

doi:10.16088/j.issn.1001-6600.2016.03.006

基于稀疏学习的 kNN 分类

宗 鸣^{1,2}, 龚永红³, 文国秋¹, 程德波^{1,2}, 朱永华⁴

(1.广西师范大学 计算机科学与信息工程学院, 广西 桂林 541004; 2.广西区域多源信息集成与智能处理协同创新中心, 广西 贵港 537000; 3.桂林航天工业学院 信息工程系, 广西 桂林 541004; 4.广西大学 计算机与电子信息学院, 广西 南宁 530004)

摘 要:在 kNN 算法分类问题中, k 的取值一般是固定的, 另外, 训练样本中可能存在的噪声能影响分类结果。针对以上存在的两个问题, 本文提出一种新的基于稀疏学习的 kNN 分类方法。本文用训练样本重构测试样本, 其中, L_1 -范数导致的稀疏性用来对每个测试样本用不同数目的训练样本进行分类, 这解决了 kNN 算法固定 k 值问题; L_{21} -范数产生的整行稀疏用来去除噪声样本。在 UCI 数据集上进行实验, 本文使用的新算法比原来的 kNN 分类算法能取得更好的分类效果。

关键词:稀疏学习; 重构; L_1 -范数; L_{21} -范数; 噪声样本

中图分类号: TP181

文献标志码: A

文章编号: 1001-6600(2016)03-0039-07

0 引言

模式分类是人工智能和模式识别中的一个基本问题。其主要考虑的问题: 给定 N 个已知类别的 C 类样本, 对于任意给定的测试样本, 如何确定其类别, 从而完成模式分类。迄今为止, 已有很多分类器被提出, 其中最具代表性的是最近邻(NN)分类器, 其将测试样本归为与其距离最近的样本所在的类^[1]。kNN (k-Nearest Neighbor)算法是一种应用广泛的分类方法, 它是最近邻算法的推广形式。kNN 算法的核心思想: 如果一个样本在特征空间中的 k 个最相似的样本中的大多数属于同一个类别, 则该样本也属于这个类别。

本文发现传统的 kNN 算法每次都用同样 k 个数目的训练样本对测试样本进行分类, 这在应用中通常不合实际。图 1 以二类问题为例, 在 kNN 算法中, k 固定取 5, 对于左边的测试样本分类, 显然被分为第一类; 对于右边的测试样本分类, 其也被分为第一类; 然而与其最邻近的两个训练样本都是属于第二类, 此时由于 k 固定取 5, 将其分为第一类可能不合理, 应将其分为第二类更准确, 即此时 k 取 2 比较合理。因此, 本文认为固定 k 对每一个测试样本不合理, k 的取值应该由数据的分布决定, 而且决定的方法不是人为设定, 应是

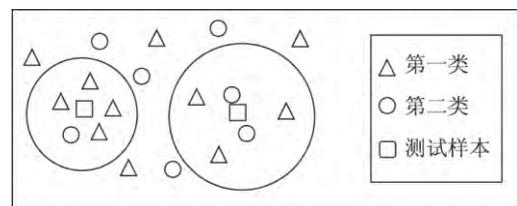


图 1 二类分类问题, $k=5$ 时情形
Fig.1 Binary classification with a fixed k value, i.e. $k=5$

学习数据而得到的。此外, 认为样本之间是存在相关性的, 考虑样本间的相关性更能揭示数据特点, 对于每个测试样本, 应该用与之相关的训练样本对其进行分类。为此, 本文用训练样本重构测试样本得到训练

收稿日期: 2015-09-09

基金项目: 国家自然科学基金资助项目(61450001, 61263035, 61573270); 国家 973 计划项目(2013CB329404); 中国博士后科学基金资助项目(2015M570837); 广西自然科学基金资助项目(2012GXNSFGA060004, 2015GXNSFCB139011, 2015GXNSFAA139306)

通信联系人: 文国秋(1987—), 女, 广西桂林人, 广西师范大学讲师。E-mail: wenguoqiu2008@163.com; 龚永红(1970—), 女, 广西永福人, 桂林航天工业学院副教授。E-mail: zysjd2015@163.com

