

一种基于数值逼近的 KPCA 改进算法

赵英男^{1,2} 王水平^{1,2} 郑玉^{1,2}

摘要

核方法广泛应用于模式识别等领域,但其存在着特征抽取效率和样本集的大小成反比的瓶颈问题. 因此提出一种基于数值逼近的方法确定虚拟样本矢量,以此代替训练样本,提高 KPCA(Kernel Principle Component Analysis)特征抽取效率. 在确定虚拟样本矢量时,只需将样本矢量的初值设定为随机变量,算法实现简单、高效. 在基准数据集上的实验结果显示该算法优于同类算法.

关键词

数值逼近;主成分分析;核主成分分析;特征提取

中图分类号 TP391

文献标志码 A

0 引言

核方法广泛应用于模式识别、机器学习等领域. 该类方法在理论上能够将原始数据映射到一个新空间,这个新空间可以增加数据的线性可分性,而且这种映射不需要显式实现,因此核方法可以处理一些较为复杂的非线性模式识别问题,如人脸识别等^[1-3]. 然而,核方法的特征抽取效率和训练样本集的大小成反比,这是阻碍核方法应用的一个瓶颈问题^[4-6]. 在实际应用中,常常需要较大的训练样本集来保证训练模型的推广能力^[7-9],在这种情形下,核方法是低效甚至是不实用的. 为了提高核方法的特征抽取效率,目前出现了一系列的改进算法,大致可分为数据集减少和数值逼近 2 种方法.

对于第 1 类方法,主要依据一定的原则,从中选择部分重要的训练样本矢量代替原训练样本集,从而提高核方法的特征抽取效率. 如 Xu 等^[10-11]主要依据主成分分析(PCA)的物理模型,进行显著节点的选取. Tipping^[12]利用 PCA 的性质提出了一种稀疏 KPCA(Sparse KPCA)方法. 拓扑削减方案(Prototype Reduction Schemes)^[13]和建立在均方误差最小化原则上的算法^[14]也可认为是该类方法. 第 2 类方法中比较典型的如 Scholkopf 等^[15]提出的数值逼近方法. 该方法基于特征空间的几何特征,根据特征空间和输入空间的联系,对输入空间进行近似,从而降低计算量,但是该算法复杂度较高.

上述的改进算法和原始 KPCA 方法相比,虽然可以提高效率,但是它们的训练过程计算代价很高,这个性质依然会削弱这些改进算法的实际应用. 本文认为在特征空间中的鉴别矢量能够被一组构造的虚拟样本矢量近似表示,据此应用一种简单、高效的数值逼近迭代算法逐个进行虚拟样本的确定,并以此来代替原训练样本. 该算法不仅可以提高 KPCA 的特征抽取效率,而且在训练阶段的花费时间也比同类算法要少.

1 核方法的重建

在表示特征空间的鉴别矢量时,一些虚拟样本的线性组合可以很好地近似表示特征矢量. 虚拟样本越少,特征抽取过程越高效.

在特征空间,2 个样本 $\varphi(\mathbf{x}_i)$, $\varphi(\mathbf{x}_j)$ 之间的距离的平方可以表示为

$$\|\varphi(\mathbf{x}_i) - \varphi(\mathbf{x}_j)\|^2 = (\varphi(\mathbf{x}_i) - \varphi(\mathbf{x}_j))^T (\varphi(\mathbf{x}_i) - \varphi(\mathbf{x}_j)). \quad (1)$$

收稿日期 2011-07-20

资助项目 国家自然科学基金(60702076);江苏高校优势学科建设工程资助项目

作者简介

赵英男,女,博士,副教授,主要研究方向为人脸识别、特征抽取以及图像处理等.

