文章编号: 1001-0920 (2015) 01-0032-07 DOI: 10.13195/j.kzyjc.2013.1673

基于概率类和不相关判别的半监督局部 Fisher 方法

王寅同, 王建东, 陈海燕, 孙 博

(南京航空航天大学 计算机科学与技术学院, 南京 210016)

摘 要: Fisher 判别分析是统计模式识别中经典的有监督维数约简方法, 可以在最大化类间散度的同时最小化类内散度, 但存在分析过程中仅使用有标记数据而忽略无标记数据的问题. 鉴于此, 提出基于概率类和不相关判别的半监督局部 Fisher (SLFisher) 方法, 以实现半监督学习的高维映射到低维的类间数据对尽可能地分离, 且类内邻近数据尽可能地紧凑. 采用 2 组标准数据集进行实验, 结果表明了 SLFisher 方法能够有效提高识别率.

关键词: Fisher 判别分析; 维数约简; 概率类; 不相关判别; 半监督学习

中图分类号: TP181 **文献标志码:** A

Semi-supervised local Fisher method based on probability class and uncorrelated discriminant

WANG Yin-tong, *WANG Jian-dong*, *CHEN Hai-yan*, *SUN Bo*

(School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China. Correspondent: WANG Yin-tong, E-mail: wangyintong@nuaa.edu.cn)

Abstract: The Fisher discriminant analysis(FDA) is a classical supervised dimensionality reduction method in statistical pattern recognition. The FDA can maximize the scatter between different classes, while minimizing the scatter within each class, but the analysis process of the FDA only utilizes the labeled data and ignores the unlabeled data. Therefore, a semi-supervised local fisher method based on probability class and uncorrelated discriminant(SLFisher), which enables the data pairs in different classes to be separated from each other and the nearby data pairs in the same class to be closed after dimensionality reduction. Two benchmark datasets are applied in the experiment, and the results show that the SLFisher can greatly improve recognition rate.

Keywords: Fisher discriminant analysis; dimensionality reduction; probability class; uncorrelated discriminant; semi-supervised learning

0 引 言

随着数据数量和维度的不断增长, 从数据中发掘知识变得越来越困难. 高维数据的低维表示已成为必不可少的预处理过程, 这样不仅可以提高学习算法的速度、鲁棒性和精确度, 而且有助于最终结果的可解释性. 同时, 模式识别和机器学习的实践也证明了维数约简是一种有效方法, 即从高维数据中抽取出反映数据本质特征的低维结构[1].

根据是否需要利用数据本身的类别判别信息, 数据的维数约简方法可以划分为无监督[2-5]、有监督[6-8]和半监督[9-13]3 个主要类别. 在工程实践中, 对数据的标记工作是费时又昂贵的, 而无标记数据的获取则容易得多. 值得注意的是, 相对于无监督和有监督维数约简方法, 研究仅部分数据有标记的数据集上的半监督学习方法尤为重要. 半监督判别分析 (SDA) 方法[14]在维数约简过程中同时利用有标记数据和无标记数据, 对有标记数据保持类间散度最大化, 通过无标记数据估计数据的内在空间结构. 半监督局部 Fisher 判别分析 (SELF)[15]作为另一种重要的半监督维数约简方法, 保持了无标记数据的全局结构, 并使得类间的有标记数据相互分离. 这两种方法有一个共同的特点, 均需要一个额外的参数用于权衡无监督和

收稿日期: 2013-12-02; **修回日期:** 2014-02-26.

基金项目: 国家自然科学基金重点项目(61139002); 国家 863 计划项目(2012AA063301); 中央高校基本科研业务费专项资金项目(NS2012134, NZ2013306); 江苏省博士后计划项目(1301013A); 中国民航信息技术科研基地开发基金项目(CAAC-ITRB-201203).

作者简介: 王寅同(1987−), 男, 博士生, 从事数据挖掘、人工智能的研究; 王建东(1945−), 男, 教授, 博士生导师, 从事人工智能、信息安全等研究.

有监督学习的比重. 事实上, 参数的选择会因所处理数据的不同而改变, 因此难以确定. 此外, 降维后的特征之间存在统计相关性, 阻碍了利用更少的特征表示足够的原始数据信息.

为了解决上述问题, 本文提出一种基于概率类和不相关判别的半监督局部Fisher (SLFisher) 方法. 该方法在有监督的局部Fisher判别分析的基础上, 引入概率类和不相关判别实现的半监督维数约简, 实现了半监督学习的高维映射到低维的类间数据对尽可能分离, 且类内邻近数据尽可能紧凑. 实验分析结果表明了SLFisher方法能够有效提高识别率.

