Y = \{-1, 1\}$ 为 \mathbf{x}_i 的相应输出值, n 为训练集样本的规模。多核学习最基本的构造是多个基本核函数的线性组合,其中最常用的是加权求和核函数。现有 m 个映射

$$\Phi_l: \mathbf{R}^p \rightarrow \mathcal{H}_l \quad l = 1, 2, \dots, m \quad (2)$$

将样本空间 \mathbf{R}^p 映射到 m 个特征空间中 $(\mathcal{H}_1, \mathcal{H}_2, \dots, \mathcal{H}_m)$,令映射为

$$\Phi(\mathbf{x}) = (\Phi_1(\mathbf{x}), \Phi_2(\mathbf{x}), \dots, \Phi_m(\mathbf{x})) \quad (3)$$

参数向量为 $\mathbf{w} = (w_1, w_2, \dots, w_m)$,核组合系数为 d_l ,即

$$\sum_{l=1}^m d_l = 1 \quad l = 1, 2, \dots, m \quad (4)$$

由此得到多核学习块 L_1 -范数正则化形式^[4]

$$\min_{\mathbf{w} \in \mathcal{H}_1 \times \dots \times \mathcal{H}_m} \sum_{i=1}^n L(y_i, \mathbf{w}^T \Phi(\mathbf{x}_i)) + \lambda \sum_{l=1}^m d_l \|\mathbf{w}_l\| \quad (5)$$

式中: $\|\mathbf{w}\| = \sqrt{\mathbf{w}^T \mathbf{w}}$ 为 \mathbf{w}_l 的欧氏范数, $L(\cdot)$ 为损失函数, $\lambda > 0$ 为正则化参数。

多核学习的损失函数采用

$$L_\sigma(y, y') = \frac{1}{\sigma} \log_2(1 + \exp((1 - yy')\sigma)) \quad (6)$$

L_σ 为严格凸函数,在变量 y' 上连续二次可导,其Fenchel共轭函数为

$$\psi(y, \nu) = \begin{cases} (1 + \nu y) \log_2(1 + \nu y) - \nu y \log_2(-\nu y) & \nu y \in (-1, 0) \\ +\infty & \text{其他} \end{cases} \quad (7)$$

多核学习正则化进而表示为

$$\min_{\mathbf{w} \in \mathcal{H}_1 \times \dots \times \mathcal{H}_m} \sum_{i=1}^n L_\sigma(y_i, \mathbf{w}^T \Phi(\mathbf{x}_i)) + \lambda \sum_{l=1}^m d_l \|\mathbf{w}_l\| \quad (8)$$

令 $k_l(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi_l(\mathbf{x}_i), \Phi_l(\mathbf{x}_j) \rangle$ 表示映射 Φ_l 对应的核函数,由锥约束得到优化问题式(8)的对偶问题为

$$\begin{aligned} \max_{\alpha} & - \sum_{i=1}^n \psi(y_i, \lambda \alpha_i) \\ \text{s.t.} & \quad \forall l, \alpha^T \mathbf{K}_l \alpha \leq d_l^2 \end{aligned} \quad (9)$$

\mathbf{K}_l 为第 l 个核函数 k_l 相关联的核矩阵

$$(\mathbf{K}_l)_{ij} = k_l(\mathbf{x}_i, \mathbf{x}_j) \quad (10)$$

为了跟踪所有正则化路径,采用数值正则化技术,利用凸对数障碍法将约束问题转化为非约束问题,式(9)的等价问题为

$$F(\alpha, \lambda) = \min_{\alpha} \sum_{i=1}^n \frac{1}{\lambda^2} \psi(y_i, \lambda \alpha_i) - \frac{\mu}{2\lambda} \sum_{l=1}^m \log_2(d_l^2 - \alpha^T \mathbf{K}_l \alpha) \quad (11)$$

式中: μ 为值较小的固定常数。对于任意给定的正则化参数 λ , $F(\alpha, \lambda)$ 在可行域 $\{\alpha, \forall l, \alpha^T K_l \alpha < d_l^2\}$ 为严格凸光滑函数, 则全局最优解唯一且可由 $\frac{\partial F}{\partial \alpha} = 0$ 得到。令 $\eta_l(\alpha) = \frac{\mu}{d_l^2 - \alpha^T K_l \alpha}$, 则

$$\frac{\partial F}{\partial \alpha_i} = \frac{1}{\lambda} \psi'(y_i, \lambda \alpha_i) + \frac{1}{\lambda} \sum_{l=1}^m \eta_l(\alpha) (K_l \alpha)_i \quad (12)$$

1.1.2 多核学习正则化路径

建立正则化路径的基本步骤是变化 λ 的值, 使其从 $+\infty$ 到 0, 从而获得全部的解路径。多核学习正则化路径在正则化参数上不再是分段线性而是分段光滑, 采用数值延拓技术跟踪求解, 其路径跟踪算法的核心是预测校正法。

