

文章编号:1672-3961(2010)05-0066-06

一种 ReliefF 特征估计方法在无监督流形学习中的应用

谭台哲, 梁应毅, 刘富春

(广东工业大学计算机学院, 广东 广州 510006)

摘要:针对流形学习存在的对噪声敏感、易受缺失值影响问题以及现实世界数据的结构复杂性和稀疏程序大等问题,提出引入 ReliefF 特征估计,即应用 ReliefF 在流形学习中。实验分4种情况进行:一是不使用特征提取方法;二是仅使用 ReliefF 特征估计方法;三是仅使用有代表性的局部线性嵌入算法;四是使用改进算法。结果表明,改进算法得到的分类准确率分别比单纯使用 ReliefF 特征估计方法和局部线性算法都要高。

关键词:特征选择;ReliefF;流形学习;分类

中图分类号:TP301 **文献标志码:**A

Application of the ReliefF feature evaluation in un-supervised manifold learning

TAN Tai-zhe, LIANG Ying-yi, LIU Fu-chun

(Faculty of Computer, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: As regards to the noise-sensitive, vulnerable to the missing values problem, the complexity and the large sparseness of real world data, This study introduces the ReliefF feature evaluation, which could be applied into manifold learning. The experiments were divided into four cases; one is not to use any feature selection algorithm; one is to use only ReliefF feature evaluation; one is to use only the representative locally linear embedding algorithm; and the last one is to use both. Results showed that the classifying accuracy rate obtained by the improved algorithm was higher than that the obtained by the ReliefF or locally linear embedding respectively.

Key words: feature selection; ReliefF; manifold learning; classification

0 引言

在机器学习和数据挖掘领域,虽然已经提出了很多方法,如关联规则和相关分析、分类和预测、聚类等,但是在实际应用中,数据预处理仍影响着这些方法性能的稳定。特征选择或特征抽取处于数据预处理阶段,成为继数据清理(填充缺失值、光滑噪声并识别离群点和纠正不一致性)^[1]之后知识发现的又一重要步骤。特征选择是从一个给定的数据集中找出一组强相关性的特征子集或者能够代表原来大部分重要信息的数据子集。通常,按照算法是否独立,分为过滤式(filter)和封装式(wrapper)两种。过滤式算法作为一个过滤器筛选出需要的特征,如有信息增益、增益率、Gini 指标和 Relief 特征估计^[2]等;而封装式算法使用启发式算法作为函数的一部分来估计特征子集,如有神经网络、流形学习等^[3]。过滤式方法在计算上有较大优势,但性能上却比封装式好。给定标准函数,特征选择退化为一个搜索问题。穷举搜索虽然是优先考虑的方法,但随

收稿日期:2010-04-15

基金项目:国家自然科学基金资助项目(60974019);广东省自然科学基金资助项目(9451009001002686)

作者简介:谭台哲(1975-),男,山东莱阳人,副教授,博士,主要研究方向为数字图像处理,模式识别与机器学习,生物特征识别。

E-mail: taizhetan@gdut.edu.cn

着问题规模的增加,计算量很快就变得不可控制;启发式组合搜索也是通常使用的一种方法并且已经成功应用在实际中^[4],这些方法都不能保证是最优的。过滤式方法完全忽略了归纳算法的偏倚以及特征子集对归纳算法的性能影响,因此不能有效地去掉冗余、不相关甚至不利于泛化的特征;封装式方法在计算性能上增加了时间复杂度,对于大型数据库以及要求计算量大的算法,则不能应用在实时系统。本文利用二者的优点提出一种过滤-封装式的方法,即在无监督流形学习基础上引入 ReliefF 特征估计。

在现有的特征选择方法中,Relief 方法因为简单易于理解以及算法的有效性,被认为是一种成功的方法。估计特征质量好坏的启发式方法主要是假设特征与目标变量之间是条件独立的,在可能涉及许多特征相互作用的问题中,这会造成特征估计的结果不准确。而 Relief 方法考虑了特征之间的相互信息,因而也是有效的。在特征间存在强依赖性的问题中,Relief 能够正确估计特征的质量。在特征间有相关性很强的问题上,它是一种有效地能够正确估计特征质量的方法^[5]。但是,Relief 仅限于数据完整、没有噪声这两种类别的情况。因此有人提出了其扩展方法 ReliefF^[6]。从 ReliefA 到 ReliefD,仅介绍了处理不完整数据的方法;而从 ReliefE 到 ReliefF 则是在之前基础上的改进,使其能够在多类别的数据集中得到应用。Relief 及其扩展 ReliefF 在算法上均容易理解,运行效率较高,适合使用在实时系统设计中。但同时存在不足之处:ReliefF 不能处理在现实世界数据库中范畴类型的特征以及需设置从数据集抽样得到的样本规模大小^[5]。