样本与测试样本之间的相关性关系,并用 lasso(the least absolute shrinkage and selection operator)正则化因子产生稀疏性^[2-4],实现对每个测试样本用经过稀疏选择的训练样本来对其进行分类,从而解决经典 kNN 算法中 k 值固定问题。另外,基于数据集中存在噪声数据^[5]的可能性,本文通过 l_{21} -范数产生整行稀疏性,实现用经去除噪声的训练数据集对测试样本进行分类。因此,本文提出基于稀疏学习的 kNN 分类算法——kNN classification based on sparse learning,简记为 SL-kNN。该算法不仅考虑了样本之间的相关性,并利用稀疏学习实现训练样本子集选择,从而可以较好地解决 kNN 算法中 k 值固定问题和样本中含噪声问题。

1 基于稀疏学习的 kNN 分类

1.1 重构

本文假设有训练样本空间 $X \in \mathbf{R}^{n \times d}$, n 为训练样本数目, d 为特征数目;测试样本空间 $Y \in \mathbf{R}^{m \times d}$, m 为测试样本数目, d 为特征数目;投影矩阵 $W \in \mathbf{R}^{n \times m}$ 。一般我们用线性模型拟合训练数据集,常用最小二乘法,即:

$$\min_W \sum_{i=1}^c \|\mathbf{y}_i^T - \mathbf{w}_i^T X\|_2^2 = \min_W \|\mathbf{Y} - \mathbf{W}^T X\|_F^2. \quad (1)$$

这里 $\|\cdot\|_F$ 是 Frobenius 矩阵范数, $\mathbf{y}_i \in \mathbf{R}^{d \times 1}$, \mathbf{w}_i 是 W 的第 i 列向量。

所谓重构指的是用一组线性无关的向量的线性组合来近似地表达一个给定的向量^[6-7],这里就是用训练样本重构测试样本,分类问题中 \mathbf{y}_i 一般为类标签, W 表示类标签 Y 与训练特征 X 之间的函数关系。而在本文中 \mathbf{y}_i 表示第 i 个测试样本,经过重构过程得到的矩阵 W 表示训练样本 X 和测试样本 Y 之间的相关性,即相关系数。 W 中元素值 w_{ij} 表示第 i 个训练样本与第 j 个测试样本之间的相关性大小, $w_{ij} > 0$ 时代表正相关, $w_{ij} < 0$ 时代表负相关, $w_{ij} = 0$ 时代表不相关;如果 W 中某列向量有 r 个元素值不为 0,则分类时 k 应取 r ,并用对应的这 r 个训练本来对相应的测试样本进行分类,对于其他测试样本,同理可得到各自的 k 值,这就解决了 kNN 算法中 k 值固定问题,但目标函数(1)得出的 W 不是稀疏的。

1.2 稀疏学习

我们通常用最小二乘损失函数来确定线性模型中的向量参数 W ,即:

$$\min_W \|\mathbf{Y} - \mathbf{W}^T X\|_F^2. \quad (2)$$

该损失函数是凸的,其解为 $W^* = (XX^T)^{-1}XY^T$,但是其中的 XX^T 不一定可逆,需要加上一正则化因子对其正则化以使 XX^T 可逆。本文添加 lasso 解决此问题,同时 lasso 被证明能使结果 W 产生稀疏性,这称为元组稀疏(element sparsity)^[2,8-9],对每个测试样本稀疏选择相应的训练本来进行分类,从而解决了上面提到的 k 值固定问题。另外,添加 l_{21} -范数用来使 W 产生整行的稀疏,即行稀疏(row sparsity)^[10],用来去除噪声样本。因此可以得到以下经过正则化的目标函数:

$$\min_W \frac{1}{2} \|\mathbf{Y} - \mathbf{W}^T X\|_F^2 + \lambda_1 \|W\|_1 + \lambda_2 \|W\|_{2,1}, \quad (3)$$

其中 $\|W\|_1 = \sum_{i=1}^n \sum_{j=1}^m |w_{ij}|$, $\|W\|_{2,1} = \sum_{i=1}^n \sqrt{\sum_{j=1}^m |w_{ij}|^2}$ 。参数 λ_1 调控 W 整体的稀疏性, λ_1 值越大越稀疏; λ_2 调控 W 整行稀疏性^[11], λ_2 值越大整行稀疏越明显。

目标函数(3)可以求解,求出的 W 一般具有如下形式,简单举例:

$$W = \begin{pmatrix} 0.1 & 0 & 0.5 \\ 0 & 0.62 & 0 \\ 0.27 & 0 & 0 \\ 0 & 0 & 0 \\ 0.7 & 0.4 & 0 \end{pmatrix}.$$