1 Fisher判别分析

维数约简问题描述如下: 对于n个样本的数据集X, 每个样本$x_i \in \boldsymbol{R}^d (i = 1, 2, \cdots, n)$包含$d$个特征且仅属于一个类别. 记类别标号为$l_i \in \{1, 2, \cdots, c\}$, 其数学表达形式为$X \in \{(x_i, l_i)\}_{i=1}^n$, 记类别标号为$l$的类别中包含$n_l$个样本, 所有类别的样本之和为$\sum\limits_{l=1}^{c} n_l = n$. 假设高维空间样本$x_i$对应的低维空间数据为$y_i \in \boldsymbol{R}^r (1 \leqslant r \ll d)$, 线性维数约简的目标是寻找到一个$d \times r$转换矩阵$T$, 使得等式$Y = T^{\mathrm{T}} X$成立. 其中: r为低维空间维度, T^{T}为矩阵T的转置.

Fisher判别分析(FDA)[16-17]是维数约简和模式分类的一种有监督学习方法, 其主要思想是使得降维后的数据在类间散度最大化的同时类内散度最小化. 因为有监督学习使用的监督信息既可以是各个数据均有属于自身类别的标记约束, 也可以是一对数据属于同一类的成对约束, 所以本文引入标记约束和成对约束两种等价的Fisher判别分析, 区别在于它们的类间散度和类内散度的表示方式不同.

1.1 标记约束Fisher判别分析

在标记约束Fisher判别分析中, 类间散度矩阵$S^{(b)}$和类内散度矩阵$S^{(w)}$的计算分为两步: 1) 以类为单位, 计算各个类的散度; 2) 将所有类的散度进行合并, 从而得到所需的类间散度矩阵和类内散度矩阵. 具体定义为

$$S^{(b)} = \sum_{l=1}^{c} n_l \big(\mu_l - \mu\big)\big(\mu_l - \mu\big)^{\mathrm{T}}, \tag{1}$$

$$S^{(w)} = \sum_{l=1}^{c} \sum_{i:l_i=l} \big(x_i - \mu_l\big)\big(x_i - \mu_l\big)^{\mathrm{T}}. \tag{2}$$

其中: $\sum\limits_{i:l_i=l}$为数据集中类别标号为l的数据集合, $\mu_l = \dfrac{1}{n_l} \sum\limits_{i:l_i=l} x_i$为数据集中类别标号为$l$的数据平均值, $\mu = \dfrac{1}{n} \sum\limits_{i}^{n} x_i$为所有数据平均值.

Fisher准则函数的定义与实际情况结合得较紧密, 采用的准则函数定义如下:

$$J_{\mathrm{FDA}}(T) = \arg \max_{T \in R^{d \times r}} \frac{T^{\mathrm{T}} S^{(b)} T}{T^{\mathrm{T}} S^{(w)} T}. \tag{3}$$

准则函数$J_{\mathrm{FDA}}(T)$的求解可变换为特征分解问题, $S^{(b)}T = \lambda S^{(w)}T$, 其中Fisher判别分析的转换矩阵$T$由$r$个最大特征值所对应的特征向量$t_1, t_2, \cdots, t_r$组成.

1.2 成对约束Fisher判别分析

与标记约束相比, 成对约束是一种更一般、更易获取的有监督信息, 分为正约束和负约束, 具有正约束关联的数据表明属于同一类, 具有负约束关联的数据表明属于不同类. 成对约束Fisher判别分析的类间散度矩阵$S^{(b)}$和类内散度矩阵$S^{(w)}$的计算都是在数据对的基础上进行的, 具体定义如下.

首先, 计算成对约束的类内散度矩阵$S^{(w)}$, 对式(2)进行代数变换, 有

$$\begin{aligned} S^{(w)} = & \\ & \sum_{l=1}^{c} \sum_{i:l_i=l} \Big(x_i - \frac{1}{n_l} \sum_{j:l_j=l} x_j\Big)\Big(x_i - \frac{1}{n_l} \sum_{j:l_j=l} x_j\Big)^{\mathrm{T}} = \\ & \sum_{i=1}^{n} x_i x_i^{\mathrm{T}} - \sum_{l=1}^{c} \frac{1}{n_l} \sum_{i,j:l_i=l_j=l} x_i x_j^{\mathrm{T}}. \end{aligned} \tag{4}$$

假设同类数据x_i和x_j的类内权重为$W_{i,j}^{(w)}$, 且$\sum\limits_{j:l_j=l_i}^{n} W_{i,j}^{(w)} = 1$, 代入式(4)得到

$$\begin{aligned} S^{(w)} = & \\ & \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} W_{i,j}^{(w)}\Big) x_i x_i^{\mathrm{T}} - \sum_{i,j=1}^{n} W_{i,j=1}^{(w)} x_i x_j^{\mathrm{T}} = \\ & \frac{1}{2} \sum_{i,j=1}^{n} W_{i,j}^{(w)} (x_i x_i^{\mathrm{T}} - x_i x_j^{\mathrm{T}} - x_j x_i^{\mathrm{T}} + x_j x_j^{\mathrm{T}}). \end{aligned} \tag{5}$$

然后, 结合总散度矩阵、类内散度矩阵和类间散度矩阵3者之间的等式关系, 计算类间散度矩阵$S^{(b)}$, 有

$$S^{(t)} \equiv S^{(w)} + S^{(b)} = \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^{\mathrm{T}}. \tag{6}$$