yingnanzhao@163.com

1 南京信息工程大学 江苏省网络监控中心, 南京, 210044

2 南京信息工程大学 计算机与软件学院, 南京, 210044

应用核函数的定义,可以将式(1)转换为

$$\|\varphi(\mathbf{x}_i) - \varphi(\mathbf{x}_j)\|^2 = k(\mathbf{x}_i, \mathbf{x}_i) + k(\mathbf{x}_j, \mathbf{x}_j) - 2k(\mathbf{x}_i, \mathbf{x}_j). \quad (2)$$

特征空间中的鉴别矢量和第1个节点之间的距离为

$$d_1^0 = \|\varphi(\mathbf{x}_1^0) - \sum_{i=1}^N \alpha_i \varphi(\mathbf{x}_i)\|^2. \quad (3)$$

其中 $\varphi(\mathbf{x}_1^0)$ 和 $\sum_{i=1}^N \alpha_i \varphi(\mathbf{x}_i)$ 分别表示第1个节点和特征空间中的鉴别矢量.显然, d_1^0 越小, $\varphi(\mathbf{x}_1^0)$ 越接近特征空间中的鉴别矢量.根据核函数的定义,有

$$\begin{aligned} \|\varphi(\mathbf{x}_1^0) - \sum_{i=1}^N \alpha_i \varphi(\mathbf{x}_i)\|^2 = & k(\mathbf{x}_1^0, \mathbf{x}_1^0) - 2 \sum_{i=1}^N \alpha_i \varphi(\mathbf{x}_1^0, \mathbf{x}_i) + \\ & \sum_{i=1}^N \alpha_i^2 \varphi(\mathbf{x}_i, \mathbf{x}_i) + 2 \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j). \end{aligned} \quad (4)$$

显然,假定最优的 \mathbf{x}_1^0 应使 d_1^0 具有最小值是合理的.因此可以利用梯度下降法迭代计算 \mathbf{x}_1^0 .迭代计算 \mathbf{x}_1^0 的公式如下:

$$\mathbf{x}_1^0(t+1) = \mathbf{x}_1^0(t) - \mu_1 \frac{\partial d_1^0}{\partial \mathbf{x}_1^0} \Big|_{\mathbf{x}_1^0 = \mathbf{x}_1^0(t)}. \quad (5)$$

其中 μ_1 是学习比率.易得如下公式:

$$\frac{\partial d_1^0}{\partial \mathbf{x}_1^0} = \frac{\partial k(\mathbf{x}_1^0, \mathbf{x}_1^0)}{\partial \mathbf{x}_1^0} - 2 \sum_{i=1}^N \alpha_i \frac{\partial k(\mathbf{x}_1^0, \mathbf{x}_i)}{\partial \mathbf{x}_1^0}. \quad (6)$$

一旦对 \mathbf{x}_1^0 设定了一个初值,就可以利用式(5)和(6)迭代计算 \mathbf{x}_1^0 的最优值.

当确定第2个虚拟样本矢量时,需要它和第1个虚拟样本尽量不相似.用内积 $g_0 = \varphi(\mathbf{x}_1^0) \cdot \varphi(\mathbf{x}_2^0) = k(\mathbf{x}_1^0, \mathbf{x}_2^0)$ 来衡量差异性,其中 \mathbf{x}_2^0 代表第2个虚拟样本矢量.认为最优的 \mathbf{x}_2^0 应该具有最小的内积 $k(\mathbf{x}_1^0, \mathbf{x}_2^0)$.类似地,假定前 $r-1$ 个虚拟已经确定,那么第 r 个虚拟矢量 \mathbf{x}_r^0 应该和前 $r-1$ 个虚拟样本矢量尽可能不相似.因此,让 $g = \sum_{i=1}^{r-1} \varphi(\mathbf{x}_i^0) \cdot \varphi(\mathbf{x}_r^0) = \sum_{i=1}^{r-1} k(\mathbf{x}_i^0, \mathbf{x}_r^0)$ 尽可能地小.