多核学习正则化路径算法包含一系列预测校正步骤, 令 $\sigma = \log_2 \lambda$, $J(\alpha, \sigma) = \frac{\partial F}{\partial \sigma}(\alpha, \sigma)$, 算法主要步骤如下:

(1) 算法初始化。根据对偶问题几何表示确定

$$\begin{aligned} \beta_i &= -\frac{y_i}{2} \quad i = 1, 2, \dots, n \\ \lambda &= \max_l \left(\frac{\beta^T K_l \beta}{d_l^2} \right)^{\frac{1}{2}} \quad \alpha = \frac{\beta}{\lambda} \end{aligned} \quad (13)$$

(2) 由 $J(\alpha, \sigma) = \frac{\partial F}{\partial \alpha}(\alpha, \sigma) = 0$ 得到 (α_0, σ_0) ;

(3) 从 (α_0, σ_0) 开始预测, 对 $\alpha(\sigma_0 + \lambda)$ 使用一阶展开, 如 $\lambda_1 = \lambda_0 + h$, $\alpha_1 = \alpha_0 + h \frac{d\alpha}{d\sigma}(\sigma_0)$, h 为预测步长;

(4) 若 (α_1, σ_1) 不满足 $J(\alpha_1, \sigma_1) = 0$, 从 $\alpha = \alpha_1$ 开始采用牛顿法求解非线性方程组 $J(\alpha, \sigma_1) = 0$ 。 h 足够小, 牛顿法将二次收敛于 $J(\alpha, \sigma_1) = 0$ 的解 α_2 ;

(5) 若 $\lambda = 0$, 算法结束; 否则, $\alpha_0 = \alpha_1, \sigma_0 = \sigma_1$, 返回步骤(3)。

求解多核学习正则化路径的关键是计算 $K_l \alpha$ 。 $K_l \alpha$ 的简化计算能提高算法的效率, 因此降低矩阵的维度, 提高矩阵的计算效率, 能有效改进多核学习正则化路径的学习效率。

1.2 CUR 矩阵分解算法

CUR 矩阵分解算法^[16-23]是处理大规模高维数据的常用方法, 算法的思想是采用抽样技术抽取大规模高维矩阵的部分行或列, 构建并分解为 3 个低维度矩阵, 然后利用 3 个矩阵相乘得到原矩阵的低秩近似矩阵。 CUR 算法得到的低秩近似矩阵类似于 SVD 分解算法降维求解得到的, 而且 CUR 算法基于原矩阵行列抽样降维, 分解效率比 SVD 分解算法高, 同时还保留着原矩阵的重要性质。

定义 1 对任意给定的矩阵 $A \in \mathbf{R}^{m \times n}$, 从 A 中选择 c 列构成矩阵 $C \in \mathbf{R}^{m \times c}$, 从 A 中选择 r 行构成矩阵 $R \in \mathbf{R}^{r \times n}$, C 和 R 学习计算出交叉矩阵 $U \in \mathbf{R}^{c \times r}$, $m \times n$ 维矩阵 $\tilde{A} \approx CUR$ 为 A 基于行列的近似矩阵, 或者称为 CUR 矩阵分解算法。

对于矩阵 A 行、列的抽样策略, 首先计算 A 中每行和每列的概率分布, 利用式(14)计算每行 p_i 的概率分布, 有

$$p_i = \frac{\sum_{j=1}^n A_{i,j}^2}{\sum_{i=1}^m \sum_{j=1}^n A_{i,j}^2} \quad (14)$$

利用式(15)计算每列 $q_j|_{j=1}^n$ 的概率分布,有

$$q_j = \frac{\sum_{i=1}^m A_{i,j}^2}{\sum_{i=1}^m \sum_{j=1}^n A_{i,j}^2} \quad (15)$$

根据计算出的行、列概率分布,按从大到小的顺序从矩阵 A 中抽取 c 列构造矩阵 $C \in \mathbb{R}^{m \times c}$,从矩阵 A 中抽取 r 行构造矩阵 $R \in \mathbb{R}^{r \times n}$,然后通过 C 和 R 的学习计算出交叉矩阵 $U \in \mathbb{R}^{c \times r}$,矩阵 U 的计算公式为

$$U = C^+ AR^+ \quad (16)$$

式中: C^+ 和 R^+ 分别为 C 和 R 的广义逆。近似矩阵 \tilde{A} 比原始矩阵 A 低秩、稀疏,同时保留着原矩阵的重要性质,进而将对原矩阵的操作转化为对低秩近似矩阵的运算,降低了矩阵计算的复杂度,提高了算法的效率。