对式(6)进行代数变换并整理, 得到

$$\begin{aligned} S^{(b)} = S^{(t)} - S^{(w)} = & \\ & \sum_{i=1}^{n} x_i x_i^{\mathrm{T}} - \frac{1}{n} \sum_{i,j=1}^{n} x_i x_j^{\mathrm{T}} - S^{(w)} = \\ & \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} \frac{1}{n}\Big) x_i x_i^{\mathrm{T}} - \sum_{i,j=1}^{n} \frac{1}{n} x_i x_j^{\mathrm{T}} - S^{(w)}. \end{aligned} \tag{7}$$

最后, 整理式(5)和(7), 得到如下成对约束Fisher判别分析的类间散度矩阵、类内散度矩阵、类间数据

对权重和类内数据对权重:

$$S^{(b)}=\frac{1}{2}\sum_{i,j=1}^{n}W_{i,j}^{(b)}(x_i-x_j)(x_i-x_j)^{\mathrm{T}}. \tag{8}$$

$$S^{(w)}=\frac{1}{2}\sum_{i,j=1}^{n}W_{i,j}^{(w)}(x_i-x_j)(x_i-x_j)^{\mathrm{T}}. \tag{9}$$

$$W_{i,j}^{(b)}=\begin{cases}1/n-1/n_{l_i},\ l_i=l_j;\\ 1/n,\ l_i\neq l_j.\end{cases} \tag{10}$$

$$W_{i,j}^{(w)}=\begin{cases}1/n_{l_i},\ l_i=l_j;\\ 0,\ l_i\neq l_j.\end{cases} \tag{11}$$

2 概率类和不相关判别的半监督局部Fisher方法

2.1 半监督维数约简问题

半监督维数约简问题描述如下: 对于$n+m$个样本的数据集$X=\{x_1,x_2,\cdots,x_{n+m}\}$, 每个样本$x_i\in\boldsymbol{R}^d\,(i=1,2,\cdots,n+m)$包含$d$个特征, 前$n$个样本是有标记数据, 记类别$l_i\in\{1,2,\cdots,c\}$, 后$m$个样本是无标记数据. 假设高维空间样本$x_i$对应的低维空间数据为$y_i\in\boldsymbol{R}^r(1\leqslant r\ll d)$, 半监督维数约简的目标是既利用有标记数据又利用无标记数据寻找到一个$d\times r$转换矩阵T, 使得等式$Y=T^{\mathrm{T}}X$成立. 其中: r为低维空间维度, T^{T}为矩阵T的转置.

SLFisher方法的提出主要基于如下两个问题:

1) SELF等[15]半监督降维方法需要一个额外的参数用于权衡无监督和有监督学习的比重. 事实上, 该参数的选择会因所处理的数据不同而改变, 因此难以确定.

2) 由式(3)得到的转换向量矩阵, 并使得高维数据映射到低维数据时, 低维数据的特征之间存在统计相关性, 这阻碍了利用更少的特征来表示足够的原始数据信息.

对于上述两个问题, SLFisher方法分别引入了重构概率类和不相关判别向量. 重构概率类使得无标记数据以概率方式属于对应类别, 从而建立一个与真实的数据类别较为一致的概率类. 不相关判别向量给出了Fisher准则函数最大化的低维数据特征不相关的充分必要条件, 并在SLFisher方法中通过Lagrange乘子优化得到不相关的Fisher准则函数.

2.2 重构概率类

重构概率类是SLFisher方法的重要组成部分, 计算无标记数据$x_i(n+1\leqslant i\leqslant n+m)$属于最邻近的有标记数据所属类别的概率$p_i(l)$, 进而得到概率矩阵$P_{(n+m)\times C}$, 其中每一行代表数据集的一个数据, 每一列代表一个类别. 概率矩阵中, 每个数据x_i的计算可以分两种情况: 1) 作为有标记的数据x_i所属类别为l, 概率值$p_i(l)=1$表示数据x_i一定是类别l的数据成员, 并设定概率矩阵第i行的其他概率值为0; 2) 作为无标记的数据x_i, 属于类别$l(0\leqslant l\leqslant c)$的概率值应满足$0\leqslant p_i(l)\leqslant 1$和$\sum\limits_l p_i(l)=1$.

关于概率矩阵$P_{(n+m)\times C}$的计算, 考虑相关权重图[3,8,15,18], 假设同类数据对之间的距离相对于非同类数据对更靠近, 仅计算无标记数据属于最邻近有标记数据的类别概率, 即寻找无标记数据的两个最邻近非同类的标记数据, 将它们的欧氏距离比值作为无标记数据属于最邻近类别的概率, 并设定无标记数据属于其他类别的概率为0. 无标记数据$x_i(n+1\leqslant i\leqslant n+m)$所属类别的概率计算步骤如下.