W 表示训练样本与测试样本之间的相关性,行代表训练样本,列代表测试样本。 W 第一列有 3 个值不为

零,说明第一个测试样本和第一、三、五个训练样本相关,应该用这3个训练样本对第一个测试样本进行分类,即此时 $k=3$ 。同理分别对第二、三个测试样本进行分类,得到相应的 $k=2, k=1$,这里 k 的取值是通过lasso对训练样本稀疏选择得到的。另外 W 中第四行全为0,即出现整行稀疏性,这是 l_{21} -范数导致的结果,这意味着第四个训练样本与所有的测试样本都不相关,显然为噪声样本,在分类时予以去除。

目标函数(3)在重构过程中利用lasso稀疏学习出与每个测试样本相关的训练样本,即对每个测试样本用不同数目的训练样本进行分类,并且用 l_{21} -范数导致的整行稀疏性去除数据集中存在的噪声,有效解决了kNN算法中存在的 k 值固定问题和数据集中的噪声问题。

1.3 SL-kNN 算法

重构、lasso和 l_{21} -范数这三者被整合到一个函数里,成为了一个优化问题,即目标函数(3)。求解目标函数(3)可以从整体上得出一个最优解 W ,学习 W 的过程是由数据驱动的,因此能确保得到的关系最优,并且通过lasso使 W 具有稀疏性,从而使得每个测试样本需要的 k 值不同。 l_{21} -范数使 W 具有整行稀疏性,用来去除噪声样本,然后根据 W 的每一列取出与该测试样本相关的训练样本构造出分类器,最后对每个测试样本进行分类。上面的过程我们称为SL-kNN算法,下面我们给出具体的算法步骤,见算法1。

算法1 SL-kNN 算法。

输入:训练样本和测试样本;

输出:进行分类后的测试样本的类别;

- 1.将输入样本数据进行正规化;
- 2.根据算法2(下文给出)求解目标函数(3),得到最优解 W ;
- 3.根据 W 每列非零元素的个数得到每个测试样本的 k 值;
- 4.根据上面得到的 k 值,选出相应的这 k 个训练样本,并用这 k 个训练样本来对该测试样本进行分类;
- 5.根据这 k 个训练样本中的大多数属于同一个类别,则判定该测试样本也属于这个类别;
- 6.对所有的测试样本执行步骤4~5进行分类;
- 7.算法结束。

1.4 算法优化分析

目标函数(3)是凸的但非光滑的,本文通过设计一个新的加速近似梯度法来求解该函数,首先在目标函数(3)上进行近似梯度法,通过如下设置^[12]:

$$f(W) = \frac{1}{2} \|Y - W^T X\|_F^2, \quad (4)$$

$$L(W) = f(W) + \lambda_1 \|W\|_1 + \lambda_2 \|W\|_{2,1}, \quad (5)$$

注意到 $f(W)$ 是凸且可微的, $\lambda_1 \|W\|_1$ 和 $\lambda_2 \|W\|_{2,1}$ 是凸但非平滑的。使用近似梯度法优化 W ,通过下面的优化规则迭代地更新 W :

$$W(t+1) = \arg \min_w G_{\eta(t)}(W, W(t)), \quad (6)$$

其中 $G_{\eta(t)}(W, W(t)) = f(W(t)) + \langle \nabla f(W(t)), W - W(t) \rangle + \frac{\eta(t)}{2} \|W - W(t)\|_F^2 + \lambda_1 \|W\|_1 + \lambda_2 \|W\|_{2,1}$, $\nabla f(W(t)) = XX^T W(t) - XY^T$, $\eta(t)$ 和 $W(t)$ 分别是调优参数和在第 t 次迭代获得的 W 值。

通过忽略公式(6)中独立的 W ,我们可以将其改写为:

$$W(t+1) = \pi_{\eta(t)}(W(t)) = \arg \min_w \frac{1}{2} \|W - U(t)\|_2^2 + \frac{\lambda_1}{\eta(t)} \|W\|_1 + \frac{\lambda_2}{\eta(t)} \|W\|_{2,1}. \quad (7)$$