2 基于CUR矩阵分解的多核学习正则化路径近似算法

本文从提高多核学习正则化路径中涉及核矩阵的相关矩阵计算效率角度提出了MKLRPCUR算法。该算法将CUR矩阵分解方法和MKLRP结合,在MKL正则化路径求解中应用矩阵抽样近似技术。首先采用CUR算法得到核矩阵 K_l 的近似矩阵 \tilde{K}_l 的分解矩阵,然后利用低维分解矩阵的乘积近似代替核矩阵,调整所有矩阵计算顺序,优化算法计算效率。MKLRPCUR算法缩小矩阵规模,降低矩阵计算复杂度,提高模型的学习效率。

2.1 基于CUR分解算法矩阵近似

在多核学习正则化路径中涉及核矩阵计算的 $K_l \alpha$ 为 K_l 和 α 两部分相乘, K_l 采用CUR矩阵分解算法得到近似矩阵 $\tilde{K}_l \approx CUR$,然后将得到的近似矩阵和 α 相乘得到 $K_l \alpha$ 的近似,有

$$\tilde{K}_l \alpha = \tilde{K}_l \alpha \approx CUM \quad (17)$$

式中: $M = R\alpha$; $M \in \mathbb{R}^{r \times 1}$ 为临时矩阵。

矩阵计算 $K_l \alpha$, $K_l \in \mathbb{R}^{n \times n}$,采用CUR矩阵分解算法得到近似矩阵的主要步骤。

输入:核矩阵 K_l 、矩阵 α ,采样列数 c 和采样行数 r 。

输出:近似矩阵 $\tilde{K}_l \alpha$

(1) 根据式(14)计算矩阵 K_l 中每一行被选择的概率 p_i ;根据式(15)计算矩阵 K_l 中每一列被选择的概率 q_j ;

(2) 从矩阵 K_l 所有列中按 q_j 从大到小选择 c 列,构造矩阵 $C \in \mathbb{R}^{n \times c}$;从矩阵 K_l 所有行中按 p_i 从大到小选择 r 行,构造矩阵 $R \in \mathbb{R}^{r \times n}$;

(3) 利用步骤(2)中得到的矩阵 C 和 R 根据式(16)计算交叉矩阵 U ;

(4) 首先利用步骤(2)中得到的矩阵 R 和矩阵 α 相乘得到矩阵 M ,然后利用步骤(2)中得到的矩阵 C 、步骤(3)中得到的矩阵 U 和 M 相乘得到近似计算 $\tilde{K}_l \alpha = \tilde{K}_l \alpha \approx CUM$ 。

MKLRPCUR算法利用矩阵行列概率抽样方法得到低秩近似矩阵,通过矩阵分解缩小了计算矩阵的规模,降低了矩阵计算的复杂度。

2.2 相对误差界分析

对任意给定的矩阵 $A \in \mathbb{R}^{m \times n}$,可设定 $S \in \mathbb{R}^{n \times c}$ ($S \in \mathbb{R}^{r \times m}$)为采样矩阵,缩放矩阵 $D \in \mathbb{R}^{c \times c}$ ($D \in \mathbb{R}^{r \times r}$)为对角矩阵。从 A 中选择 c 列构成矩阵 $C \in \mathbb{R}^{m \times c}$,从 A 中选择 r 行构成矩阵 $R \in \mathbb{R}^{r \times n}$,则

$$C = ASD, R = (SD)^T A = DS^T A \quad (18)$$

将 S 和 D 添上 C 和 R 的下标, 即 $C = AS_c D_c, R = D_R S_R^T A$ 。

定义 2 对任意给定的矩阵 $A \in \mathbf{R}^{m \times n}$, 从 A 中选择 c 列构成矩阵 $C \in \mathbf{R}^{m \times c}$, $m \times n$ 维矩阵 $A' \approx CX$ 为 A 基于列的近似矩阵, 或者称为 CX 矩阵分解算法。

引理 1 给定矩阵 $A \in \mathbf{R}^{m \times n}$, $k \ll \min\{m, n\}$ 为一个秩参数, 令 $\epsilon \in (0, 1]$, 选取 A 的 c 列构造矩阵 C , 设置 $c = O(k^2 \log(1/\delta)/\epsilon^2)$, 近似矩阵 $A' = P_C A = CC^+ A$, 则存在

$$\min_{X \in \mathbf{R}^{c \times n}} \|A - CX\|_F = \|A - CC^+ A\|_F \leq (1 + \epsilon) \|A - A_k\|_F \quad (19)$$

式中: $P_C A$ 为矩阵 A 在子空间 C 的投影, C^+ 为 C 的广义逆。当 $c = O(k \log k/\epsilon^2)$ 时也同样满足上式成立。