Step 1: 寻找数据x_i的最近邻有标记数据x_j, $\{(x_j,l_j)|x_j\in\boldsymbol{R}^d,\ l_j\in\{1,2,\cdots,c\}\}_{j=1}^{n}$, 并计算数据$x_i$到数据$x_j$的欧氏距离$d_{il_j}$.

Step 2: 寻找数据x_i的次近邻有标记数据x_k, $\{(x_k,l_k)|x_k\in\boldsymbol{R}^d,\ l_k\in\{1,2,\cdots,l_j-1,l_j+1,\cdots,c\}\}_{k=1}^{n}$, 并计算数据$x_i$到数据$x_k$的欧氏距离$d_{il_k}$.

Step 3: 计算无标记数据x_i属于最邻近类别l_j的概率值

$$p_i(l_j)=d_{il_k}/(d_{il_j}+d_{il_k}). \tag{12}$$

由式(12)可知, 无标记数据x_i处于两个非同类的最邻近有标记数据x_j与x_k之间, 若数据x_i接近于数据x_j与x_k的中心位置, 则概率值p_{il}近似等于0.5, 表明数据x_i所属类别的状态不明确. 随着数据x_i偏向于某一个有标记数据, 概率值逐渐接近于1. 综上所述, 所有数据以大于等于0.5的概率属于某一类别, 属于其他类别的概率为0.

2.3 不相关判别向量

假设Fisher判别向量集为T, 其中$t_1,t_2,\cdots,t_k$$(k\geqslant 1)$为$T$已知的前$k$个特征向量. 现计算第$k+1$个特征向量$t_{k+1}$, 并使得准则函数最大化, 则第$k+1$个向量应满足条件

$$t_{k+1}^{\mathrm{T}}t_i=0,\ i=1,2,\cdots,k. \tag{13}$$

由上述约束计算r个优化判别特征向量$t_1,t_2,\cdots,t_r$, 并实现由高维空间$\boldsymbol{R}^d$到低维空间$\boldsymbol{R}^r$的线性转换, 有

$$Y=\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_r\end{bmatrix}=\begin{bmatrix}t_1^{\mathrm{T}}\\ t_2^{\mathrm{T}}\\ \vdots\\ t_r^{\mathrm{T}}\end{bmatrix}X. \tag{14}$$

称式(14)在(13)约束下的线性变换为Foley-Sammon判别转换[19].

定理1 低维空间数据集$Y=T^{\mathrm{T}}X$中, 任何两

个特征向量 y_i 和 $y_j(i \neq j)$ 都是统计不相关的, T 使得优化准则函数最大化的条件是当且仅当转换向量满足如下共轭正交约束:

$$t_i^{\mathrm{T}} S^{(t)} t_j = 0,\ i \neq j. \tag{15}$$

证明　设 Fisher 判别降维后的两个特征向量 y_i 和 y_j 的统计相关计算为

$$E[(y_i - Ey_i)(y_j - Ey_j)] =$$
$$t_i^{\mathrm{T}} E[(X - E(X))(X - E(X))^{\mathrm{T}}] t_j = t_i^{\mathrm{T}} S^{(t)} t_j, \tag{16}$$

其中 $S^{(t)}$ 为数据集 X 的总方差. 一般情况下, 式 (13) 约束下的 (16) 不等于 0. 换言之, 降维后的两个特征向量 y_i 和 $y_j(i \neq j)$ 是统计相关的, 并不能使得优化准则函数在 T 下最大化. 所以, 为了实现 Fisher 判别降维后的两个特征 y_i 和 $y_j(i \neq j)$ 统计不相关, 需要满足

$$E[(y_i - Ey_i)(y_j - Ey_j)] = t_i^{\mathrm{T}} S^{(t)} t_j = 0. \tag{17}$$

另外, 规范化 Fisher 判别向量 t_i, 使得 $t_i^{\mathrm{T}} S_t t_i = 1$, 结合式 (15) 可知, 不相关的 Fisher 判别向量应满足

$$T^{\mathrm{T}} S^{(t)} T = I. \tag{18}$$

综上, 定理 1 得证. □

2.4　SLFisher 方法

SLFisher 作为一种半监督的线性维数约简方法, 综合考虑有标记和无标记数据, 使得降维后的类间数据尽可能分离, 类内邻近数据尽可能紧凑, 且类内非邻近数据尽可能地保持. 在成对约束 FDA 方法的基础上, SLFisher 的半监督局部类间散度矩阵 $S^{(slw)}$ 和半监督局部类内散度矩阵 $S^{(slb)}$ 定义为

$$S^{(slb)} = \frac{1}{2} \sum_{i,j=1}^{n+m} W_{i,j}^{(slb)} (x_i - x_j)\big(x_i - x_j\big)^{\mathrm{T}}, \tag{19}$$

$$S^{(slw)} = \frac{1}{2} \sum_{i,j=1}^{n+m} W_{i,j}^{(slw)} \big(x_i - x_j\big)\big(x_i - x_j\big)^{\mathrm{T}}. \tag{20}$$