利用梯度下降的思想,可以设计如下公式进行 \mathbf{x}_r^0 的迭代计算:

$$\mathbf{x}_r^0(t+1) = \mathbf{x}_r^0(t) - \mu_2 \frac{\partial g}{\partial \mathbf{x}_r^0} \Big|_{\mathbf{x}_r^0 = \mathbf{x}_r^0(t)}, \quad (7)$$

$$\frac{\partial g}{\partial \mathbf{x}_r^0} = \sum_{i=1}^{r-1} \frac{\partial k(\mathbf{x}_i^0, \mathbf{x}_r^0)}{\partial \mathbf{x}_r^0}. \quad (8)$$

其中 μ_2 是学习比率.

2 算法概述

提出下列算法来确定虚拟样本矢量.

第1步.确定第1个虚拟样本矢量.

将一个随机矢量赋给第1个虚拟样本矢量,每一个元素的取值范围从0到1.然后用式(5)和(6)计算 \mathbf{x}_1^0 .

第2步.确定第2个虚拟样本矢量.

设第2个虚拟样本的初值为一个随机样本矢量.然后利用式(7)和 $\frac{\partial g}{\partial \mathbf{x}_2^0} = \frac{\partial k(\mathbf{x}_1^0, \mathbf{x}_2^0)}{\partial \mathbf{x}_2^0}$ 迭代计算第2个虚拟样本矢量.

第 r 步.确定第 r 个虚拟样本矢量.

假设前 $r-1$ 个虚拟样本矢量 $\mathbf{x}_1^0, \mathbf{x}_2^0, \dots, \mathbf{x}_{r-1}^0$ 经过步骤1,2, $\dots, r-1$ 已经被确定,对第 r 个虚拟样本矢量设定一个随机变量初值.然后利用式(7)和(8)来迭代计算 \mathbf{x}_r^0 .

需要指出的是:也可以用距离衡量2个矢量 $\varphi(\mathbf{x}_r^0)$ 、 $\varphi(\mathbf{x}_i^0)$ 之间的差异性.即

$$g_0 = \varphi(\mathbf{x}_1^0) \cdot \varphi(\mathbf{x}_2^0) = k(\mathbf{x}_1^0, \mathbf{x}_2^0) \text{ 和}$$

$$g = \sum_{i=1}^{r-1} \varphi(\mathbf{x}_i^0) \cdot \varphi(\mathbf{x}_r^0) = \sum_{i=1}^{r-1} k(\mathbf{x}_i^0, \mathbf{x}_r^0) \text{ 可以用}$$

$$g_0 = \|\varphi(\mathbf{x}_1^0) - \varphi(\mathbf{x}_2^0)\|^2 = k(\mathbf{x}_1^0, \mathbf{x}_1^0) + k(\mathbf{x}_2^0, \mathbf{x}_2^0) - 2k(\mathbf{x}_1^0, \mathbf{x}_2^0) \text{ 和}$$

$$g = \sum_{i=1}^{r-1} \|\varphi(\mathbf{x}_i^0) - \varphi(\mathbf{x}_r^0)\|^2 = \sum_{i=1}^{r-1} (k(\mathbf{x}_i^0, \mathbf{x}_i^0) +$$

$$k(\mathbf{x}_i^0, \mathbf{x}_r^0) - 2k(\mathbf{x}_i^0, \mathbf{x}_r^0))$$

分别代替.如果采纳的核函数是高斯函数

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\eta}\right),$$

那么式(4)等价于

$$\frac{\partial d_1^0}{\partial \mathbf{x}_1^0} = \frac{4}{\eta} \sum_{i=1}^N \alpha_i k(\mathbf{x}_1^0, \mathbf{x}_i) (\mathbf{x}_1^0 - \mathbf{x}_i). \quad (9)$$

而且,式(7)等价于

$$\frac{\partial g}{\partial \mathbf{x}_r^0} = -\frac{2}{\eta} \sum_{i=1}^{r-1} k(\mathbf{x}_i^0, \mathbf{x}_r^0) (\mathbf{x}_i^0 - \mathbf{x}_r^0). \quad (10)$$

上述迭代过程的终止条件是虚拟样本的数量已经达到预定数目或者是满足了其他条件.

