

# 一种非负稀疏近邻表示的多标签学习算法

陈思宝, 徐丹洋, 罗斌

(安徽大学计算机科学与技术学院 合肥 230601)

**【摘要】**针对训练数据中的非线性流形结构以及基于稀疏表示的多标签分类中判别信息丢失严重的问题, 该文提出一种非负稀疏近邻表示的多标签学习算法。首先找到待测试样本每个标签类上的 $k$ -近邻, 然后基于LASSO稀疏最小化方法, 对待测试样本进行非负稀疏线性重构, 得到稀疏的非负重构系数。再根据重构误差计算待测试样本对每个类别的隶属度, 最后实现多标签数据分类。实验结果表明所提出的方法比经典的多标签 $k$ -近邻分类(ML-KNN)和稀疏表示的多标记学习算法(ML-SRC)方法性能更优。

**关键词** 多标签学习; 稀疏近邻表示; LASSO稀疏最小化; 非负重构

中图分类号 TP391.4 文献标志码 A doi:10.3969/j.issn.1001-0548.2015.06.018

## A Non-Negative Sparse Neighbor Representation for Multi-Label Learning Algorithm

CHEN Si-bao, XU Dan-yang, and LUO Bin

(School of Computer Science and Technology, Anhui University Hefei 230601)

**Abstract** In order to avoid the influence of the nonlinear manifold structure in training data and preserve more discriminant information in the sparse representation based multi-label learning, a new multi-label learning algorithm based on non-negative sparse neighbor representation is proposed. First of all, the  $k$ -nearest neighbors among each class are found for the test sample. Secondly, based on non-negative the least absolute shrinkage and selectionator operator (LASSO)-type sparse minimization, the test sample is non-negative linearly reconstructed by the  $k$ -nearest neighbors. Then, the membership of each class for the test sample is calculated by using the reconstruction errors. Finally, the classification is performed by ranking these memberships. A fast iterative algorithm and its corresponding analysis of converging to global minimum are provided. Experimental results of multi-label classification on several public multi-label databases show that the proposed method achieves better performances than classical ML-SRC and ML-KNN.

**Key words** LASSO sparse minimization; multi-label learning; non-negative reconstruction; sparse neighbor representation

多标签数据在现实世界中广泛存在。与传统的单标签数据不同, 多标签数据可以同时包含多个标签。例如一张图片中存在房屋、树木、湖泊等多个景物, 每类景物作为一个标签可以为这张图片贴上多个标签。每个标签可以作为一个单独类别, 因此, 多标签数据可以同时被划分到多个类别中, 且比单标签数据分类更为复杂。

早期有关多标签数据分类的研究主要集中在文本数据方面<sup>[1-2]</sup>。近十年来, 随着图像检索、数据挖掘、标签推荐等方面的发展, 多标签学习逐渐成为一个热门的研究方向。目前, 研究者们已提出了一些有效的多标签学习算法。如文献[3]提出的ML-KNN算法将传统的 $k$ -近邻算法和贝叶斯理论相

结合, 来实现多标签数据的分类; 文献[4]联合多个分类器, 提出一套针对多标签数据分类的整体方案, 成功地对音乐流派进行多标签分类; 文献[5]提出基于核的多标签分类算法Rank-SVM, 实现了对基因数据集的有效分类。

另一方面, 稀疏表示与压缩感知<sup>[6]</sup>也受到广泛重视。其中, 文献[7]利用训练样本稀疏线性表示待测试样本, 提出稀疏表示分类方法(SRC), 实现了高效并且鲁棒的人脸识别; 文献[8]对稀疏系数加入非负性约束, 将非负稀疏表示分类(non-negative SRC, NSRC)应用于人脸识别; 文献[9]采用近邻数据进行稀疏线性重构, 提出基于稀疏近邻表示的分类(sparse neighbor representation classification, SNRC)

收稿日期: 2014-04-22; 修回日期: 2015-06-12

基金项目: 国家863项目(2014AA015104); 国家自然科学基金(61202228, 61472002); 安徽省高校自然科学研究重点项目(KJ2012A004)

作者简介: 陈思宝(1979-), 男, 副教授, 主要从事模式识别和图像处理方面的研究。

方法。文献[10]将SRC算法应用于多标签数据分类,提出基于稀疏表示的多标记学习算法(ML-SRC)。

ML-SRC算法首先利用全部训练样本在LASSO<sup>[11]</sup>稀疏最小化约束下线性表示待测试样本,然后对稀疏系数的元素进行筛选,大于0的稀疏系数所对应的训练样本被认为与待测试样本有正相关而被保留下来,小于0的系数被直接置0;最后,通过计算待测试样本对于每个标签类别的隶属度来实现多标签分类。然而,该算法用全部训练数据组成字典,给后面稀疏系数的计算带来巨大的计算量,并且当训练数据中存在非线性流形结构时,所得到的稀疏重构及分类效果会受到严重影响。此外,将小于0的稀疏系数强制改为0,丢弃了稀疏系数中的所有负数,仅保留非负稀疏系数进行多标签分类,导致稀疏重构系数中判别信息的严重丢失,从而使得分类效果下降。