证明 如果 $C = AS_c D_c, X_f = C^+ A$ 是使 $\|A - CX\|_F$ 最小的矩阵, 则

$$\|A - CC^+ A\|_F = \|A - (AS_c D_c)(AS_c D_c)^+ A\|_F \leq \|A - (AS_c D_c)(P_{A,k} AS_c D_c)^+ P_{A,k} A\|_F \quad (20)$$

式中: $P_{A,k} = U_{A,k} U_{A,k}^T$ 为在 SVD 分解最大的 k 个左奇异向量的投影。为了约束式(20), 根据文献[16]的方法抽取矩阵 A 和 A_k 的列, 求解 $\min_{X \in \mathbf{R}^{m \times m}} \|XA_k - A\|_F$ 的最优近似解, 由此得到推论

$$\|A - (AS_c D_c)(A_k S_c D_c)^+ A_k\|_F \leq (1 + \epsilon) \|A - AA_k^+ A_k\|_F = (1 + \epsilon) \|A - A_k\|_F \quad (21)$$

引理 2 给定矩阵 $A \in \mathbf{R}^{m \times n}$, C 为选取矩阵 A 部分列构成的矩阵, 令 $\epsilon \in (0, 1]$, 选取 A 的 r 行构造矩阵 R , 设置 $r = O(c^2 \log(1/\delta)/\epsilon^2)$, 则存在

$$\|A - CUR\|_F \leq (1 + \epsilon) \|A - CC^+ A\|_F \quad (22)$$

当 $r = O(c \log c/\epsilon^2)$ 时, 上式也成立。

证明 根据文献[16]的方法抽取矩阵 A 和 C 的行, 求解 $\min_{X \in \mathbf{R}^{c \times n}} \|CX - A\|_F$ 近似解的问题, 由此得到推论

$$\|A - C(D_R S_R^T C)^+ D_R S_R^T A\|_F \leq (1 + \epsilon) \|A - CC^+ A\|_F \quad (23)$$

式中: $U = (D_R S_R^T C)^+, R = D_R S_R^T A$ 。

引理 3 给定矩阵 $A \in \mathbf{R}^{m \times n}$ 和一个整数 $k \ll \min\{m, n\}$, A_k 为秩 k 的近似于 A 的矩阵, 如果按概率选择的列数 c 或行数 r 是关于 k 和 ϵ 的多项式 (ϵ 为误差元素), $c = O(k^2 \log(1/\delta)/\epsilon^2)$, 构造矩阵 $C, r = O(c^2 \log(1/\delta)/\epsilon^2)$, 构造矩阵 R , 则满足概率至少为 $1 - \delta$ 使得

$$\|A - CUR\|_F \leq (1 + \epsilon) \|A - A_k\|_F \quad (24)$$

证明 综合引理 1, 2 得到

$$\|A - CUR\|_F \leq (1 + \epsilon) \|A - CC^+ A\|_F \leq (1 + \epsilon)^2 \|A - A_k\|_F = (1 + \epsilon') \|A - A_k\|_F \quad (25)$$

令 $\epsilon' = \epsilon$, 得证 $\|A - CUR\|_F \leq (1 + \epsilon) \|A - A_k\|_F$ 。

定理 1 给定矩阵 $K \in \mathbf{R}^{n \times n}$ 和一个整数 $k \ll n, \tilde{K}_k$ 为秩 k 的近似于 K 的矩阵, 按概率选择 K 的 c 列构造矩阵 C , 选取 K 的 r 行构造矩阵 R , 行数 r 是关于 c 和 ϵ 的多项式 (ϵ 为误差元素, 令 $\epsilon \in (0, 1]$), $r = O(c \log c/\epsilon^2)$, 则满足

$$\|K - CUR\|_F \leq (1 + \epsilon) \|K - \tilde{K}_k\|_F \quad (26)$$

证明 由于列数 c 或行数 r 关于 k 或 ϵ 的多项式 $c = O(k \log k/\epsilon^2), r = O(c \log c/\epsilon^2)$ 分别适用于引理 1 和引理 2, 则同样适用于引理 3。由此可以得出结论当 $r = O(c \log c/\epsilon^2)$ 时, 由引理 3 的式(24)得到

$\|K - CUR\|_F \leq (1 + \epsilon) \|K - \tilde{K}_k\|_F$, 即 $K \approx \tilde{K} \approx CUR$ 。

综合上述引理, 由引理 1 $\|A - CC^+A\|_F \leq (1 + \epsilon) \|A - A_k\|_F$ 和引理 2 有 $\|A - CUR\|_F \leq (1 + \epsilon) \|A - CC^+A\|_F$, 可以得到引理 3, 即

$$\|A - CUR\|_F \leq (1 + \epsilon) \|A - A_k\|_F \quad (27)$$

误差元素 ϵ 足够小时, 有

$$\|A - CUR\|_F \approx \|A - A_k\|_F, CUR \approx A_k \quad (28)$$

所以当 A 为 $K \in \mathbb{R}^{n \times n}$, ϵ 足够小, CUR 和 \tilde{K} 的相对误差也很小, $K \approx \tilde{K} \approx CUR$ 。