3 实验与数据分析

在2个基准数据集^[11]上测试了本文的算法.每个数据集包含100个子集.每个子集包含一个训练集和一个测试集.采用的高斯核函数 $k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\eta}\right)$.参数 η 设为第1个训练集的

协相关矩阵的 Frobenius 模的平方. 抽取特征的维数和构造的虚拟样本的维数相同. 应用最近邻分类器进行测试样本的分类. 当对每一个数据集进行实验的时候, 第 1 个训练子集来确定虚拟样本, 并重新计算 KPCA 并测试所有的测试子集. 由于虚拟样本具有随机性质, 每一个数据集上的实验运行 10 遍. 对每一个测试子集, 10 次识别错误率的均值为最后的平均识别错误率. 由于每一个测试子集有一个平均识别错误率, 故给出所有测试子集的平均识别错误率, 如表 1 和 2 所示. 这里, 为了观察所提算法的性能, 本文给出了不同虚拟样本数量带来的识别错误率的变化. 可以看出, 本文的方法在识别率方面要高于文献[13]所提算法.

表 1 基于 CANCER 数据集的 KPCA 平均识别错误率

Table 1 Average classification error rate of KPCA on CANCER dataset

构造的虚拟样本数	平均识别错误率	
	本文方法	文献[13]的 IKPCA 方法
6	0.095 4	0.103 4
7	0.100 1	0.124 2
8	0.093 0	0.108 3
9	0.090 8	0.118 3
10	0.088 7	0.121 4
15	0.091 1	0.101 8
20	0.091 3	0.104 2
25	0.089 1	0.093 1
30	0.090 1	0.100 9
35	0.089 7	0.102 6
40	0.086 0	0.093 5

表 2 基于 HEART 数据集的 KPCA 平均识别错误率

Table 2 Average classification error rate of KPCA on HEART dataset

构造的虚拟样本数	平均识别错误率	
	本文方法	文献[13]的 IKPCA 方法
3	0.115 1	0.119 0
4	0.115 5	0.119 1
5	0.102 0	0.109 9
6	0.098 0	0.109 3
7	0.098 1	0.108 1
8	0.099 2	0.106 3
9	0.104 2	0.110 7
10	0.101 2	0.105 1
15	0.091 8	0.095 6
20	0.094 3	0.100 9

4 结语

目前核方法在模式识别、机器学习等领域应用较为广泛, 但是核方法在应用中存在一个瓶颈问题, 即特征抽取效率和训练样本集的大小成反比, 因为在实际应用中需要较大的训练样本集来保证训练模型的推广能力, 在这种情形下, 核方法是低效甚至是不实用的. 不同于以前方法从核模式上进行特征抽取过程的计算量削减, 本文从一个崭新的角度来解决类似问题. 提出的算法具有下列优点: 首先, 利用虚拟样本的构建来近似特征空间中的真正的鉴别矢量, 不仅符合核方法近似的本质, 而且具有一个封闭解; 其次, 所提算法的训练过程具有很低的计算复杂度; 最后, 相比同类算法, 当虚拟样本矢量较少时, 本文方法能够产生一个较高的识别率.

本文的方法也显示了提出的准则, 即虚拟样本矢量差异度应尽可能较大的准则是合理的. 当我们看待一个近似真正鉴别矢量问题为一个问题, 即寻找一个特征矢量集来正确替代整个训练集合的作用, 容易看出它的合理性. 显然, 如果矢量集中的元素非常不相似, 那么这个集合在矢量空间可以提供更多的信息. 事实上, 当构造的虚拟样本空间相互之间差异性很大的话, 这个虚拟样本矢量集合可以替代原始训练样本集合.