本文针对上述两个现象,提出一种非负稀疏近邻表示的多标签学习算法(ML-NSNRC)。

## 1 相关方法

### 1.1 基于稀疏表示的多标签数据分类

假设训练集包含 $n$ 个多标签训练样本 $\mathbf{y}_i \in R^d$ ,  $i=1,2,\dots,n$ ,  $T=\{1,2,\dots,L\}$ 为其所对应的标签集合。所有训练样本拼成矩阵形式 $\mathbf{D}=[\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n] \in R^{d \times n}$ ,  $d < n$ 。根据训练样本包含的标签信息,构造训练样本 $\mathbf{y}_i$ 的标签信息 $\mathbf{b}_i$ ,若 $\mathbf{y}_i$ 包含标签 $l$ ,则令 $b_{li}=1$ ,否则令 $b_{li}=0$ 。构造训练样本矩阵 $\mathbf{D}$ 对应的标签矩阵 $\mathbf{B}=[\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n] \in R^{L \times n}$ 。多标签学习的目标是学习一个多标签分类器 $h: \mathbf{y} \rightarrow 2^L$ ,来对待测试样本 $\mathbf{y}$ 所包含的标签进行估计。ML-SRC利用训练集 $\mathbf{D}$ 稀疏线性重构待测试样本 $\mathbf{y}$ :

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s}} \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{s}\|_2^2 + \lambda \|\mathbf{s}\|_1 \quad (1)$$

式中, $\mathbf{s}$ 表示稀疏线性重构系数; $\lambda > 0$ 为调节参数。

ML-SRC<sup>[10]</sup>算法采用函数 $f(\mathbf{y}, l)$ 来计算待测试样本属于每个标签的隶属度:

$$f(\mathbf{y}, l) = \frac{\sum_{i=1}^n (b_{li} \hat{s}_i)}{\sum_{i=1}^n \hat{s}_i} \quad l=1,2,\dots,L \quad (2)$$

式中, $\hat{s}_i = \begin{cases} s_i, & s_i > 0 \\ 0, & s_i \leq 0 \end{cases}, i=1,2,\dots,n$ 。通过对 $f(\mathbf{y}, l)$ 与阈值函数的比较,得到 $\mathbf{y}$ 的预测标签集 $h(\mathbf{y})$ 。

### 1.2 稀疏近邻表示分类算法

稀疏近邻表示分类算法(SNRC)<sup>[9]</sup>在训练集中先

找到待测试样本在每个类别上的 $k$ -近邻,然后用每个类别的 $k$ -近邻样本数据对待测试样本进行稀疏线性重构,并根据其在每个类别上 $k$ -近邻的重构误差最小对待测试样本分类。

SNRC通过优化目标函数 $\mathbf{s}_i$ 得到待测试样本 $\mathbf{y}$ 的稀疏线性重构系数:

$$\mathbf{s}_i = \arg \min_{\mathbf{s}_i} \|\mathbf{y} - \boldsymbol{\psi}_i \mathbf{s}_i\|_2 + \lambda \|\mathbf{s}_i\|_1 \quad (3)$$

式中, $\boldsymbol{\psi}_i$ 表示待测试样本 $\mathbf{y}$ 在训练样本集 $\mathbf{D}$ 的第 $i$ 类训练样本上的 $k$ -近邻样本集合; $i=1,2,\dots,c$ ;  $c$ 表示训练样本集中的类别总数;调节参数 $\lambda > 0$ 。根据每个类别的重构误差最小对待测试样本进行分类。SNRC采用近邻数据进行稀疏线性重构,可有效地避免训练数据中非线性流形结构带来的影响。

## 2 非负稀疏近邻分类

借鉴SNRC采用近邻数据进行稀疏线性重构,可有效地避免训练数据中非线性流形结构给稀疏线性重构带来的影响。同时借鉴非负稀疏表示分类<sup>[8]</sup>在优化过程中直接限制其稀疏系数为非负,可使待测试样本用局部近邻数据进行凸的稀疏线性重构。由此,本文提出非负稀疏近邻分类(NSNRC)方法。

### 2.1 NSNRC目标函数

NSNRC算法通过KNN方法得到待测试样本 $\mathbf{y}$ 在训练样本集 $\mathbf{D}$ 的第 $i$ 类上的 $k$ -近邻样本集 $\boldsymbol{\psi}_i$ 。将 $\boldsymbol{\psi}_i$ 作为基,对稀疏系数加入非负约束,得到待测试样本 $\mathbf{y}$ 在 $\boldsymbol{\psi}_i$ 上的非负稀疏解:

$$\hat{\mathbf{s}}_i = \arg \min_{\mathbf{s}_i} \frac{1}{2} \|\mathbf{y} - \boldsymbol{\psi}_i \mathbf{s}_i\|_2^2 + \lambda \|\mathbf{s}_i\|_1 \quad (4)$$

s.t.  $i=1,2,\dots,c$

式中, $\|\mathbf{s}_i\|_1 = \sum_{j=1}^k |s_{ij}|$ ,  $s_{ij} \geq 0$ ,  $j=1,2,\dots,k$ ;  $\lambda > 0$ 。