其中 $W_{i,j}^{(slb)}$ 和 $W_{i,j}^{(slw)}$ 分别表示类间数据对权值和类内数据对权值, 定义为

$$W_{i,j}^{(slb)} = \begin{cases} \widetilde{A}_{i,j}(1/(n+m) - 1/n_{l_i}),\ l_i = l_j; \\ 1/(n+m),\ l_i \neq l_j. \end{cases} \tag{21}$$

$$W_{i,j}^{(slw)} = \begin{cases} \widetilde{A}_{i,j}/n_{l_i},\ l_i = l_j; \\ 0,\ l_i \neq l_j. \end{cases} \tag{22}$$

$\widetilde{A}_{i,j}$ 为同类数据对的启发式局部度量, 当数据 x_i 与 x_j 之间的距离较远而成为同类非邻近数据对时, 其相似度接近于 0; 当两个数据距离不断缩小而成为同类邻近数据对时, 其相似度更接近于 1; 其相似度等于 1 当且仅当两个数据完全一样. 本文采用启发式局部度量方式计算相似度矩阵 $\widetilde{A}$, 定义为

$$\widetilde{A}_{i,j} = p_{i,j} \exp\Big(- \frac{\|x_i - x_j\|^2}{\sigma_i \sigma_j} \Big). \tag{23}$$

其中: $\sigma_i = \|x_i - x_j^{(k)}\|$ 为数据 x_i 的局部度量; $x_i^{(k)}$ 为数据 x_i 的第 k 个非自身的最近邻数据, k 为超参数, 取值为 7 可以获得较好的局部缩放的高斯相似性[20]; 数据 x_i 和 x_j 属于类别标号 l 的概率值 $p_{i,j} = p_i(l)p_j(l)$.

SLFisher 方法实现高维数据映射到低维时, 在半监督局部类间数据对尽可能地分离的同时, 半监督局部类内邻近数据对尽可能地紧凑, 其形式化表达式为

$$\max \sum_{i,j=1}^{n+m} W_{i,j}^{(slb)} \big(y_i - y_j\big)\big(y_i - y_j\big)^{\mathrm{T}} =$$
$$\max \sum_{i,j=1}^{n+m} W_{i,j}^{(slb)} \big(T^{\mathrm{T}} x_i - T^{\mathrm{T}} x_j\big)\big(T^{\mathrm{T}} x_i - T^{\mathrm{T}} x_j\big)^{\mathrm{T}} =$$
$$\max T^{\mathrm{T}} S^{(slb)} T, \tag{24}$$

$$\min \sum_{i,j=1}^{n+m} W_{i,j}^{(slw)} \big(y_i - y_j\big)\big(y_i - y_j\big)^{\mathrm{T}} =$$
$$\min \sum_{i,j=1}^{n+m} W_{i,j}^{(slw)} \big(T^{\mathrm{T}} x_i - T^{\mathrm{T}} x_j\big)\big(T^{\mathrm{T}} x_i - T^{\mathrm{T}} x_j\big)^{\mathrm{T}} =$$
$$\min T^{\mathrm{T}} S^{(slw)} T. \tag{25}$$

结合式 (24)、(25) 和不相关的 Fisher 判别向量 (18), 得到 SLFisher 方法的目标函数

$$\max T^{\mathrm{T}} S^{(slb)} T,\ \min T^{\mathrm{T}} S^{(slw)} T;$$
$$\text{s.t. } T^{\mathrm{T}} S^{(t)} T = I. \tag{26}$$

对式 (26) 采用 Lagrange 乘子优化, 得到对应特征分解问题

$$(S^{(slb)} - S^{(slw)}) T = \lambda S^{(t)} T, \tag{27}$$

进而得到 SLFisher 的转换矩阵 T, 由式 (27) 的 r 个最大特征值 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_r$ 所对应的特征向量 $t_1, t_2, \cdots, t_r$ 组成.

算法 1　基于概率类和不相关判别的 SLFisher 方法.

输入: 高维数据集 $X = \{(x_i, l_i) | l_i \in \{1, 2, \cdots, c\}\}_{i=1}^{n} \bigcup \{x_j | x_j \in \boldsymbol{R}^d\}_{j=n+1}^{n+m}$;

输出: 1) $d \times r$ 的转换矩阵 $T^{(\mathrm{SLFisher})}$, 2) 对应于高维数据集 X 的 r- 维数据集 Y.

Step 1:　由式 (12) 和 (23) 计算重构概率类 $P_{(n+m)\times C}$ 和相似度矩阵 $\widetilde{A}$.

Step 2: 由式 (21) 和 (22) 计算类间数据对权重矩阵和类内数据对权重矩阵.

Step 3: 由式 (19) 和 (20) 计算半监督局部类间散度矩阵和半监督局部类内散度矩阵.

Step 4: 由式 (27) 计算各最大特征值对应的特征向量, 从而得到转换矩阵 T.