其中 $U(t) = W(t) - \frac{1}{\eta(t)} \nabla f(W(t))$ 和 $\pi_{\eta(t)}(W(t))$ 是 $W(t)$ 在凸集 $\eta(t)$ 上的欧几里德投影,由于 $W(t+1)$ 在每一行的可分离性,也就是 $W^i(t+1)$,则能单独地对每一行更新权值:

$$w^i(t+1) = \arg \min_{w^i} \frac{1}{2} \|w^i - u^i(t)\|_2^2 + \frac{\lambda_1}{\eta(t)} \|w^i\|_1 + \frac{\lambda_2}{\eta(t)} \|w^i\|_{2,1}, \quad (8)$$

其中 $u^i(t) = w^i(t) - \frac{1}{\eta(t)} \nabla f(w^i(t))$ 和 $w^i(t)$ 分别是 $U(t)$ 和 $W(t)$ 的第 i 行。根据公式(8) $w^i(t+1)$ 可以

得到闭式解如下:

$$\textcircled{1} \text{ 若 } |w^i| - \lambda_1 \geq 0 \text{ 且 } \|u^i(t)\|_2^2 \geq \frac{\lambda_2}{\eta(t)}, \text{ 则 } w^{i*} = (|w^i| - \lambda_1) \cdot \text{sgn}(w^i) + \left(1 - \frac{\lambda_2}{\eta(t) \|u^i(t)\|_2^2}\right) u^i(t);$$

$$\textcircled{2} \text{ 若 } |w^i| - \lambda_1 \geq 0 \text{ 且 } \|u^i(t)\|_2^2 < \frac{\lambda_2}{\eta(t)}, \text{ 则 } w^{i*} = (|w^i| - \lambda_1) \cdot \text{sgn}(w^i);$$

$$\textcircled{3} \text{ 若 } |w^i| - \lambda_1 < 0 \text{ 且 } \|u^i(t)\|_2^2 \geq \frac{\lambda_2}{\eta(t)}, \text{ 则 } w^{i*} = \left(1 - \frac{\lambda_2}{\eta(t) \|u^i(t)\|_2^2}\right) u^i(t);$$

$$\textcircled{4} \text{ 若 } |w^i| - \lambda_1 < 0 \text{ 且 } \|u^i(t)\|_2^2 < \frac{\lambda_2}{\eta(t)}, \text{ 则 } w^{i*} = 0.$$

同时,为了加速公式(6)中的近似梯度法,本文进一步引进辅助变量 $V(t+1)$,即:

$$V(t+1) = W(t) + \frac{\alpha(t) - 1}{\alpha(t+1)} (W(t+1) - W(t)), \quad (9)$$

其中系数 $\alpha(t+1)$ 通常被设置为 $\alpha(t+1) = \frac{1 + \sqrt{1 + 4\alpha(t)^2}}{2}$.

算法 2 中列出了优化算法的伪代码,并在定理 1 中说明了其收敛性。

算法 2 解决公式(3)的伪代码。

输入: $\eta(0), \alpha(1) = 1, \gamma$;

输出: W

1 初始化 $t = 1$;

2 初始化 $W(1)$ 为一个随机的对角矩阵;

3 repeat

4 while $L(W(t)) > G_{\eta(t-1)}(\pi_{\eta(t-1)}(W(t)), W(t))$ do

5 Set $\eta(t-1) = \gamma\eta(t-1)$;

6 end

7 Set $\eta(t) = \eta(t-1)$;

8 计算 $W(t+1) = \arg \min_w G_{\eta(t)}(W, V(t))$;

9 计算 $\alpha(t+1) = \frac{1 + \sqrt{1 + 4\alpha(t)^2}}{2}$;

10 计算公式(9);

11 until 公式(3)收敛。

定理 1 设 $\{W(t)\}$ 是由算法 2 产生的序列,那么对于 $\forall t \geq 1$,式(10)成立:

$$L(W(t)) - L(W^*) \leq \frac{2\gamma L \|W(1) - W^*\|_F^2}{(t+1)^2}, \quad (10)$$

其中 γ 是事先定义的常量, L 是公式(5)中 $f(W)$ 梯度的 Lipschitz 常数以及 $W^* = \arg \min_w L(W)$ 。

定理 1 的证明详见文献[13]。定理 1 表明本文提出的加速近似梯度算法的收敛率为 $O\left(\frac{1}{t^2}\right)$,其中 t 是算法 2 中的迭代次数。