流形学习最初以等距映射 ISOMAP^[7]和局部线性嵌入 LLE (locally linear embedding)^[8-9]为代表,随后得到广泛的关注。Laplacian Eigen-map^[10]、Semi-definite Embedding^[11]、Hessian LLE^[12]以及 Local Tangent Space Alignment^[13]等方法相继提出。由于在现实世界中,数据的结构复杂,且多不呈现流形,以及流形学习方法对噪声、不完整数据敏感,从而不能直接使用流形学习的方法进行低维嵌入。本文把 ReliefF 应用在无监督流形学习中,先去除部分不相关或冗余的特征而得到一个特征子空间,再在这个子空间中使用流形学习的方法,从而实现了一种新的过滤-封装式特征选择方法。

1 算法实现

1.1 Relief 特征估计

Relief 特征估计方法由 Kira 和 Rendell 于 1992 年提出,是基于特征权重的一种算法。Relief 方法能够检测那些在统计上与目标属性不相关的特征。其算法公式如下:

$$W = W - \text{diff}(X_R, \text{Near_Hit})^2 / N + \text{diff}(X_R, \text{Near_Miss})^2 / N, \quad (1)$$

其中, W 为一组特征权重向量, A 为属性, N 为样本数。

如果 x_i 和 x_j 是序数类型,则

$$\text{diff}(x_i, x_j) = \begin{cases} 0, & x_i = x_j, \\ 1, & x_i \neq x_j. \end{cases} \quad (2)$$

如果 x_i 和 x_j 是数值类型,则

$$\text{diff}(x_i, x_j) = (x_i - x_j) / \text{nu}, \quad (3)$$

其中, nu 是标准化单位,使其属性之间的差值归一化到 $[0, 1]$ 区间。

Relief 方法对两类情况进行讨论。算法首先随机抽取样本点 X_R , 找到最近邻的两类点 Near_Hit (同类) 和 Near_Miss (异类), 然后分别计算 X_R 与 Near_Hit 和 Near_Miss 之间特征值差异: 对同一类点 Near_Hit 的属性值相减; 对异类点 Near_Miss 的特征值相加, 如此进行与数据规模 N 同样大小的次数循环, 最后得到所有特征的平均权重, 这些权重反映了与目标属性的相关性 (值越大相关性越强)。这是 Relief 方法的基本思想。

虽然 Relief 方法能够处理离散的和连续的特征, 但仅限于二分类的情况。对于处理不完整数据以及多类别的情况, Relief 还不能很好地解决。Igor 于 1993 年提出 ReliefF, Relief 的一种扩展, 使得 Relief 能够处理噪声、不完整特征以及多类别属性的数据集。其改进的算法公式如下:

$$W[A] = W[A] - \sum_{i=1}^k \text{diff}(A, X_R, \text{Near_Hit}_i) / (N \cdot k) + \sum_{C \neq \text{class}(C)} \left[\frac{P(C)}{1 - P(C)} \sum_{i=1}^k \text{diff}(A, X_R, \text{Near_Miss}_i(C)) / (N \cdot k) \right], \quad (4)$$

其中, k 为近邻点个数, $P(C)$ 为类别 C 的概率, 其它字符与上述相同。

如果 x_i 为未知特征,则

$$\text{diff}(A, x_i, x_j) = 1 - P(\text{value}(A, x_j) / \text{class}(x_i)). \quad (5)$$

如果 x_i 和 x_j 均为未知特征,则

$$\text{diff}(A, x_i, x_j) = 1 - \sum_{V \in \text{value}(A)} (P(V | \text{class}(x_i)) \times P(V | \text{class}(x_j))), \quad (6)$$

其中, V 是属性 A 的所有值集合。

ReliefF 方法通过计算不同类别的 k 个近邻点而不是一个近邻点的平均值,达到平滑噪声的目的。这是与 Relief 方法不同的地方。但同时因为对每一类都计算 k 个近邻点而增加了计算量。