当训练样本集 $\mathbf{D}$ 存在非线性流形时, $\boldsymbol{\psi}_i$ 可看成符合局部线性流形的训练样本子集。加入非负性限制后,对局部近邻数据点进行凸性重构,进一步挖掘局部结构信息,从而提高重构效果。

### 2.2 NSNRC的优化算法

由于NSNRC的目标函数式(4)是一个凸函数,因此可以证明存在全局最优解。一般,全局最优解可通过二次规划或梯度下降得到。然而,梯度下降收敛较慢,二次规划实施又比较复杂,因此采用类似文献[12]提出的乘法迭代更新公式来求式(4)中的最优非负稀疏重构系数:

$$\mathbf{s}_i^{(t+1)} = \mathbf{s}_i^{(t)} \boldsymbol{\psi}_i^T \mathbf{y} / (\boldsymbol{\psi}_i^T \boldsymbol{\psi}_i \mathbf{s}_i^{(t)} + \lambda \mathbf{1}) \quad (5)$$

式中,  $\mathbf{1}$  表示一个全为1的向量。当  $\mathbf{s}_i$  中每个初始元素全部为正时,  $\mathbf{s}_i$  中的元素在式(5)更新下的结果均是非负的。

### 2.3 算法收敛性分析

本文利用辅助函数对NSNRC的优化算法进行收敛性分析。

**定义 1** 若  $G(\mathbf{s}_i, \mathbf{s}_i^{(t)})$  是  $F(\mathbf{s}_i)$  的辅助函数, 则需要满足以下条件:

$$G(\mathbf{s}_i, \mathbf{s}_i^{(t)}) \geq F(\mathbf{s}_i), \quad G(\mathbf{s}_i, \mathbf{s}_i) = F(\mathbf{s}_i) \quad (6)$$

**引理 1** 如果  $\mathbf{K}(\mathbf{s}_i^{(t)})$  是对角矩阵:

$$\mathbf{K}_{ab}(\mathbf{s}_i^{(t)}) = \delta_{ab} \frac{(\boldsymbol{\psi}^T \boldsymbol{\psi} \mathbf{s}_i^{(t)})_a + \lambda}{\mathbf{s}_{ia}^{(t)}} \quad (7)$$

那么  $F(\mathbf{s}_i)$  的辅助函数可定义为:

$$G(\mathbf{s}_i, \mathbf{s}_i^{(t)}) = F(\mathbf{s}_i^{(t)}) + (\mathbf{s}_i - \mathbf{s}_i^{(t)})^T \nabla F(\mathbf{s}_i^{(t)}) + \frac{1}{2} (\mathbf{s}_i - \mathbf{s}_i^{(t)})^T \mathbf{K}(\mathbf{s}_i^{(t)}) (\mathbf{s}_i - \mathbf{s}_i^{(t)}) \quad (8)$$

引理1证明: 显然  $G(\mathbf{s}_i, \mathbf{s}_i) = F(\mathbf{s}_i)$  成立, 只需要证明  $G(\mathbf{s}_i, \mathbf{s}_i^{(t)}) \geq F(\mathbf{s}_i)$ 。其中:

$$F(\mathbf{s}_i) = F(\mathbf{s}_i^{(t)}) + (\mathbf{s}_i - \mathbf{s}_i^{(t)})^T \nabla F(\mathbf{s}_i^{(t)}) + \frac{1}{2} (\mathbf{s}_i - \mathbf{s}_i^{(t)})^T (\boldsymbol{\psi}^T \boldsymbol{\psi}) (\mathbf{s}_i - \mathbf{s}_i^{(t)}) \quad (9)$$

与式(8)比较, 结果为:

$$0 \leq (\mathbf{s}_i - \mathbf{s}_i^{(t)})^T [\mathbf{K}(\mathbf{s}_i^{(t)}) - \boldsymbol{\psi}^T \boldsymbol{\psi}] (\mathbf{s}_i - \mathbf{s}_i^{(t)}) \quad (10)$$

为了证明式(10)的半正定性, 只需考虑矩阵:

$$\mathbf{M}_{ab}(\mathbf{s}_i^{(t)}) = \mathbf{s}_{ia}^{(t)} (\mathbf{K}(\mathbf{s}_i^{(t)}) - \boldsymbol{\psi}^T \boldsymbol{\psi})_{ab} \mathbf{s}_{ib}^{(t)} \quad (11)$$