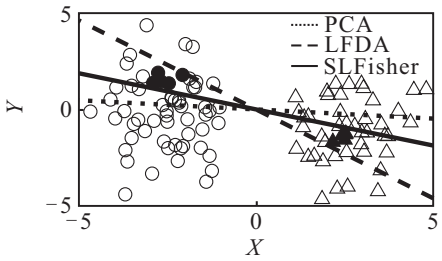
Step 5: 由式 $Y = T^{\mathrm{T}} X$ 计算高维数据集 X 所对应的 r- 维数据集 Y.

3 实验分析

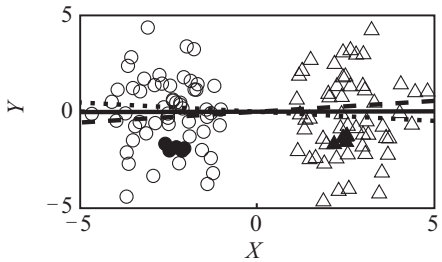
本节对SLFisher方法的维数约简问题进行性能评估, 参与比较的4个代表性维数约简方法分别为主成份分析(PCA)[21-22]、局部保持映射(LPP)[18]、半监督判别分析(SDA)[14]和半监督局部Fisher判别分析(SELF)[15]. 分别在人工数据集DS1和标准人脸数据集(Extended Yale-B和CMU PIE)上进行比较, 具体执行分为3个步骤: 1) 每个算法在训练集上学习得到一个投影矩阵; 2) 由每个投影矩阵获得测试集的低维数据表示; 3) 使用1-近邻分类器计算测试集低维数据的分类精确度. 本文SELF方法取得的维数约简结果与实验选取参数相关.

3.1 人工合成数据

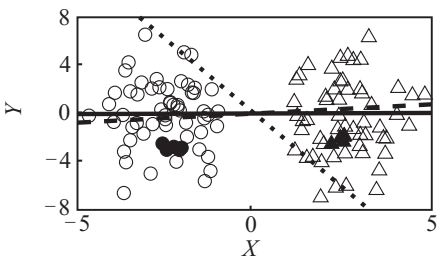
图1为数据集DS1的数据和一维约简结果的可视化, 其中圆和三角图形分别代表正例和反例样本, 实心和空心图形分别代表有标记和无标记数据, 图1(a)和图1(b)是同一数据集只是标记的数据不同, 图1(b)和图1(c)是同一数据集, 只是后者Y轴坐标度量放大了1.5倍. 另外, 无监督PCA、有监督LFDA和半监督SLFisher三种方法在数据集DS1上得到的一维投影向量分别对应于点虚线、线虚线和实线. 图2和图3分别对应图1各分图的重构概率类3D和2D的可视化.



(a) 部分标记的数据1

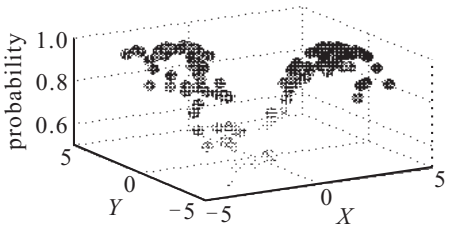


(b) 部分标记的数据2

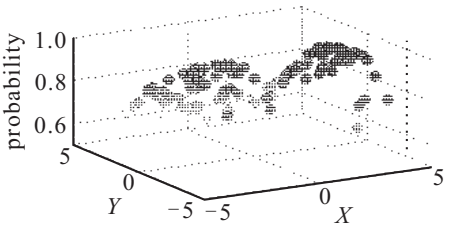


(c) 部分标记的数据3

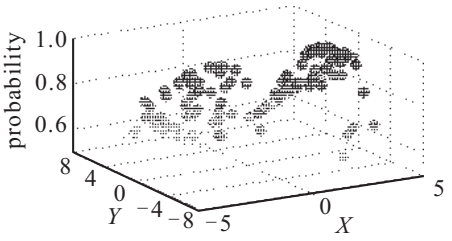
图 1 数据集DS1的数据和一维约简结果的可视化



(a) 重构概率类3D(数据1)

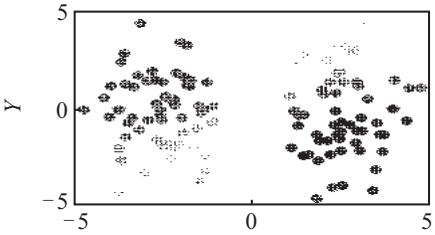


(b) 重构概率类3D(数据2)

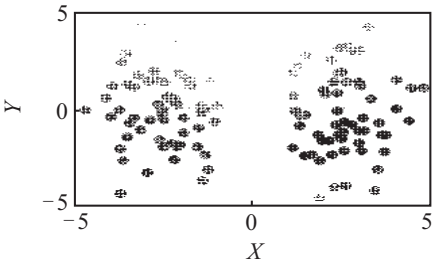


(c) 重构概率类3D(数据3)

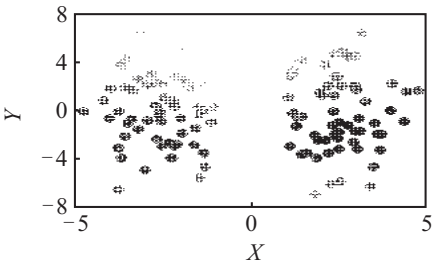
图 2 重构概率类3D的可视化



(a) 重构概率类2D(数据1)



