

智能识别中的降维算法简述

皋 军

(盐城工学院 信息工程学院, 江苏 盐城 224051)

摘要:近年来降维方法作为智能识别中关键的数据预处理技术得到了较为成功地运用。在总结比较经典的降维方法的基础上,分别从特征选择和特征提取两个方面来阐述各种方法所具有的特点和优势,并在一定程度上指出相应方法在目前所存在的问题和挑战。

关键词:特征降维;特征选择;特征提取;智能分析

中图分类号:TP18

文献标识码:A

文章编号:1671-5322(2010)03-0014-07

随着信息化的发展,在具体的智能识别过程中需要处理的数据越来越多地呈现出高维特征,造成这种现象的主要原因在于:在智能识别过程中只有当样本已经包含了足够多的模式分类信息,才能得到较好的智能识别效果。然而,如何确定特征中是否已经包含足够多的类别信息本身就是个很难解决的问题。为了得到较好的智能识别效果,在通常情况下人们通过采集尽可能多的特征去体现样本的类别信息,这导致原始样本空间的维数可能达到几千维甚至上万维。而如果在如此高维的原始空间直接使用模式分类方法,那么所得到的智能识别效果将受到较大的影响。同时,由于随着样本特征维数的增加将使得对样本的统计特性更加难以估计,从而会影响分类算法的泛化能力,呈现出所谓过学习的现象。因此,在智能识别过程中必须要对高维数据进行相应的预处理,以达到在保持样本信息量的基础上尽可能降低特征的维数,提高模式分类的效果,而特征降维就是一种较为有效的数据预处理方法。

特征降维作为智能识别系统中非常关键的数据预处理方法,近年来在许多领域得到有效的运用。比如在数据可视化领域中,由于人类不能直接处理和感知真实世界的高维数据,而对低维数据的处理能力目前还强于机器智能。因此可以使用降维技术得到高维数据在低维空间上(二维或

三维)的可视化投影,从而可以较为容易地发现隐藏在高维数据中的类别信息和空间分布结构;在图像识别领域,当要求处理的数据以矢量形式出现时,该数据就是一高维数据,而通常在处理如此高维数据时为了在一定程度上防止过拟合现象的出现,一般要求具有足够大的训练集,显然这是不可行的,因此必须对高维的数据进行有效降维,虽然在一定程度上会丢失部分信息量,然而相对于智能识别的效率来说是合适的;在微阵列数据基因选择方面,DNA微阵列数据具有明显的超高维超小样本的特点,正是由于这个特点造成了严重的维数灾难现象,因此为了有效的发现特定癌症的基因模式、找出最利于分类的基因个体,必须对此类型的基因数据进行特征降维处理。

特征降维方法在过去的几十年中被广泛地加以研究,但总体上可以已有的方法分为两大类,即特征选择和特征提取。所谓特征选择就是在原始的特征集中选取最有代表性的特征子集,而特征提取则是对原始特征空间采用某种具体的映射操作以获取低位的投影空间。因此,特征选择和特征降维不管从算法依据的原理上还是算法的具体构造上都存在着一定的差异,这就需要分别对上述两种特征降维分别加以讨论。

1 特征选择方法

特征选择作为一种特征降维的方法,其主要

收稿日期:2010-07-08

基金项目:盐城工学院应用基础研究资助项目(XKY2009070)

作者简介:皋军(1971-),男,江苏阜宁县人,副教授,江南大学博士研究生,主要研究方向为图像处理、生物数据识别等。

的目的在于如何在高维的原始样本空间选择一组最有代表性的特征来组成一个新的低维样本空间。为此需要解决两个问题,一是选择最优特征子集的评判标准,二是具体的特征选择方法。从目前经典的特征选择方法来看,大体可以分为基于特征加权^[1-4]、基于概率密度估计^[5-11]、基于信息论^[12-18]和基于支持向量机^[11-13],诚然在上述方法中可能还结合其他技术,比如结合 LDA 的关键技术^[14]等等。下面就以上几类来分别加以讨论。