的半正定性。当且仅当  $\mathbf{M}$  半正定时,  $\mathbf{K} - \boldsymbol{\psi}^T \boldsymbol{\psi}$  也为半正定:

$$\begin{aligned} \mathbf{x}^T \mathbf{M} \mathbf{x} &= \sum_{ab} x_a \mathbf{M}_{ab} x_b = \\ & \sum_{ab} \mathbf{s}_{ia}^{(t)} (\boldsymbol{\psi}^T \boldsymbol{\psi})_{ab} \mathbf{s}_{ib}^{(t)} x_a^2 - x_a \mathbf{s}_{ia}^{(t)} (\boldsymbol{\psi}^T \boldsymbol{\psi})_{ab} \mathbf{s}_{ib}^{(t)} x_b + \lambda = \\ & \sum_{ab} (\boldsymbol{\psi}^T \boldsymbol{\psi})_{ab} \mathbf{s}_{ia}^{(t)} \mathbf{s}_{ib}^{(t)} \left[ \frac{1}{2} x_a^2 + \frac{1}{2} x_b^2 - x_a x_b \right] + \lambda = \\ & \frac{1}{2} \sum_{ab} (\boldsymbol{\psi}^T \boldsymbol{\psi})_{ab} \mathbf{s}_{ia}^{(t)} \mathbf{s}_{ib}^{(t)} (x_a - x_b)^2 + \lambda \geq 0 \quad (12) \end{aligned}$$

式(12)中证明了  $\mathbf{M}$  的半正定性, 由此可知式(10)也是半正定的, 进而可知  $G(\mathbf{s}_i, \mathbf{s}_i^{(t)}) \geq F(\mathbf{s}_i)$ , 因此引理1成立。

**引理 2** 如果  $G$  是一个辅助函数, 那么目标函数  $F$  在以下更新中是非增的

$$\mathbf{s}_i^{(t+1)} = \arg \min_{\mathbf{s}_i} G(\mathbf{s}_i, \mathbf{s}_i^{(t)}) \quad (13)$$

引理 2 证明: 由式 (13) 易知  $G(\mathbf{s}_i^{(t+1)}, \mathbf{s}_i^{(t)}) \leq G(\mathbf{s}_i^{(t)}, \mathbf{s}_i^{(t)})$ , 又由引理1可知当  $G(\mathbf{s}_i, \mathbf{s}_i^{(t)})$  是  $F(\mathbf{s})$  的辅助函数时, 满足定义1中的条件。因此有如下结果:

$$F(\mathbf{s}_i^{(t+1)}) \leq G(\mathbf{s}_i^{(t+1)}, \mathbf{s}_i^{(t)}) \leq G(\mathbf{s}_i^{(t)}, \mathbf{s}_i^{(t)}) = F(\mathbf{s}_i^{(t)}) \quad (14)$$

则引理2成立。

由引理2可知, 当且仅当  $\mathbf{s}_i^{(t)}$  的值使  $G(\mathbf{s}_i, \mathbf{s}_i^{(t)})$  达到局部最小时,  $F(\mathbf{s}_i^{(t+1)}) = F(\mathbf{s}_i^{(t)})$ 。当  $F(\mathbf{s}_i)$  在  $\mathbf{s}_i^{(t)}$  的邻域内导数存在且连续时, 通过迭代式(13)可得到目标函数的一串有顺序的估计值, 当目标函数到达局部最小值时, 估计值收敛:

$$F(\mathbf{s}_{i\min}) \leq F(\mathbf{s}_i^{(t+1)}) \leq F(\mathbf{s}_i^{(t)}) \leq \dots \leq F(\mathbf{s}_i^0) \quad (15)$$

根据引理2, 利用梯度下降使辅助函数的导数为0, 推导稀疏系数的更新规则:

$$\begin{aligned} \nabla_{\mathbf{s}_i} G(\mathbf{s}_i, \mathbf{s}_i^{(t)}) = \\ \boldsymbol{\psi}^T (\boldsymbol{\psi} \mathbf{s}_i^{(t)} - \mathbf{y}) + \lambda \mathbf{I} + \mathbf{K}(\mathbf{s}_i^{(t)}) (\mathbf{s}_i - \mathbf{s}_i^{(t)}) = 0 \quad (16) \end{aligned}$$

解得:

$$\begin{aligned} \mathbf{s}_i &= \mathbf{s}_i^{(t)} - k^{-1} (\mathbf{s}_i^{(t)}) (\boldsymbol{\psi}^T \boldsymbol{\psi} \mathbf{s}_i^{(t)} - \boldsymbol{\psi}^T \mathbf{y} + \lambda \mathbf{I}) = \\ \mathbf{s}_i^{(t)} &- (\mathbf{s}_i^{(t)}) / (\boldsymbol{\psi}^T \boldsymbol{\psi} \mathbf{s}_i^{(t)} + \lambda \mathbf{I}) (\boldsymbol{\psi}^T \boldsymbol{\psi} \mathbf{s}_i^{(t)} - \boldsymbol{\psi}^T \mathbf{y} + \lambda \mathbf{I}) = \\ &\mathbf{s}_i^{(t)} (\boldsymbol{\psi}^T \mathbf{y}) / (\boldsymbol{\psi}^T \boldsymbol{\psi} \mathbf{s}_i^{(t)} + \lambda \mathbf{I}) \quad (17) \end{aligned}$$