参考文献

References

- [1] Muller K R, Mika S, Rätsch G, et al. An introduction to kernel-based learning algorithms [J]. IEEE Transactions on Neural Network, 2001, 12(2): 181-201
- [2] Lin Y Yu, Liu T L, Fuh C S. Multiple kernel learning for dimensionality reduction [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(6): 1147-1160
- [3] Baudat G, Anouar F. Generalized discriminant analysis using a kernel approach [J]. Neural Computation, 2000, 12(10): 2385-2404
- [4] You D, Hamsici O C, Martinez A M. Kernel optimization in discriminant analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(3): 631-638
- [5] Xu Y, Yang J Y, Lu J F, et al. An efficient renovation on kernel Fisher discriminant analysis and face recognition experiments [J]. Pattern Recognition, 2004, 37(10): 2091-2094
- [6] Chu W S, Chen J C, Lien J J J. Kernel discriminant transformation for image set-based face recognition [J]. Pattern Recognition, 2011, 44(8): 1567-1580
- [7] 杜卓明, 屠宏, 耿国华. KPCA 方法过程研究与应用

- [J]. 计算机工程与应用, 2010, 46(7): 8-10
DU Zhuoming, TU Hong, GENG Guohua. KPCA method research and application process[J]. Computer Engineering and Applications, 2010, 46(7): 8-10
- [8] Zheng W M, Zou C R, Zhao L. An improved algorithm for kernel principal component analysis[J]. Neural Processing Letters, 2005, 22(1): 49-56
- [9] Schölkopf B, Smola A, Mllerr K-R. Nonlinear component analysis as a kernel eigenvalue problem [J]. Neural Computation, 1998, 10(5): 1299-1319
- [10] Xu Y, Zhang D, Song F, et al. A method for speeding up feature extraction based on KPCA[J]. Neurocomputing, 2007, 70(4/5/6): 1056-1061
- [11] Xu Y, Zhang D, Jin Z, et al. A fast kernel-based nonlinear discriminant analysis for multi-class classification [J]. Pattern Recognition, 2006, 39(6): 1026-1033
- [12] Tipping M E. Sparse kernel principal component analysis [M] // Leen T K, Dietterich T G, Tresp V. NIPS 2000: Neural Information Processing Systems, MIT Press, 2000
- [13] Kim S W, Oommen B J. On using prototype reduction schemes to optimize kernel-based fisher discriminant analysis[J]. IEEE Transactions on Systems, Man, and Cybernetics. Part B: Cybernetics, 2008, 38(2): 564-570
- [14] Xu Y. A new kernel MSE algorithm for constructing efficient classification procedure[J]. International Journal of Innovative Computing, Information and Control, 2009, 5(8): 2439-2447
- [15] Schölkopf B, Mika S, Burges C J C, et al. Input space versus feature space in kernel-based methods [J]. IEEE Transactions on Neural Networks, 1999, 10(5): 1000-1017

Improved KPCA algorithm based on numerical approximation

ZHAO Yingnan^{1,2} WANG Shuiping^{1,2} ZHENG Yu^{1,2}

1 Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science & Technology, Nanjing 210044

2 School of Computer & Software, Nanjing University of Information Science & Technology, Nanjing 210044

Abstract Though kernel methods have been widely used for pattern recognition, they suffer from the problem that the extraction efficiency is in inverse proportion to the size of the training sample set. To solve it, we propose a novel improvement to Kernel Principle Component Analysis (KPCA) based on numerical approximation. The method is on the base of the assumption that the discriminant vector in the feature space can be approximately expressed by a certain linear combination of some constructed virtual sample vectors. We determine these virtual sample vectors one by one by using a very simple and computationally efficient iterative algorithm. When they are dissimilar to each other, this set is able to well replace the role of the whole training sample set in expressing the discriminant vector in the feature space. It is remarkable that the determined virtual sample vectors lead to a good improvement to KPCA, which allows an efficient feature extraction procedure to be obtained. Also, we need only to set the initial values of the virtual sample vectors to random values. The experiments on two benchmark datasets show that our method can achieve the goal of efficient feature extraction as well as good and stable classification accuracy.

Key words numerical approximation; principal component analysis (PCA); Kernel PCA (KPCA); feature extraction