2.3 时间复杂度

本文提出的 MKLRPCUR 算法时间复杂度分为多核学习正则化路径和采用 CUR 矩阵分解算法降低矩阵计算规模两部分。多核学习正则化路径是分段光滑的, 当训练集规模为 n_1 , 核函数个数规模为 m_1 , 求解每一步的复杂度为 $O(n_1^3 + m_1 n_1^2)$, 步骤总数复杂度为 $O(m_1)$ 。由此, 多核学习正则化路径时间复杂度为 $O(m_1 n_1^3 + m_1^2 n_1^2)$ 。 CUR 矩阵分解算法中原始矩阵为 $A \in \mathbb{R}^{m \times n}$, 计算 F 范数的时间复杂度为 $O(mn)$, 计算矩阵行列概率的时间复杂度为 $O(m + n)$, 计算 C 的广义逆的时间复杂度为 $O(2mc^2 + 2c^3)$, R 的广义逆的时间复杂度为 $O(2nr^2 + 2r^3)$, 计算 U 的时间复杂度为 $O(mcn + ncr)$ 。由此, CUR 算法的时间复杂度为

$$O(mn + m + n + 2mc^2 + 2c^3 + 2nr^2 + 2r^3 + mcn + ncr) \quad (29)$$

c, r 相对于数据规模很小, 时间复杂度为 $O(cmn)$, 又因为 MKLRPCUR 算法中 A 为 $K_l^{n \times n}$, 故矩阵计算部分时间复杂度为 $O(cn^2)$ 。

综上所述, 实现矩阵降维计算的 MKLRPCUR 算法时间复杂度为 $O(m_1(n_1^3 + m_1 n_1^2 + cn^2))$ 。

3 实验与分析

为了验证 MKLRPCUR 算法的有效性, 本文选取常用的 UCI 标准数据集和两个图像数据库 ORL 和 COIL 进行实验。实验环境为 4 核 Intel Core i7 个人 PC 机, 主频 3.60 GHz, 内存 32 GB, 算法实现采用 Python 语言, 编程环境使用 Pycharm 4.0.4, 同时在进行图像数据实验时, 使用了开源计算机视觉库 OpenCV。由于核函数的组合不是本文研究重点, 所以选择学习和分类经验较好的高斯核和多项式核构造多核函数。同时, 采用 MKLRPCUR 算法、多核学习正则化路径精确算法 (MKLRP) 以及文献 [7] 中的 MKLRPMC 算法从算法运行时间和分类准确率两方面进行了实验对比。实验中正则化参数初始值设置为

$$\lambda = \max_k \left(\frac{\beta^T K_k \beta}{d_k^2} \right)^{\frac{1}{2}}, \alpha = \frac{\beta}{\lambda}, \beta_i = -\frac{y_i}{2} \quad (30)$$

路径计算终止值 $\lambda_{\min} = 0$, 采样列数 $c = \frac{1}{2}n$, 采样行数 $r = O(c \lg(c)/\epsilon^2)$, $\epsilon = 1$ 。

3.1 UCI 数据集

选取 UCI 数据库中 3 种数据集进行实验, 实验数据集统计信息如表 1 所示。实验中对所有属性应用一个次数为 2 的多项式核; 对每个属性分别应用次数为 1, 2 的多项式核; 对所有属性应用一个 γ 为 2 的高斯核; 对每个属性分别应用 10 个

表 1 UCI 数据集

Table 1 UCI dataset

Dataset	Sample size	Attribute	Class
Liver	345	6	2
Sonar	208	60	2
Diabetes	768	8	2

系数 γ 不同的高斯核, $\gamma \in \{0.1, 0.3, 0.5, 0.8, 1, 1.1, 1.3, 1.5, 1.8, 2\}$ 。实验采用 5-折交叉验证法, 从数据集中随机选择 80% 样本作训练数据, 20% 样本作测试数据。同时, 为了避免实验中的偶然误差, 在每种数据集上重复实验 5 次, 计算运行时间和分类准确率的平均值作为算法评价指标的结果。

表 2 给出了 3 种算法在 UCI 数据集上的平均分类准确率, 可以看出近似算法 MKLRPCUR、MKLRPMC 的分类准确率接近于精确算法 MKLRP 的分类准确率, 三者的预测性能基本一致。表 3 给出了 3 种算法在 UCI 数据集上的平均运行时间, 可以看出近似算法 MKLRPCUR 和 MKLRPMC 的平均运行时间明显都比精确算法 MKLRP 的时间少, 而且, MKLRPCUR 算法的平均运行时间比 MKLRPMC 算法的时间更少。