1.2 无监督流形学习

等距映射 ISOMAP 在人脸、手势、手写数字“2”数据集以及 Swiss Roll 数据集中都能得到很好的低维嵌入,这种低维结构表现了原数据空间的显著特征。在人脸数据集中,低维嵌入表现了人脸在水平和垂直方向的自由摆动程度以及光照明暗程度;在手势数据集中,低维嵌入表现了手掌在同一水平方向上的伸展程度;在数字“2”数据集中,低维嵌入表现了“2”在书写时出现顶部圆圈和底部弧形的程度;而在 Swiss Roll 数据集中,低维嵌入则更易于表现原数据空间中的流形结构。

ISOMAP 方法算法的基本思想:在使用多维尺度变换 MDS(multidimensional scaling)^[14](一种把相异信息嵌入到欧几里德空间的经典方法)前对其相异度量矩阵(即欧氏距离矩阵)改进,引入测地距矩阵(用 Dijkstra 求最短路径可得),测地距矩阵保持了流形结构中任意两点间的距离,而用欧氏距离矩阵却不能保持。原文使用 Swiss Roll 数据集能很好地解释这种距离关系。ISOMAP 可以看成是 MDS 的扩展,其描述分 3 步骤:

步骤 1 构造 k 近邻点图;

步骤 2 利用最短路径求取点与点之间的测地距离(流形结构中的距离);

步骤 3 使用 MDS 找到低维欧几里德空间中的点,这些点之间的距离与步骤 2 的距离相同。

局部线性嵌入 LLE 算法能够从高维数据点 \vec{X}_i 映射到低维嵌入向量 \vec{Y}_i 。在 S-Curve 数据集中得到的低维嵌入明显地表现了原数据空间流形结构中特征间的关系;在人脸数据集中,低维嵌入在同一区域反映了相似的姿势和表情;在百科文档数据中,低维嵌入能够发现连续的语义相关空间。

LLE 算法的基本思想:在原数据空间中的点通过其 k 近邻点来构造,首先重构得到一组满足特定条件的含 k 个 W_i 的值,然后利用这组值再重构低维空间中的点,这些点从局部上保留了原空间中的特征关系。其描述分 3 步骤:

步骤 1 计算每一个点 \vec{X}_i 的近邻;

步骤 2 计算能够最好重构每一个点 \vec{X}_i 的权重 W_{ij} ,使其最小化代价函数

$$\varepsilon(W) = \sum_i |\vec{X}_i - \sum_j W_{ij} \vec{X}_j|^2; \quad (7)$$

步骤 3 利用权重 W_{ij} 来重构向量 \vec{X}_i ,使得最小化二次型:

$$\Phi(Y) = \sum_i |\vec{Y}_i - \sum_j W_{ij} \vec{Y}_j|^2. \quad (8)$$

1.3 改进算法

由于 Relief 对噪声、不完整和多类别数据不能很好地进行特征估计,以及流形学习对噪声敏感或原数据空间不呈现流形结构,本文提出在 ReliefF 的基础上进行流形学习,即先用 ReliefF 估计特征的权重大小,据此选择出权重大与阈值的特征(用户通过观察设定值或者使用下文的参数选择方法),然后对特征选择后的数据空间再进行无监督流形学习,进而得到低维嵌入,实现了过滤-封装式特征选择。由于这两种方法的使用都涉及到参数选择,故一次运行过程达不到效果,可循环往复地进行实验,直到得到满意结果为止,算法描述如下:

输入 原始数据 $X_{N \times D}$, 阈值 τ , 低维数 d , 近邻数 k_1, k_2

输出 低维嵌入 Y

算法

//使用 ReliefF 进行特征估计

(1) 进入 N 次循环;

- (2) 每次随机选择一个样本点 X_R ;
- (3) 计算一次每个属性的特征权重(每次循环将重复计算一次)

$$W[A] = W[A] - \sum_{i=1}^k \text{diff}(A, X_R, \text{Near_Hit}_i) / (N \cdot k) + \sum_{C \neq \text{class}(C)} \left[\frac{P(C)}{1 - P(C)} \sum_{i=1}^k \text{diff}(A, X_R, \text{Near_Miss}_i(C)) / (N \cdot k) \right];$$

- (4) 进入无条件循环 while;
- //对权重大于阈值的属性进行选择(或者对权重进行递增减排序,按百分比选择)
- (5) 对于每个属性的权重,如果 $W[i] > \tau$,则 f_i 留下;//相关的属性留下
- (6) 否则 f_i 去掉;//不相关的属性去除
- (7) 使用流形学习方法进行维度归约得到低维嵌入 Y ;
- (8) 如果 Y 不满足,则对 τ, d, k_2 重新赋值;
- (9) 否则跳出 while 循环;
- (10) 结束 N 次循环;