(b) 重构概率类2D(数据2)



(c) 重构概率类2D(数据3)

图 3 重构概率类2D的可视化

由图 1(a) 和图 1(b) 可见, 无监督方法未受到选取不同有标记数据的影响, 有监督方法对于选择不同有标记数据出现较大波动, 即有监督方法可能出现过拟合于有标记数据. 图 1(b) 和图 1(c) 表明, 无监督方法与有监督方法相比, 前者更容易受到不同度量的影响, 出现满足全局数据的低维嵌入子空间而与真实低维嵌入子空间相差甚远的维数约简结果. 图 2(a) 和图 3(a) 存在不符合实际分类的 3 个紧靠数据块的数据点, 即重构概率类得到 3 个错误分类的数据. 因为这 3 个数据所属类别的概率值接近 0.5, 所以在下述的SLFisher方法中对维数约简影响较小.

3.2　人脸图像识别

对 Extended Yale-B 和 CMU PIE 两个人脸数据集进行维数约简实验. 为了平衡处理的复杂度和精确度, 对人脸图像作统一的预处理, 如表 1 所示. 人脸图像上的两个眼睛位置保持水平, 图像大小缩放到 32×32 像素, 因此每个图像都可以表示成图像空间中的 1 024 维列向量. 对各数据集随机选择 50 % 的数据作为训练数据, 其余数据作为测试数据, 并将训练数据的 25 % 作为有标记训练数据, 剩余训练数据作为无标记训练数据. 在对应测试数据和训练数据上依次执行 4 种方法, 并相互独立地执行 20 次. 最后计算 4 种方法在各数据集上的平均识别精确度和重复执行识别精度的标准方差.

表 1　人脸数据集信息

Database	Dimensionality	Samples	Class
Extended Yale-B	32×32	2 414	38
CMU PIE	32×32	11 554	68

Extended Yale-B 人脸数据集包括 38 个不同个体的 2 414 张人脸图像, 这些图像拍摄于不同的光线角度、表情和配戴眼镜的情况下. 图 4 给出了 4 种不同个体的 20 张示例照片; 表 2 列出 4 种方法在 Extended Yale-B 人脸数据集约简到 10- 维、20- 维和 30- 维的平均识别精确度和重复执行识别精度的标准方差, 并加粗了表中最大识别精度的字体.

表 2　Extended Yale-B 数据集的算法执行比较

Algorithm	$r = 10$	$r = 20$	$r = 30$
LPP	24.67 ± 1.07	31.71 ± 1.59	36.22 ± 1.62
SDA	64.56 ± 1.98	72.32 ± 1.29	75.19 ± 1.17
SELF	34.59 ± 1.81	42.72 ± 1.54	46.65 ± 1.42
SLFisher	$\mathbf{65.06 \pm 1.79}$	$\mathbf{73.12 \pm 1.29}$	$\mathbf{75.59 \pm 1.01}$

CMU PIE 人脸数据集包括 68 个不同个体的 41 368 张人脸图像, 这些图像拍摄于不同角度、光线和表情下, 实验选取 C27 的脸部正面图像集. 图 5 为 4 种不同个体的 20 张示例照片; 表 3 为 4 种方法在 CMU PIE 人脸数据集约简到 10- 维、20- 维和 30- 维的平均识别精确度和重复执行识别精度的标准方差, 并加粗了表中最大识别精度的字体.

图 4　Extended Yale-B 数据集中 4 个个体的部分图像

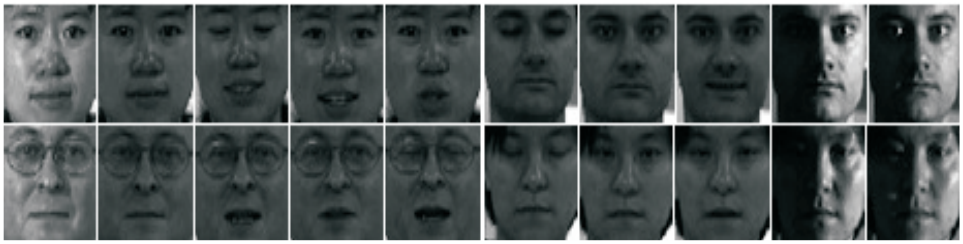


图 5　CMU PIE 数据集中 4 个个体的部分图像

表 3　CMU PIE 数据集的算法执行比较

Algorithm	$r = 10$	$r = 20$	$r = 30$
LPP	40.34 ± 1.52	47.90 ± 0.98	52.47 ± 1.00
SDA	$\mathbf{77.16 \pm 1.42}$	85.76 ± 1.12	88.03 ± 1.06
SELF	61.61 ± 2.11	72.92 ± 1.28	79.46 ± 0.99
SLFisher	76.75 ± 1.81	$\mathbf{86.18 \pm 1.25}$	$\mathbf{89.00 \pm 1.03}$