1.1 基于特征加权的特征选择方法

特征加权的方法出发点比较明确,就是通过对每一特征赋予相应的权值来表征不同特征对模式分类的贡献大小^[15]。文献[1]提出的加权 K-均值类型聚类(Weighting in K-Means Type Clustering, W-K-Means),该方法通过无监督的模式分类方法(聚类)来得到每个特征所对应的权值,并对相应的权值进行排序,并使用聚类的有效性来作为特征选择的标准。然而,该方法在处理大型稀疏数据时在一定程度上存在某些特征权值不可计算问题,同时表现出在处理含有噪音数据时特征选择效果鲁棒性较差。文献[2]提出的 RELIEF 特征选择方法,该方法根据识别相邻模式的区分能力来迭代产生相应特征的特征权值,算法简单有效。而文献[3]在充分分析了原始 RELIEF 理论上和算法构造上的不足后,依据最大期望原理重新构造迭代目标函数,提出了新的迭代 RELIEF 算法:I-RELIEF,该方法在一定程度上继承了原 RELEF 算法的优点,同时可以实现多类模式分类的特征选择,提高了算法的适应性。然而,上述两种方法都在不同程度上存在着抗噪性不强、对无关特征敏感的缺陷。文献[4]在分析已有的基于特征加权的特征选择方法的基础上,通过对样本和特征分别加权的方式来重新构造目标函数,并较系统的讨论了特征权指数对收敛性的影响,在一定程度上提高了算法抗噪能力,增强了特征选择的稳定性。通过对以上几种具有代表性的特征加权的特征选择方法的分析,可以看出该类方法在两个方面面临着挑战:一是如何引入新的机制较大幅度的提高算法的抗噪能力,以提高特征选择的稳定性;二是目前决大部分特征加权的特征选择方法都要使用迭代优化的方法,在迭代过程中虽然在一定程度上能保证权值的收敛性,但是否可以收敛到全局最优解或局部

最优解目前在理论上并没有得到很好的解决。

1.2 基于概率密度估计的特征选择方法

近年来,研究者成功的将各种概率密度估计的方法运用到特征选择领域,提出了许多基于概率密度估计的特征选择方法。这些方法从使用概率密度估计的方法来讲可以分为两大类:参数估计和非参数估计。在参数估计的特征选择方法中,比如文献[5]提出一种基于 Bayesian 估计的特征选择方法,该方法基于分层的 Bayesian 线性模型,并结合马尔可夫链蒙特卡罗方法来估计每一特征在模型出现的概率,通过概率的大小来选择特征,文献[6]使用均值和方差构造的统计量来作为特征选择的测度,文献[7,8]直接使用统计量来进行特征选择,以上这些方法都能在模式分类中较好地处理特征降维问题。然而作为基于参数估计的方法,总是要求已知总体分布的形式,而在很多实际问题并不知道总体分布形式,或总体分布不是一些通常遇到的典型分布,即不能写成某些参数的函数,从而使得该类方法在一定程度上表现出适应性不强的缺陷,同时特征选择效果的鲁棒性较差。而相对参数估计特征选择方法,基于非参数估计的特征选择方法可以处理任意概率分布的特征选择问题,而不必要求概率密度的参数形式已知。在目前基于非参数估计的特征选择方法中,Parzen 窗口估计被广泛的使用^[9-11],该类方法在处理特征选择问题时表现出算法适应性较强,即对规则分布还是不规则分布,单峰还是多峰分布都可以使用该方法得到密度估计,而且只要样本足够多,总可以保证收敛于任何复杂的未知密度。然而,Parzen 窗口估计的特征选择方法由于在密度估计过程中需要大量的训练样本,因此具有较大的时间与空间复杂度,同时由于在使用 Parzen 窗口估计来处理实际的特征选择问题时,特征选择的效果往往还受到窗口函数、窗口宽度的影响,因此如何根据实际问题选择相应的窗口函数和窗口宽度也是基于 Parzen 窗口估计的特征选择方法面临的一个挑战。

1.3 基于信息论的特征选择方法

基于信息论的特征选择方法主要依据从每个特征中获取信息量的大小来进行特征选择,目前使用最广泛的信息度量可以分为熵度量和互信息度量。在特征选择方法中使用熵度量的目的在于表明特征空间中特征变量的随机性,反映每个特征所包含的信息量。文献[12]通过将特征权值