其中,

τ : 阈值决定了被选择的特征,大于它的特征留下,否则去掉。有两种方式可以设定阈值: $\tau \leq \frac{1}{\sqrt{\alpha \cdot N}}$ (α 为假设检验的置信区间值,可为 0.01, 0.05, 0.1 等; N 为数据集的规模大小); 通过观察 τ 的分布(对权重进行递增减排序但不改变特征顺序,然后选择部分权重大的特征)。

k_1 : ReliefF 算法中的近邻数,根据数据集特征数及规模大小设定其大小, k_1 越大计算量就越大。

d : 流形学习中低维嵌入的维度,从低维往高维逐渐增加。

k_2 : 流形学习使用的近邻数,值越大计算量也越大。

2 实验结果与分析

2.1 数据来源

为了体现算法的有效性,实验数据源的属性尽可能大于 1 000 个。如此,用 Relief 估计特征并按特征权重递减排列或者设定阈值,按一定方法选择属性得到原数据的特征子空间,保证再进行流形学习有足够的属性。本文以 UCI(university of California irvine 机器学习数据库,搜集了各种数据库、相关领域理论研究成果以及数据生成器等,作为机器学习算法的经验分析工具,内容涵盖了生命科学、物理科学、计算机科学/工程、社会科学、经济、文本等领域)^[15] 的 Arcene 数据集为实验对象,具体信息如表 1(其中,900 个实例有 100 个训练集,700 个无类别标号测试集,100 个 valid 集)。

表 1 Arcene 数据集具体信息
Table 1 Information of Arcene data set

数据集	数据类型	属性类型	样本数	属性数
Arcene	多变量	实数值	900	10 000

2.2 实验步骤

实验以 100 个训练集进行(目的仅是使用不同方法下的分类结果进行评判改进算法的有效性,而不是预测未知分类的测试集),用 WEKA(waikato environment for knowledge analysis,是用 Java 语言编写的一套机器学习工具,包含了各种数据分析和预测模型的可视化工具和算法,并提供一个图形用户界面以方便使用)^[16] 中的 ReliefFAttributeEval 和 LibSVM^[17] 工具以及现有的流形学习方法 LLE,并加以 10 层交叉验证。具体步骤如下:

- 步骤 1 设置 τ, d, k_1, k_2, X 参数;
- 步骤 2 使用本文提出的算法对数据集进行实验;
- 步骤 3 对结果分析并加以 10 层交叉验证;
- 步骤 4 若分类结果不满足,返回步骤 1,否则结束。

本文分别对数据集进行4种不同情况的实验:(1)不使用任何特征选择方法;(2)仅使用 ReliefF 进行属性子集选择;(3)仅使用流形学习方法 LLE 进行维度归约;(4)同时使用 ReliefF 和 LLE。由于 SVM 是目前很好的分类器,故实验中用 SVM 对上面4种情况下得到的数据进行分类,验证所提出方法的有效性。

2.3 结果分析

表2最左列表示流形学习的低维嵌入,从2维逐渐增加到45维;顶行表示 ReliefF 估计特征后,按权重递减顺序排列后选择权重大的部分特征百分比(不改变原数据空间中特征的位置,或者通过设定阈值选择权重大于它的特征),从100%逐渐递减到5%。该表显示了4种不同情况的实验数据:

情况1 当 d 为 10 000(即不降维)和不选择特征(100%)时,对应于不使用任何特征选择方法;

情况2 当 d 为 10 000(即不降维)和选择特征(5%~95%)时,对应于仅使用 ReliefF 方法进行属性子集选择;

情况3 当 $0 < d < 10\,000 (d \in \mathbf{N})$ 和不选择特征(100%)时,对应于仅使用流形学习方法 LLE 进行维度归约;

情况4 当 $0 < d < 10\,000 (d \in \mathbf{N})$ 和选择特征(5%~95%)时,对应于使用 ReliefF 和 LLE 同时进行。

表2 ReliefF 特征估计+流形学习的实验数据
Table 2 Experimental data of ReliefF feature evaluation & manifold learning