由表 2 和表 3 可见, 半监督的方法在很大程度上优于无监督的 LPP 方法, SLFisher 方法在两个人脸数据库上的维数约简结果几乎完全优于参与比较的其他方法. 另外, LPP、SDA、SELF 和 SLFisher 四种方法维数约简到 10- 维、20- 维和 30- 维的平均识别精度不

断增大, 并且10- 维到20- 维的精度增大幅度超过20- 维到30 维的精度增大幅度. 这表明特征维数的增加对于数据分析存在两方面的影响:

1) 更全面地描述数据, 为数据分析的决策和判断提供充分的依据, 即识别精度随着约简维数的增加呈现递增趋势;

2) 更多的冗余或干扰特征包含在数据中, 影响数据分析结果的性能和效率, 导致识别精度的增幅呈放缓趋势, 甚至出现精度下降的现象.

4 结 论

本文在经典维数约简Fisher判别分析的基础上提出了一种基于概率类和不相关判别的半监督局部Fisher (SLFisher) 方法. 通过重构概率类使得数据获得一个所属类别的概率, 结合不相关判别向量优化Fisher准则函数, 实现半监督高维映射到低维的数据特征之间是不相关的, 并且类间数据对尽可能地分离和类内邻近数据对尽可能地紧凑. 最后, 通过实验分析表明了SLFisher方法在标准人脸数据集优于参与比较的其他算法. SLFisher方法的不足之处是, 容易受到数据对相似度度量的影响, 要求同类近邻数据对的相似度尽可能不小于非同类数据对的相似度.

参考文献(References)

[1] Yan S, Xu D, Zhang B, et al. Graph embedding and extensions: A general framework for dimensionality reduction[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 29(1): 40-51.

[2] Tenenbaum J, Silva V, Langford J. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000, 290(5500): 2319-2322.

[3] Roweis S, Saul L. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500): 2323-2326.

[4] Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural Computation, 2003, 15(6): 1373-1396.

[5] Zhang L, Qiao L, Chen S. Graph-optimized locality preserving projections[J]. Pattern Recognition, 2010, 43(6): 1993-2002.

[6] Li X, Tao J, Zhang K. Efficient and robust feature extraction by maximum margin criterion[J]. Neural Networks, 2006, 17(1): 157-165.

[7] Sugiyama M. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis[J]. The J of Machine Learning Research, 2007, 8: 1027-1061.

[8] Li W, Fowler J, Bruce L. Locality-preserving dimensionality reduction and classification for hyperspectral image analysis[J]. IEEE Trans on Geoscience and Remote Sensing, 2012, 50(4): 1185-1198.

[9] Song Y, Nie F, Zhang C, et al. A unified framework for semi-supervised dimensionality reduction[J]. Pattern Recognition, 2009, 41(1): 2789-2799.

[10] Zhang Z, Zhao M, Tommy W. Marginal semi-supervised sub-manifold projections with informative constraints for dimensionality reduction and recognition[J]. Neural Network, 2012, 36: 97-111.

[11] Song Y, Nie F, Zhang C. Semi-supervised sub-manifold discriminant analysis[J]. Pattern Recognition, 2008, 29(13): 1806-1813.

[12] Huang H, Li J, Liu J. Enhanced semi-supervised local fisher discriminant analysis for face recognition[J]. Future Generation Computer Systems, 2008, 28(1): 244-253.

[13] Liu Z, Wang J, Man J, et al. Self-adaptive local fisher discriminant analysis for semi-supervised image recognition[J]. Int J of Biometrics, 2012, 4(4): 338-356.

[14] Cai D, He X, Han J. Semi-supervised discriminant analysis[C]. IEEE 11th Int Conf on Computer Vision. Rio de Janeiro: Brazil, 2007: 1-7.

[15] Sugiyama M, Ide T, Nakajima S, et al. Semi-supervised local fisher discriminant analysis for dimensionality reduction[J]. Machine Learning, 2010, 78(1/2): 35-61.

[16] Fisher R. The use of multiple measurements in taxonomic problems[J]. Annals of Eugenics, 1936, 7(2): 179-188.

[17] Huang H, Feng H, Peng C. Complete local fisher discriminant analysis with laplacian score ranking for face recognition[J]. Neurocomputing, 2012, 89: 64-77.

[18] He X, Niyogi P. Locality preserving projections[C]. Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2004: 153-160.

[19] Zhong J, Yang J, Hu Z, et al. Face recognition based on the uncorrelated discriminant transformation[J]. Pattern Recognition, 2001, 34(7): 1405-1416.

[20] Zelnik L, Perona P. Self-tuning spectral clustering[C]. Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2005: 1601-1608.

[21] Jolliffe I. Principal component analysis[M]. New York: Springer, 2002: 111-147.

[22] Abdi H, Williams L. Principal component analysis[J]. Wiley Interdisciplinary Reviews: Computational Statistics, 2010, 4(2): 433-459.

(责任编辑: 郑晓蕾)