因此, 非负稀疏近邻分类方法的目标函数可通过迭代优化算法式(5)来求解。

### 2.4 算法复杂度分析

本文提出的NSNRC算法计算复杂度可通过分析式得到。式(5)中  $(\boldsymbol{\psi}_i)^T \mathbf{y}$  和  $(\boldsymbol{\psi}_i)^T \boldsymbol{\psi}_i$  的计算复杂度分别为  $O(kd)$  和  $O(k^2d)$ , 其中,  $d$  为数据维数,  $k$  为待测试样本  $\mathbf{y}$  的近邻数。优化算法每一步的计算复杂度为  $O(2k + k^2 + k)$ 。因此, NSNRC的计算复杂度为  $O(kd + k^2d + t(3k + k^2))$ , 其中,  $t$  为优化迭代次数。采用快速迭代收缩门限算法 (fast iterative shrinkage-thresholding algorithm, FISTA) 方法对SRC算法求其稀疏解, 根据文献[13]可得其计算复杂度为  $O(tn^2d)$ , 其中,  $n$  为训练样本集  $\mathbf{D}$  的样本总数。

## 3 基于非负稀疏近邻表示的多标签分类

虽然基于稀疏表示的多标签分类方法可以达到不错的分类效果, 但仍存在以下不足: 1) 在求解标签隶属度时, ML-SRC对稀疏解中的负值进行的置0操作势必导致稀疏解包含的判别信息严重丢失, 从而对多标签分类的结果造成一定影响; 2) ML-SRC在所有训练数据集上对待测试样本进行稀疏表示。当训练数据集中存在非线性流形时, 如果仍在所有训练数据上进行稀疏表示则会产生严重影响。本文提出的非负稀疏近邻分类方法可有效解决以上两个不足。

令待测试样本  $\mathbf{y}$  在训练集  $\mathbf{D}$  上各标签的  $k$ -近邻

集  $\Psi = [\varphi_1^0, \varphi_1^1, \varphi_2^0, \varphi_2^1, \dots, \varphi_L^0, \varphi_L^1] \in R^{d \times 2kL}$ , 其中,  $d$  为样本的维度,  $k$  为近邻数,  $L$  为数据集标签总数。  $\varphi_l^j$  由 KNN 方法求得, 表示待测试样本  $\mathbf{y}$  在  $\mathbf{D}$  中第  $l$  标签的第  $j$  类上的  $k$ -近邻, 其中, 上标  $j=0$  表示不存在标签  $l$  的类别,  $j=1$  表示存在标签  $l$  的类别。

将  $\varphi_l^j$  作为基, 对稀疏系数加入非负约束, 得到待测试样本  $\mathbf{y}$  在  $\varphi_l^j$  上的非负稀疏解:

$$\hat{s}_l^j = \arg \min_{s_l^j} \frac{1}{2} \|\mathbf{y} - \varphi_l^j s_l^j\|_2^2 + \lambda \|s_l^j\|_1$$

$$\text{s.t. } s_l^j \geq 0, i=1, 2, \dots, k, j=0, 1, l=1, 2, \dots, L \quad (18)$$

式中,  $\|s_l^j\|_1 = \sum_{i=1}^k |s_{li}^j|$ ; 调节参数  $\lambda > 0$ 。

根据 NSNRC 的优化算法, 得到最优非负稀疏重构系数的优化迭代公式:

$$s_l^{j(t+1)} = s_l^{j(t)} ((\varphi_l^j)^T \mathbf{y}) / ((\varphi_l^j)^T \varphi_l^j s_l^{j(t)} + \lambda \mathbf{1}) \quad (19)$$

式中,  $t \geq 0$ 。

根据所求得的非负稀疏解来计算待测试样本  $\mathbf{y}$  对每个标签类别的重构误差:

$$\varepsilon_l^j(x) = \|\varphi_l^j s_l^j - \mathbf{y}\|_2 \quad (20)$$

### 3.1 标签隶属度计算

根据  $\mathbf{y}$  对应各标签上每类的重构误差, 提出一个实值函数  $g(\mathbf{y}, l)$  来表示待测试样本  $\mathbf{y}$  属于标签  $l$  的隶属度:

$$g(\mathbf{y}, l) = \frac{e^{-\varepsilon_l^1(\mathbf{y})}}{e^{-\varepsilon_l^0(\mathbf{y})} + e^{-\varepsilon_l^1(\mathbf{y})}} \quad (21)$$