维度 d	ReliefF 特征估计后按百分比(%)选择的属性数																			
	100	95	90	85	80	75	70	65	60	55	50	45	40	35	30	25	20	15	10	5
10 000	56	56	56	56	56	56	56	56	56	56	56	56	56	56	56	56	56	56	56	56
2	60	60	60	60	60	63	63	60	63	60	60	60	60	60	60	60	60	63	63	63
3	60	60	60	63	60	60	60	60	60	60	61	60	60	60	60	60	60	60	60	60
4	61	62	64	61	61	61	68	63	64	62	60	62	63	62	62	63	63	65	60	62
5	66	65	61	64	66	67	57	68	61	64	61	64	66	65	65	59	65	67	61	62
6	67	65	70	65	67	67	55	67	67	65	64	65	65	64	64	65	66	67	58	65
7	69	66	61	66	67	67	70	68	68	65	65	65	66	64	64	65	66	67	65	66
8	68	77	68	69	72	70	72	72	72	73	65	72	73	72	72	62	74	75	76	68
9	74	76	75	76	78	77	64	80	78	76	73	78	78	76	76	64	78	81	60	75
10	78	75	72	77	78	78	79	66	78	78	79	78	78	78	74	72	79	80	76	77
11	78	78	71	78	78	79	80	78	78	78	78	79	79	77	79	71	79	80	82	78
12	78	79	75	80	79	79	82	80	79	81	76	78	79	75	75	79	77	80	78	76
14	81	80	76	82	79	79	79	79	80	78	77	77	78	76	76	78	79	79	75	75
16	82	80	81	82	78	80	83	83	80	79	77	75	81	79	79	76	82	81	80	78
18	81	82	82	85	77	79	84	83	79	79	79	78	82	77	77	76	82	81	78	81
20	84	82	84	87	87	86	85	86	87	79	85	87	85	78	78	78	82	80	79	82
25	82	81	85	83	82	83	83	81	81	81	81	85	82	81	81	81	83	82	79	79
30	80	78	79	77	80	78	78	79	79	80	80	82	80	83	82	81	83	81	78	76
35	78	78	78	76	78	79	79	78	80	77	78	81	81	80	80	77	80	76	75	75
40	79	79	81	80	79	77	79	75	80	79	80	79	81	79	79	77	78	78	80	75
45	80	78	76	76	76	77	77	80	77	78	77	76	82	82	82	76	82	79	77	74

粗体数字表示当前列中最大值,也即最大分类准确率。可以看出,在选择不同范围的特征数后以及低维嵌入为20维的数据空间中,准确率最高个数是最多的。这比单纯使用 ReliefF 特征估计和流形学习方法都提高了分类准确率,而且是有效的。

为了能更清晰地了解这两种方法同时使用后的分类准确率变化趋势,图1使用了表2中部分实验数据结果(即 $d = \{45, 35, 25, 20, 18, 7, 6, 5, 4, 3, 2\}$, 如此避免因线条过多而杂乱)。图1表明,并不是特征数越多或越少以及低维嵌入的维度越大或越小,分类准确率就会越高。在本次实验中,低维嵌入的维度在20的时候是最优的。

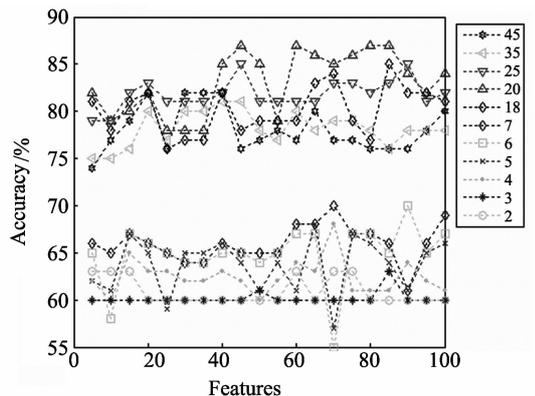


图1 实验数据部分图示
Fig. 1 Partial data of the experimental results

参数的选择在本文提出的算法中显得很重要,因此允许进行参数的循环调整。对不同的数据集,可能会出现完全不同的参数设置;以及选用不同的流形学习方法,同一实验数据集也有不同的结果。比如,在本文同样的实验环境中,使用 ISOMAP 算法比 LLE 算法耗费的时间要长 (ReliefF 特征估计的时间相同)。表明 ReliefF 特征估计和流形学习两者结合使用的方法比单一使用的效果好。但同时也存在着参数选择问题。