2 实验结果与分析

本文实验数据来自 UCI 机器学习库(<http://archive.ics.uci.edu/~ml/>),其中 Fertility 数据集、Blood Transfusion Service Center(简记 BloodTrans)数据集、German Credit(简记 German)数据集、Climate

Model Simulation Crashes(简记 Climate)数据集、Australian 数据集和 SPECTF Heart(简记 SPECTF)数据集都是二类数据集,Seeds 和 Balance Scale(简记 Balance)数据集为三类数据集,具体情况见表 1。实验环境为 Windows XP 系统环境,实现算法的软件为 Matlab7.11.0。

表 1 UCI 数据集基本情况

Tab.1 UCI dataset summarization

Dataset	Instance	Attribute	Class
Fertility	100	10	2
BloodTrans	748	5	2
Seeds	210	8	3
German	1 000	25	2
Australian	690	14	2
Climate	540	17	2
SPECTF	267	44	2
Balance	625	4	3

表 2 两种算法在 UCI 数据集上的分类准确性(均值±方差)

Tab.2 Comparison on classification accuracy (mean±STD) of two algorithms

数据集	kNN	L-kNN	SL-kNN
Fertility	0.896 3±0.037 2	0.927 8±0.038 5	0.940 7±0.033 6
BloodTrans	0.673 0±0.062 9	0.711 0±0.062 8	0.724 0±0.058 7
Seeds	0.751 4±0.054 5	0.782 4±0.064 2	0.801 4±0.061 4
German	0.782 6±0.538 0	0.813 0±0.051 8	0.842 0±0.053 1
Australian	0.842 9±0.071 2	0.885 7±0.040 2	0.890 5±0.045 2
Climate	0.744 4±0.056 4	0.781 5±0.079 0	0.807 4±0.085 2
SPECTF	0.820 6±0.046 2	0.850 8±0.040 4	0.850 8±0.048 6
Balance	0.850 0±0.085 0	0.880 0±0.103 3	0.880 0±0.103 3

在每个数据集上用十折交叉验证法^[14]做 10 次实验看算法效果,主要的对比算法是 kNN 算法和 L-kNN 算法,其中 kNN 算法是一种经典的数据挖掘分类算法,而 L-kNN 算法是只考虑稀疏性(即只使用 l_1 -范数)而没有考虑去除噪声样本的算法,即相比目标函数(3)没有考虑 l_{21} -范数正则化项。为了保证公平性,kNN 算法、L-kNN 算法和 SL-kNN 算法在每一次实验中选用相同的训练集和测试集,记录这两种算法 10 次实验取得的分类准确率。一般准确率越高,分类的准确性越好,否则,分类的准确性越低。kNN 算法、L-kNN 算法和 SL-kNN 算法在 8 个数据集上进行实验的分类准确性(Accuracy)结果的均值和方差如表 2 所示。

两种算法在数据集上 10 次实验的分类准确率结果如图 2~9 所示。

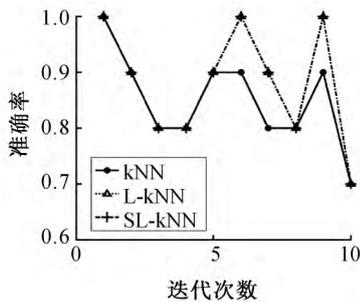


图 2 Fertility
Fig.2 Fertility

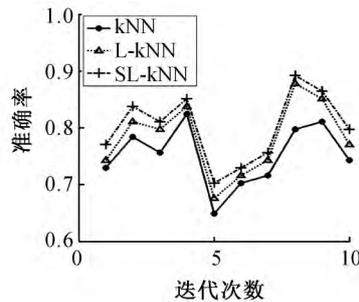


图 3 BloodTrans
Fig.3 BloodTrans

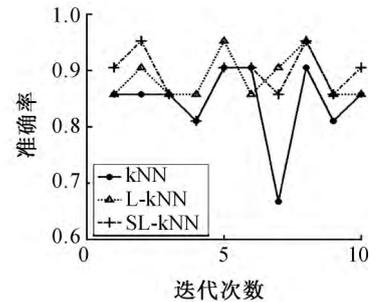


图 4 Seeds
Fig.4 Seeds

分析表 2 可知,通过和 kNN 方法、L-kNN 方法比较,本文提出的改进方法 SL-kNN 获得了最好的分类准确率。例如,SL-kNN 算法的分类准确率在 8 个数据集上平均比 kNN 算法和 L-kNN 算法分别提高了 4.69%和 1.13%。另外,根据图 2~9,我们也发现 SL-kNN 算法几乎在每一次实验中都有较高的分类准确率。