根据求得的隶属度预测待测试样本  $\mathbf{y}$  的标签集  $h(\mathbf{y}) = \{l | g(\mathbf{y}, l) > p(\mathbf{y}), l \in \mathbf{T}\}$ , 其中,  $p(\mathbf{y})$  为阈值函数。

### 3.2 ML-NSNRC 的算法流程

在 ML-NSNRC 算法中, 待测试样本  $\mathbf{y}$  被其在训练集中各标签的每一类上的  $k$ -近邻集稀疏重构。再通过稀疏重构误差来计算待测试样本  $\mathbf{y}$  属于每个标签类的隶属度进而确定  $\mathbf{y}$  的标签集。算法流程为:

1) 输入, 训练样本矩阵  $\mathbf{D} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n] \in R^{d \times n}$  及其对应的标签集信息  $\mathbf{B} \in R^{L \times n}$ , 待测试样本  $\mathbf{y} \in R^d$ ;

2) 输出, 待测试样本  $\mathbf{y}$  的隶属度  $g(\mathbf{y}, l)$  和标签集。

分为以下3个步骤:

1) 利用  $\ell_1$  范数对  $\mathbf{D}$  中每一列归一化, 通过 KNN 方法得到待测试样本  $\mathbf{y}$  在训练集中各标签的每一类上的  $k$ -近邻集, 从而得到  $\Psi$ ;

2) 利用式(20)求待测试样本  $\mathbf{y}$  在  $\varphi_l^j$  上的非负稀

疏解  $s_l^j$ ,  $s_l^j$  初始值设为  $1/n$ ;

3) 根据求得的非负稀疏系数, 利用式(20)和式(21)计算待测试样本属于每个标签的隶属度, 进而确定  $\mathbf{y}$  的标签集。

## 4 实验分析

### 4.1 实验数据简介

为了验证所提出的 ML-NSNRC 多标签学习算法的性能, 本文选取3个常用的多标签数据集: 自然场景数据集<sup>[3]</sup>、Yeast 基因数据集和 scene 数据集<sup>[14]</sup>进行多标签分类实验。

#### 1) 自然场景数据集

在自然场景数据集下, 多标签学习算法通过已标记的训练图像对测试图像的标签集进行预测。该数据集包含了2 000张自然场景的图像, 其中每张图像的标签都已经过人工交互判断标定, 标签包括沙漠、山、海洋、日落、树木。该数据集中超过22%的风景图像同时属于多个类别, 其中每张图像的平均标签个数为  $1.24 \pm 0.44$  个, 每张图像可表示为一个294维的特征向量。

#### 2) Yeast 基因数据集

多标签学习算法通过分析 Yeast 基因集中已知功能的 Yeast 基因, 对待测试样本中的 Yeast 基因的功能进行标注。Yeast 基因的功能类别被分为4层, 本文实验采用的文献[5]中的方案, 即仅仅使用顶层的功能类别。最终该数据集包含2 417个基因, 共有14个类别, 每个基因由一个103维的特征向量表示, 每个基因的平均标签个数为  $4.24 \pm 1.57$  个。

#### 3) scene 数据集

scene 数据集<sup>[14]</sup>的任务是通过已知标签集的训练图像来预测待分类样本图像的标签集。该数据集包含2 407张图像, 其中每张图像包含落日、海滩、山川、原野、落叶、都市这6个标签中的一个或多个, 每张图像可表示为一个294维的特征向量。

### 4.2 ML-NSNRC 中参数初始化的影响

本文所提出的 ML-NSNRC 迭代更新算法需要对稀疏系数进行初始化, 不同初始稀疏系数会对迭代更新算法的收敛性产生影响。为了测试参数初始化带来的影响, 分别采用全部平均  $1/n$ 、绝对标准正态随机分布随机数  $|N(0, 1)|$ 、 $(0, 1)$  均匀分布随机数、绝对最小二乘解以及最小二乘解正部对稀疏系数初始化。图1显示了在自然场景数据集下, 对所提出的 ML-NSNRC 算法采用5种不同的初始化方法, 其目标函数值的迭代更新趋势可看出, 不同的参数初始化

对ML-NSNRC迭代更新算法并未造成较大的影响,而且所提出的优化方法收敛速度非常快。为了简化算法并节省时间,将稀疏系数 $s$ 的每个元素初始值设为 $1/n$ 。