3 结论

在数据挖掘领域,数据预处理对后期关联规则和相关分析、分类和预测以及聚类等知识发现方法有很大帮助。特征选择包含在数据预处理阶段,有过滤式和封装式两种。本文提出一种改进方法是基于两者已有的算法,即 ReliefF 特征估计和无监督流形学习。利用 ReliefF 方法有较好的抗噪声、处理数据不完整和多类别特点以及无监督流形学习方法的高度降维特性,改进算法能够较好地 Arcene 数据集进行特征选择。由于改进算法是在组合两种特征选择方法的基础上提出,原来的参数选择问题仍然存在。故算法加入了调整参数的循环过程,以改进不同数据集不同参数的问题。将来的工作可以在这方面进行完善。

参考文献:

- [1] 韩家炜, 堪博. 数据挖掘: 概念与技术[M]. 范明, 孟小峰译. 2 版. 北京: 机械工业出版社, 2007.
HAN Jiawei, KAMBER Micheline. Data mining: concepts and techniques [M]. FAN Ming, MENG Xiaofeng tran. 2ed. Beijing: China Machine Press, 2007.
- [2] KIRA Kenji, RENDELL Larry A. The feature selection problem: traditional methods and a new algorithm[C]//Proceedings of the 10th National Conference on Artificial Intelligence. California, MA: The MIT Press, 1992: 129-134.
- [3] HUANG Yuan, SHIAN-SHYONG Tseng, WU Gangshan, et al. A two-phase feature selection method using both filter and wrapper[C]//IEEE International Conference on Systems. [S. l.]: IEEE, 1999: 132-136.
- [4] SUN Yijun, WU Dapeng. A RELIEF based feature extraction algorithm[C]//Proceedings of the 8th SIAM International Conference on Data Mining. [S. l.]: Curran Associates Inc, 2008: 188-195.
- [5] YUE Huang, MCCULLAGH Paul J, BLACK Norman D. An optimization of ReliefF for classification in large datasets [J]. Data & Knowledge Engineering, 2009, 68: 1348-1356.
- [6] KONONENKO Igor. Estimating attributes: analysis and extensions of Relief[C]//Proceedings of European Conference on Machine Learning. [S. l.]: Springer Verlag, 1994: 171-182.
- [7] TENENBAUM Joshua B, SILVA Vin de, LANGFORD John C. A global geometric framework for nonlinear dimensionality reduction [J]. Science, 2000, 290: 2319-2323.
- [8] ROWEIS Sam T, SAUL Lawrence K. Nonlinear dimensionality reduction by locally linear embedding [J]. Science, 2000, 290: 2323-2326.
- [9] SAUL Lawrence K, ROWEIS Sam T. Think globally, fit locally: unsupervised learning of low dimensional manifolds [J]. Journal of Machine Learning Research, 2003, 4: 119-155.
- [10] BELKIN Mikhail, NIYOGI Partha. Laplacian eigenmaps and spectral techniques for embedding and clustering[C]//Proceedings of the 2001 Neural Information Processing Systems Conference. [S. l.]: The MIT Press, 2001: 585-591.
- [11] WEINBERGER Kilian Q, SHA Fei, SAUL Lawrence K. Learning a kernel matrix for nonlinear dimensionality reduction [C]//Proceedings of the 21st International Conference on Machine Learning. [S. l.]: [s. n.], 2004: 839-846.
- [12] DONOHO David L, GRIMES Carrie. Hessian eigenmaps: new locally linear embedding techniques for high-dimensional data [C]//Proceedings of the National Academy of Sciences. [S. l.]: PNAS, 2003: 5591-5596.
- [13] ZHANG Zhenyue, ZHA Hongyuan. Principal manifolds and nonlinear dimension reduction via local tangent space alignment [J]. SIAM Journal of Scientific Computing, 2004, 26(1): 313-338.
- [14] COX T F, COX M A A. Multidimensional scaling [M]. 2ed. [S. l.]: Chapman and Hall/CRC, 2001.
- [15] ASUNCION A, NEWMAN D J. UCI machine learning repository [DB/OL]. Irvine, CA: University of California, School of Information and Computer [2010-03-26]. <http://www.ics.uci.edu/~mlern/MLRepository.html>, 2007.
- [16] HALL Mark, FRANK Eibe, HOLMES Geoffrey, et al. The WEKA data mining software: an update [J]. SIGKDD Explorations, 2009, 11(1): 10-18.
- [17] CHANG Chihchung, LIN Chihjen. LIBSVM: a library for support vector machines [CP/OL]. [2010-03-26]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.