SL-kNN 算法比 L-kNN 算法准确率高是因为使用了 l_{21} -范数消除噪声样本,例如,在 Australian 和 Climate 数据集上,该算法的分类准确率比 L-kNN 方法分别提高了 2.9%和 1.29%。这意味着 Australian 和 Climate 数据集可能含有噪声样本,进一步来说,Australian 数据集可能有更多的噪声样本。综合来看,相比于 kNN 算法和 L-kNN 算法,SL-kNN 算法具有更好的分类效果,对噪声的鲁棒性也更强。

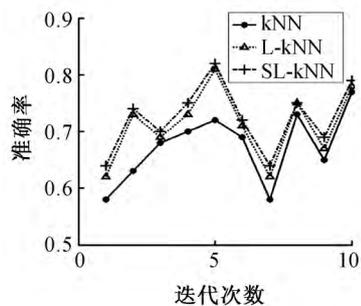


图 5 German
Fig.5 German

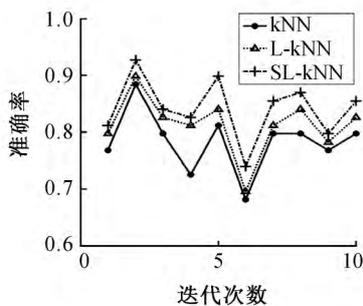


图 6 Australian
Fig.6 Australian

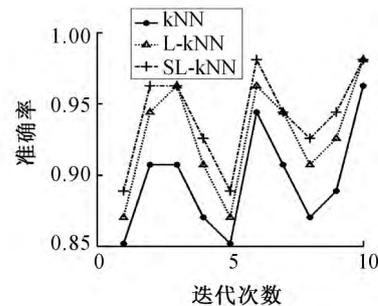


图 7 Climate
Fig.7 Climate

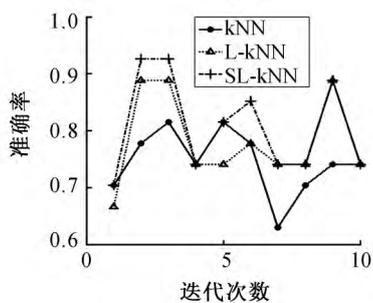


图 8 SPECTF
Fig.8 SPECTF

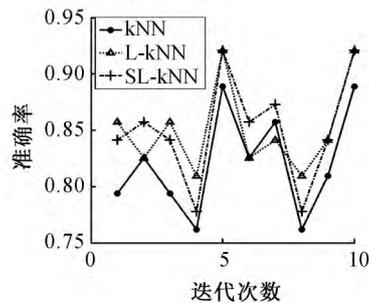


图 9 Balance
Fig.9 Balance

3 结束语

本文针对 kNN 算法在分类问题中遇到的两个问题,即 k 值固定不变问题和如何去除噪声样本问题,提出了一种新的基于稀疏学习的 kNN 分类算法(SL-kNN),用重构获得测试样本与训练样本之间的相关性,用 lasso 产生样本间相关系数的稀疏性,用 l_{21} -范数去除训练数据集中的噪声。在 UCI 数据集上进行的实验结果表明,SL-kNN 算法比 kNN 算法取得更高的分类准确率,对噪声的鲁棒性更优。

参 考 文 献:

- [1] QIN Yongsong, ZHANG Shichao, ZHU Xiaofeng, et al. Semi-parametric optimization for missing data imputation [J]. Applied Intelligence, 2007, 27(1):79-88. DOI: 10.1007/s10489-006-0032-0.
- [2] ZHU Xiaofeng, HUANG Zi, CHENG Hong, et al. Sparse hashing for fast multimedia search[J]. ACM Transactions on Information Systems, 2013, 31(2):9. DOI: 10.1145/2457465.2457469.
- [3] ZHU Xiaofeng, LI Xuelong, ZHANG Shichao. Block-row sparse multiview multilabel learning for image classification [J]. IEEE Transactions on Cybernetics, 2016, 46(2):450-461. DOI:10.1109/TCYB.2015.2403356.
- [4] ZHU Xiaofeng, ZHANG Lei, HUANG Zi. A sparse embedding and least variance encoding approach to hashing [J]. IEEE Transactions on Image Processing, 2014, 23(9):3737-3750. DOI:10.1109/TIP.2014.2332764.
- [5] LIU Huawen, ZHANG Shichao. Noisy data elimination using mutual k -nearest neighbor for classification mining[J]. Journal of Systems and Software, 2012, 85(5):1067-1074. DOI:10.1016/j.jss.2011.12.019.
- [6] ZHU Xiaofeng, HUANG Zi, SHEN Hengtao, et al. Linear cross-modal hashing for efficient multimedia search[C]// Proceedings of the 21st ACM International Conference on Multimedia. New York: ACM Press, 2013: 143-152. DOI: 10.1145/2502081.2502107.

- [7] ZHU Xiaofeng, HUANG Zi, SHEN Hengtao, et al. Dimensionality reduction by mixed kernel canonical correlation analysis [J]. *Pattern Recognition*, 2012, 45(8):3003-3016. DOI:10.1016/j.patcog.2012.02.007.
- [8] HASTIE T, TIBSHIRANI R, FRIEDMAN J. 统计学习基础:数据挖掘、推理与预测[M]. 范明, 柴玉梅, 咎红英, 译. 北京:电子工业出版社, 2003:7-8.
- [9] ZHU Xiaofeng, SUK H, SHEN D G. Matrix-similarity based loss function and feature selection for Alzheimer's disease diagnosis[C]//Proceeding of IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos, CA:IEEE Computer Society, 2014:3089-3096. DOI:10.1109/CVPR.2014.395.
- [10] ZHU Xiaofeng, HUANG Zi, YANG Yang, et al. Self-taught dimensionality reduction on the high-dimensional small-sized data[J]. *Pattern Recognition*, 2013, 46(1):215-229. DOI:10.1016/j.patcog.2012.07.018.
- [11] ZHU Xiaofeng, ZHANG Shichao, JIN Zhi, et al. Missing value estimation for mixed-attribute data sets[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2011, 23(1):110-121. DOI:10.1109/TKDE.2010.99.
- [12] ZHU Xiaofeng, HUANG Zi, CUI Jiangtao, et al. Video-to-shot tag propagation by graph sparse group lasso[J]. *IEEE Transactions on Multimedia*, 2013, 15(3):633-646. DOI:10.1109/TMM.2012.2233723.
- [13] NESTEROV Y. Introductory lectures on convex optimization: A basic course[M]. Berlin: Springer, 2004.
- [14] ZHU Xiaofeng, LI Xuelong, ZHANG Shichao, et al. Robust joint graph sparse coding for unsupervised spectral feature selection [J]. *IEEE Transactions on neural networks and learning systems*, 2016, PP(99):1-13. DOI:10.1109/TNNLS.2016.2521602.

kNN Classification Based on Sparse Learning

ZONG Ming^{1,2}, GONG Yonghong³, WEN Guoqiu¹, CHENG Debo^{1,2}, ZHU Yonghua⁴

(1.College of Computer Science and Information Technology, Guangxi Normal University, Guilin Guangxi 541004, China;
2.Guangxi Collaborative Innovation Center of Multi-source Information Integration and Intelligent Processing, Guigang
Guangxi 537000, China; 3.Department of Information Engineering, Guilin University of Aerospace Technology, Guilin
Guangxi 541004, China; 4.School of Computer, Electronics and Information, Guangxi University,
Nanning Guangxi 530004, China)

Abstract: The value of k is usually fixed in the issue of k Nearest Neighbors (kNN) classification. In addition, there may be noise in train samples which affect the results of classification. To solve these two problems, a sparse-based k Nearest Neighbors (kNN) classification method is proposed in this paper. Specifically, the proposed method reconstructs each test sample by the training data. During the reconstruction process, l_1 -norm is used to generate the sparsity and different k values are used for different test samples, which solves the issue of fixed value of k . And l_{21} -norm is used to generate row sparsity which can remove noisy training samples. The experimental results on UCI datasets show that the proposed method outperforms the existing kNN classification method in terms of classification performance.

Keywords: sparse learning; reconstruction; l_1 -norm; l_{21} -norm; noise sample

(责任编辑 黄 勇)