表 2 3 种算法在 UCI 数据集的分类准确率

Dataset	MKLRP	MKLRPCUR	MKLRPMC
Liver	87.80±1.61	87.80±1.54	87.83±1.63
Sonar	87.96±1.70	87.81±1.35	87.14±1.41
Diabetes	90.36±1.42	90.20±1.49	90.08±1.57

表 3 3 种算法在 UCI 数据集的平均运行时间

Dataset	MKLRP	MKLRPCUR	MKLRPMC
Liver	273±34	156±37	182±48
Sonar	1 332±117	1 235±110	1 267±106
Diabetes	917±83	638±94	797±94

3.2 图像数据集

ORL 是人脸图像数据库, 由 40 名工作人员的信息组成, 每个人有 10 幅脸部图像, 每张图像像素大小为 92×112 。COIL 数据库是哥伦比亚大学采集的图像, 包含杯子、小鸭玩具和小车模型等 20 种事物的图像, 每种事物还包括 72 幅每隔 5° 旋转的图像, 共计 1 440 幅图像, 每张图像像素大小为 64×64 。

实验首先对图像预处理, 然后提取图像特征作为算法的输入。在每张图像中, 对提取的所有特征应用一个次数为 2 的多项式核; 对每个特征分别应用一个次数为 1, 2 的多项式核; 对所有特征应用 1 个 γ 为 2 高斯核; 对每个特征分别应用 10 个系数 γ 不同的高斯核, $\gamma \in \{0.1, 0.3, 0.5, 0.8, 1, 1.1, 1.3, 1.5, 1.8, 2\}$ 。实验选取不同的训练样本个数, ORL 图像数据库中每个工作人员分别随机选择 3 张、5 张和 8 张图像, 即分别选择 120 张、200 张和 320 张图像做训练样本; COIL 图像数据库中每个事物分别随机选择 25 张、40 张和 50 张图像, 即分别选择 500 张、800 张和 1 000 张图像做训练样本。同时为了避免实验中的偶然误差, 在每种数据集上重复实验 5 次, 计算运行时间和分类准确率的平均值作为算法评价指标的结果。

图 1 给出了 3 种算法在 ORL 图像数据库上不同训练样本平均运行时间变化趋势, 从图中可以看出, 在小规模数据集中随着训练数据样本个数添加, 3 种算法运行时间也在不同程度的增加, 同时近似算法的运行时间比精确算法较少, MKLRPCUR 算法的运行时间比 MKLRPMC 算法较少。图 2 显示了 3 种算法在 COIL 图像数据库上不同训练样本平均运行时间变化趋势, 从图中可以看出, 在较大规模数据集中随着训练数据样本个数添加, 3 种算法运行时间也在不同程度的增加, 并且, 随着训练数据增加, 运行时间差别越来越明显, 近似算法的运行时间比精确算法较少, MKLRPCUR 算法的运行时间比 MKLRPMC 算法少。

表 4 给出了 3 种算法在两种图像数据库上不同训练样本数量情况下实验 5 次的平均运行时间, 可以看出近似算法 MKLRPCUR 的运行时间比另外两种算法少, 明显提高了算法的计算效率。表 5 给出了 3 种算法在两种图像数据库上不同训练样本数量情况下实验 5 次的平均分类准确率, 可以看

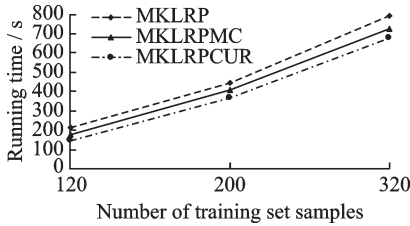


图1 3种算法在ORL图像数据库中运行时间
Fig.1 Running time of three algorithms in ORL image database

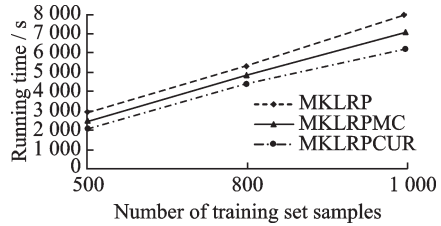


图2 3种算法在COIL图像数据库中运行时间
Fig.2 Running time of three algorithms in COIL image database

表4 3种算法在ORL和COIL图像库的运行时间

Table 4 Running time of the three algorithms in ORL and COIL image libraries

Dataset	MKLRP			MKLRPCUR			MKLRPMC		
	120	200	320	120	200	320	120	200	320
ORL	212 ± 33	446 ± 15	791 ± 42	145 ± 27	369 ± 21	677 ± 35	176 ± 27	408 ± 23	723 ± 46
COIL	500	800	1 000	500	800	1 000	500	800	1 000
	2 855 ± 591	5 277 ± 626	7 966 ± 744	2 035 ± 553	4 355 ± 411	6 144 ± 378	2 386 ± 580	4 803 ± 492	7 021 ± 633