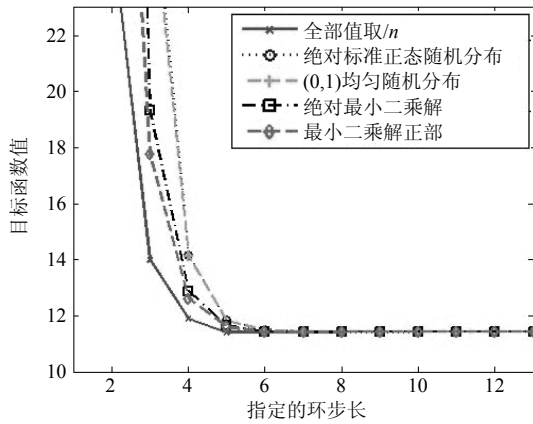


图1 不同稀疏系数初始化对ML-NSNRC收敛性的影响

### 4.3 ML-NSNRC中近邻数 $k$ 的影响

用 $k$ -近邻来近似表示数据中非线性流形的局部结构时,近邻数 $k$ 的选择至关重要。不同数据分布形状及不同的数据采集密度都需要设置不同的近邻数 $k$ 。在ML-NSNRC中,近邻数 $k$ 值受到库中每类标签正负样本个数的影响,如训练样本集中标签 $l$ 的正样本数为 $n_1$ ,负样本数为 $n_2$ ,其中 $n_1 \ll n_2$ ,那么 $k$ 必须小于等于 $n_1$ 。为了测试不同近邻数 $k$ 对所提出的ML-NSNRC的影响,分别取近邻数 $k$ 为5的正整数倍,在scene数据集上进行多标签学习。图2显示了ML-NSNRC和ML-KNN的平均分类精度随着不同近邻数 $k$ 的变化情况。为了便于比较,将相同实验配置下的ML-SRC和ML-NSRC的平均分类精度用水平线表示。从图中可知当 $k$ 取20时ML-NSNRC的平均精度达到较好水平,当 $k$ 取5时ML-KNN的平均精度达到较好水平。

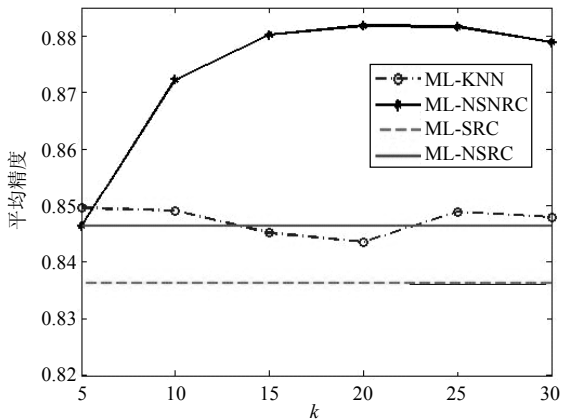


图2 近邻数 $k$ 对ML-NSNRC和ML-KNN平均精度的影响

### 4.4 多标签数据分类结果与分析

利用文献[3]中的多标签算法评价标准验证本文所提出的算法的有效性。评价标准包括汉明损失、1-错误率、排序损失、覆盖率、平均精度这5个方面,其中,平均精度的值越大表示多标签算法越好,而其他4个标准则是值越小表示算法越好。在实施ML-SRC时对稀疏系数限制为非负,提出非负稀疏表示的多标签学习算法(ML-NSRC)。为了进一步验证本文算法的性能,将所提出的ML-NSNRC与3种多标签学习算法进行比较。这3种算法分别是基于 $k$ -近邻和传统贝叶斯的多标签学习算法ML-KNN、基于稀疏表示的多标记学习算法ML-SRC和基于非负稀疏表示的多标签学习算法ML-NSRC。其中ML-KNN和ML-SRC均是当前多标签学习领域效果最好的方法之一。ML-NSNRC实验中非负稀疏系数求解的迭代次数为100次,自然场景数据集下 $k$ 取400, Yeast数据集下 $k$ 取5, scene数据集下 $k$ 取20。

表1是4种多标签学习算法在自然场景数据集上的性能比较。可看出,在自然场景数据集上,本文所提出的算法在1-错误率、覆盖率、排序损失和平均精度上比其他3种算法表现更佳。其中,汉明损失比ML-KNN低19.4%,覆盖率比ML-KNN低14.7%,平均精度比ML-KNN高4.3%。

表1 4种多标签学习算法在自然场景数据集上的性能比较

算法	指标				
	汉明损失	1-错误率	覆盖率	排序损失	平均精度
ML-KNN	0.196	0.321	0.986	0.180	0.790
ML-SRC	0.160	0.286	0.889	0.156	0.815
ML-NSRC	0.155	0.275	0.879	0.151	0.819
ML-NSNRC	0.158	0.271	0.841	0.143	0.824

表2是4种多标签学习算法在Yeast基因集上的性能比较,其中ML-SRC和ML-NSRC采用了十折交叉验证的方法。可知在Yeast基因集上,本文所提出的算法在汉明损失、覆盖率、排序损失和平均精度这4个性能指标上比其他3种算法更佳。其中,汉明损失比ML-KNN降低18.5%,排序损失比ML-KNN降低9%,平均精度比ML-KNN和ML-SRC提高1.6%。

表2 4种多标签学习算法在Yeast数据集上的性能比较

算法	指标				
	汉明损失	1-错误率	覆盖率	排序损失	平均精度
ML-KNN	0.232	0.229	6.479	0.177	0.757
ML-SRC	0.194	0.235	6.244	0.170	0.757
ML-NSRC	0.190	0.246	6.078	0.165	0.758
ML-NSNRC	0.189	0.232	6.077	0.161	0.769