表5 3种算法在ORL和COIL图像库的分类准确率

Table 5 Classification accuracy of the three algorithms in ORL and COIL image libraries

Dataset	MKLRP			MKLRPCUR			MKLRPMC		
	120	200	320	120	200	320	120	200	320
ORL	88.3 ± 0.5	90.7 ± 0.5	93.1 ± 0.6	88.30 ± 0.5	90.6 ± 0.3	93.1 ± 0.4	88.30 ± 0.4	90.1 ± 0.7	92.5 ± 0.8
COIL	500	800	1 000	500	800	1 000	500	800	1 000
	96.4 ± 0.2	97.1 ± 0.3	97.2 ± 0.2	96.4 ± 0.3	97.1 ± 0.3	97.1 ± 0.2	96.1 ± 0.3	97.3 ± 0.3	97.1 ± 0.3

出两种近似算法的分类准确率接近于精确算法MKLRP的分类准确率,3种算法具有相似的预测性能。

综合MKLRPCUR、MKLRPMC和MKLRP算法在UCI标准数据集、ORL和COIL图像数据库上的实验分析结果,3种算法的分类准确率相似,但算法的运行时间不同,MKLRPCUR算法的运行时间比MKLRPMC和MKLRP算法的运行时间较少,训练样本数据量越多运行时间区别越明显。实验结果说明了近似算法MKLRPCUR在保证实验准确率的前提下,有效缩短了算法的运行时间,提高了算法学习的效率。

4 结束语

本文提出了基于CUR矩阵分解的多核学习正则化路径近似算法,该方法将CUR算法和多核学习正则化路径结合在一起,利用抽样技术获得核矩阵的近似矩阵的分解矩阵,实现了矩阵降维,缩小了矩阵存储的内存空间,降低了矩阵计算的时间复杂度。实验结果表明,MKLRPCUR算法不仅可以将复杂问题简单化,有效降低算法的计算复杂度,还可以保证学习的准确性、整体改善算法的效率。由此多核学习正则化路径近似算法的研究可以得到广泛推广。本文没有对多核学习的模型选择进行研究,学习准确率未得到提升,如何进行模型选择提高多核学习的学习性能是未来研究的重点。

参考文献:

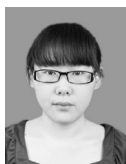
- [1] 汪洪桥, 孙富春, 蔡艳宁, 等. 多核学习方法[J]. 自动化学报, 2010, 36(8): 1037-1050.
WANG Hongqiao, SUN Fuchun, CAI Yanning, et al. On multiple kernel learning methods[J]. ACTA Automatica Sinica, 2010, 36(8): 1037-1050.
- [2] BACH F R, LANCKRIET G R G. Multiple kernel learning, conic DUALITY, and the SMO algorithm[C]//Proceedings of International Conference on Machine Learning. Banff, Canada: [s.n.], 2004: 1-8.
- [3] RAKOTOMAMONJY A, BACH F R, CANU S, et al. [J]. Journal of Machine Learning Research, 2008, 9(3): 2491-2521.
- [4] BACH F R, THIBAU R, JORDAN M I. Computing regularization paths for learning multiple kernels[C]//Proceedings of Advances in Neural Information Processing Systems 17. Cambridge, MA: MIT Press, 2004: 41-48.
- [5] 丁立中, 廖士中. 基于正则化路径的支持向量机近似模型选择[J]. 计算机研究与发展, 2012, 49(6): 1248-1255.
DING Lizhong, LIAO Shizhong. Approximate model selection on regularization path for support vector machines[J]. Journal of Computer Research and Development, 2012, 49(6): 1248-1255.
- [6] 王梅, 王莎莎, 孙莺萁, 等. SVRRP_(MCC): 一种支持向量回归机的正则化路径近似算法[J]. 计算机科学, 2017, 44(12): 42-47.
WANG Mei, WANG Shasha, SUN Yingqi, et al. SVRRP_{MCC}: A regularization path approximation algorithm of support vector regression[J]. Computer Science, 2017, 44(12): 42-47.
- [7] 王梅, 李董, 孙莺萁, 等. 求解大规模问题的多核学习正则化路径算法[J]. 模式识别与人工智能, 2018, 31(2): 190-196.
WANG Mei, LI Dong, SUN Yingqi, et al. Regularization path algorithm of multiple kernel learning for solving large scale problems[J]. Pattern Recognition and Artificial Intelligence, 2018, 31(2): 190-196.
- [8] HASTIE T, ROSSET S, TIBSHIRANI R, et al. The entire regularization path for the support vector machine[J]. Journal of Machine Learning Research, 2004, 5(4): 1391-1415.
- [9] 刘松华. 核矩阵低秩分解与核空间信息度量研究及应用[D]. 西安: 西安电子科技大学, 2011.
LIU Songhua. Low-rank decomposition of kernel matrix and informative energy metric in kernel space[D]. Xian: Xidian University of Electronic Technology, 2011.
- [10] 王臻. 核范数随机矩阵求解及其图像处理应用[D]. 南京: 南京邮电大学, 2017.
WANG Zhen. Random matrix solution of nuclear norm and its application in image processing[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2017.
- [11] 雷恒鑫, 刘惊雷. 基于行列联合选择矩阵分解的偏好特征提取[J]. 模式识别与人工智能, 2017, 30(3): 279-288.
LEI Hengxin, LIU Jinglei. Preference feature extraction based on column union row matrix decomposition[J]. Pattern Recognition and Artificial Intelligence, 2017, 30(3): 279-288.
- [12] 雷恒鑫. 基于CUR分解的偏好处理方法研究[D]. 烟台: 烟台大学, 2018.
LEI Hengxin. Research on preference processing method based on CUR decomposition[D]. Yantai: Yantai University, 2018.
- [13] 袁淦钊. 矩阵的低秩近似算法及其应用[D]. 广州: 华南理工大学, 2013.
YUAN Ganzhao. Low-rank matrix approximation algorithm and its applications[D]. Guangzhou: South China University of Technology, 2013.
- [14] 徐梦珂. 矩阵的低秩近似及应用[D]. 贵阳: 贵州大学, 2017.
XUN Mengke. The low rank approximation of matrix and its application[D]. Guiyang: Guizhou University, 2017.
- [15] 周水生, 郑颖, 穆新亮. 基于Cholesky分解的K2DPCA人脸识别研究[J]. 系统工程理论与实践, 2016, 36(2): 528-535.
ZHOU Shuisheng, ZHENG Ying, MU Xinlaing. K2DPCA methods for face recognition based on cholesky decomposition[J]. System Engineering-Theory & Practice, 2016, 36(2): 528-535.
- [16] DRINEAS P, MAHONEY M W, MUTHUKRISHNAN S. Relative-error CUR matrix decompositions[J]. SIAM Journal on Matrix Analysis & Applications, 2008, 30(2): 844-881.
- [17] 管涛, 李玉玲. 大规模矩阵降维的随机逼近方法[J]. 数学的实践与认识, 2016, 46(24): 184-193.
GUAN Tao, LI Yuling. Stochastic approximation approaches of large-scale matrix dimension reduction[J]. Practice and Understanding of Mathematics, 2016, 46(24): 184-193.

- [18] MAHONEY M W. CUR matrix decompositions for improved data analysis[J]. Proceedings of the National Academy of Sciences, 2009, 106(3): 697-702.
- [19] DRINEAS P, KANNAN R, MAHONEY M W. Fast Monte Carlo algorithms for matrices II: Computing a low-rank approximation to a matrix[J]. SIAM Journal on Computing, 2004, 36(1): 158-183.
- [20] DRINEAS P, KANNAN R, MAHONEY M W. Fast Monte Carlo algorithms for matrices III: Computing a compressed approximate matrix decomposition[J]. SIAM Journal on Computing, 2006, 36(1): 184-206.
- [21] MITROVIC N, ASIF M T, RASHEED U, et al. CUR decomposition for compression and compressed sensing of large-scale traffic data[C]//Proceedings of International IEEE Conference on Intelligent Transportation Systems.[S.l.]: IEEE, 2013: 1475-1480.
- [22] 曾琦, 李国盛, 郭云鹏, 等. 高维数据降维中SVD与CUR分解对比分析[J]. 中原工学院学报, 2014, 25(6): 80-84.
ZENG Qi, LI Guosheng, GUO Yunpeng, et al. Comparative analysis of SVD decomposition and CUR decomposition in dimensionality reduction[J]. Journal of Zhongyuan University of Technology, 2014, 25(6): 80-84.
- [23] 郑清雅. 云计算环境下基于学习风格的教学资源推荐系统设计及实现[D]. 沈阳: 沈阳师范大学, 2016.
ZHEN Qingya. The design and implementation of recommendation system of teaching resources based on the learning style under the cloud computing environment[D]. Shenyang: Shenyang Normal University, 2016.

作者简介:



王梅(1976-),女,教授,研究方向:机器学习、模型选择和核方法, E-mail: wangmei@nepu.edu.cn。



李董(1992-),女,硕士研究生,研究方向:机器学习。



薛成龙(1995-),男,硕士研究生,研究方向:机器学习。

(编辑:刘彦东)