表3是4种多标签学习算法在scene数据集上的性能比较。可知本文算法在各项指标上都优于其他3种算法。其中,本文算法的汉明损失比ML-KNN和ML-SRC分别降低18.2%和21.4%,覆盖率比ML-KNN降低32.6%,平均精度比ML-KNN和ML-SRC分别提高了4.3%和5.6%。

表3 4种多标签学习算法在scene数据集上的性能比较

算法	指标				
	汉明损失	1-错误率	覆盖率	排序损失	平均精度
ML-KNN	0.099	0.248	0.583	0.099	0.847
ML-SRC	0.103	0.275	0.566	0.096	0.836
ML-NSRC	0.097	0.256	0.543	0.091	0.846
ML-NSNRC	0.081	0.204	0.393	0.063	0.883

通过比较自然场景图像数据集、Yeast基因数据集以及scene数据集上的实验结果可知,本文所提出的ML-NSNRC算法在分类精确度上比ML-KNN、ML-SRC、ML-NSRC算法均有明显提高,在其他4项指标上也表现出较好的水平。本文所提出的算法结合了非负稀疏表示的有效鉴别能力,和 $k$ -近邻方法对非线性流形数据集的局部线性整理,可以有效解决多标签数据分类问题。

## 5 结束语

本文提出了一种非负稀疏近邻表示分类算法(NSNRC),将其成功地应用到多标签数据分类中,并提出一种非负稀疏近邻表示分类的多标签学习算法ML-NSNRC。将待测试样本在训练样本每一类中的 $k$ -近邻作为待测试样本的训练样本基,并在稀疏编码过程中对稀疏系数加入非负性限制。该算法有效克服了非线性流形数据对算法的影响,并且得到了全为非负值的稀疏系数,从而为数据的标签分类保留了更多的判别信息。提出一种快速优化迭代求解算法,并给出其收敛到全局最优的证明。在自然场景图像、Yeast基因、scene这3个数据集上的实验结果表明,所提出的算法在性能上优于其他经典算法,符合预期目标。下一阶段需要寻求更优的稀疏模型,进一步提高多标签分类效果。

## 参 考 文 献

[1] SCHAPIRE R E, SINGER Y. Boostexter: a boosting-based system for text categorization[J]. Machine Learning, 2000, 39(2-3): 135-168.

- [2] UEDA N, SAITO K. Parametric mixture models for multi-label text[J]. Advances in Neural Information Processing, 2003(15): 721-728.
- [3] ZHANG M L, ZHOU Z H. ML-KNN: a lazy learning approach to multi-label learning[J]. Pattern Recognition, 2007, 40(7): 2038-2048.
- [4] SANDEN C, ZHANG J Z. Enhancing multi-label music genre classification through ensemble techniques[C] // Proceedings of the 34th international ACM SIGIR Conference on Research and development in Information Retrieval. New York: ACM, 2011: 705-714.
- [5] ELISSEEFF A, WESTON J. A kernel method for multi-labelled classification[J]. Advances in Neural Information Processing, 2002(14): 681-687.
- [6] CANDÈS E J, ROMBERG J, TAO T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information[J]. IEEE Transactions on Information Theory, 2006, 52(2): 489-509.
- [7] WRIGHT J, YANG A Y, GANESH A, et al. Robust face recognition via sparse representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210-227.
- [8] JI Y, LIN T, ZHA H. Mahalanobis distance based non-negative sparse representation for face recognition[C] // International Conference on Machine Learning and Applications. Miami, FL: IEEE, 2009: 41-46.
- [9] HUI K, LI C, ZHANG L. Sparse neighbor representation for classification[J]. Pattern Recognition Letters, 2012, 33(5): 661-669.
- [10] 宋相法, 焦李成. 基于稀疏表示的多标记学习算法[J]. 模式识别与人工智能, 2012, 25(1): 124-129.
- SONG Xiang-fa, JIAO Li-cheng. A multi-label learning algorithm based on sparse representation[J]. Pattern Recognition and Artificial Intelligence, 2012, 25(1): 124-129.
- [11] TIBSHIRANI R. Regression shrinkage and selection via the lasso[J]. Journal of the Royal Statistical Society (Series B, Methodological), 1996, 58(1): 67-88.
- [12] LEE D D, SEUNG H S. Algorithms for non-negative matrix factorization[J]. Advances in Neural Information Processing, 2001(2): 556-562.
- [13] BECK A, TEOULLE M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems[J]. SIAM Journal on Imaging Sciences, 2009, 2(1): 183-202.
- [14] BOUTELL M R, LUO J, SHEN X, et al. Learning multi-label scene classification[J]. Pattern Recognition, 2004, 37(9): 1757-1771.