对应的熵引入目标函数,这样做的目的在于最小化类内散度同时最大化负的权值熵,从而实现特征选择。文献[13]提出的 FCBF 方法则是通过特定的信息增益补偿的机制来克服信息增益计算会偏向于具有更多取值的特征的弱点,并通过定义特征之间显著的相关性度量,来实现特征选择。而在特征选择方法中使用互信息的基本方法,一般来说只是来表明特征之间的相关性,互信息越大说明特征之间的依赖性越强,否则说明特征越独立。文献[14]提出的特征选择方法则依据最大互信息的原则,该方法通过特征均值来选择具有一定独立性的特征,而没有充分考虑这些被选特征是否含有分类信息,同时也没有充分考虑特征之间相对分类来说是否具有弱相关性。文献[15]提出的 CMIM 特征选择方法则依据最大条件信息增益原则,该方法一定程度上可以保证被选特征的独立性,同时可以保证被选特征之间的弱相关性。CMIM 方法在选择特征时,不但考虑特征具有的独立性,同时还要考虑该特征是否包含有利于分类的分类信息,否则即使该特征独立性再强也不会被选择,这一点区别于文献[14]的方法,而在判别特征相关性方面则和 FCBF 较为相似,但 CMIM 只通过唯一的耗散函数来判别被选特征是否含有有助于分类的信息以及与其他被选特征之间的相关性大小,同时在判别过程中并不需要调节参数,这一点显然区别于 FCBF 方法。文献[16]提出的 MIFS 方法则在神经网络中引入互信息来实现特征选择,该方法只考虑了单个特征相对于目标类之间的依赖性而没有充分考虑特征之间相关性评测,从而在一定程度上没有较好地解决特征冗余的问题。文献[17-19]提出的 MIFS-U、MIFSBS 依然是使用互信息来对特征进行选择,与 MIFS 方法不同的是 MIFS-U、MIFSBS 方法研究特征空间与目标类之间的相关性,从而大大降低了冗余特征与无关特征,提高了特征选择的效果。从以上分析可知,基于信息论的特征选择方法更侧重于被选特征的独立性以及特征之间的弱相关性,然而,由于特征选择过程中需要首先知道特征变量、目标变量所对应的先验概率密度函数,以及特征与目标之间的联合概率密度函数,而如何通过一定的估计方法得到以上的概率密度函数使该类方法在处理实际特征选择方法所面临的挑战,特别是当特征变量或目标变量呈现出连续性特点时,在使用具体的概率密度估计方

法之前,则首先要寻找合适的离散化方法来处理这些连续性变量,因此,这一点显得更为困难。

1.4 基于支持向量机的特征选择方法

基于支持向量机特征选择方法一般依赖结构风险最小化原理,具有较强的泛化能力。因此,在特征选择问题上较之于基于经验风险最小的众多方法具有更好的鲁棒性。文献[20]提出的支持向量机的回归特征消除法(the SVM recursive feature elimination, SVM-RFE),该方法首次使用留一交叉验证(leave-one-out method)误差率作为信息准则(information criterion)来实现特征选择(选择误差最小的特征子集),然而该方法在在评价不同特征子集的误差率消耗了大量的额外计算(特别是处理高维数据集时),因此 SVM-RFE 方法具有较高的时间复杂度(复杂度为代表特征的维数)。为了在一定程度上降低 SVM-RFE 算法的时间复杂度,文献[21]则通过合理的简化 SVM-RFE 算法使用的误差信息准则,从而使得时间复杂度为线性。文献[22]则根据 SVM 方法用于放松边界条件的松弛因子来构造新的特征选择判别信息准则:支持向量机信息准则(support vector machine information criterion, SVMIC),该信息准则由于依据的是 SVM 本身的信息,因而在一定程度上并不需要消耗大量的额外计算,从而使得特征选择的时间复杂度转化为只和测试样本数成线性关系,从这一点上看,该方法在处理小样本高维数据具有明显的优势。然而,通过分析发现以上几种方法在进行特征选择时还是通过选择支持向量来具体实现特征选择。文献[23]提出的势支撑向量机(potential support vector machines, P-SVMs)则是通过定义新的目标函数和相应的边界条件直接选取支撑特征,从而提高了特征选择的效率。同时由于定义了新的边界条件,在一定程度上减少了边缘误差的传播。文献[24]则通过分析发现 P-SVMs 方法的目标函数只是类内散度的一种特殊情况,因此使用类内散度作为新的目标函数并提出广义的势支撑向量机(generalized potential support vector machines, GPSFM),该方法不但提高了 P-SVMs 方法的适应性,同时还具有冗余度低、特征选择速度快的特点。然而,由于 P-SVMs、GPSFM 方法在特征选择过程中直接选取的是支撑特征,使得该类算法的时间复杂度依赖于特征空间的维数,同时类内散度在一定情况下会出现奇异性问题,因此,该类方法在处理小样本

高维数据时特征选择效率和精度会受到一定程度上的影响。从以上分析可知,目前大部分基于支持向量机的特征选择方法在一定程度上只能解决一些特定问题,如何提高方法的泛化能力、直接实现多分类的特征选择将是该类方法面临的挑战。

2 特征提取方法

特征选择是通过在高维的特征集中选取相应特征子集的方法来实现特征降维,而特征提取则是通过寻找某一特定变换将高维的特征空间转换为低维空间来实现特征降维。因此,在使用特征提取方法来解决实际问题时,如何根据问题实际选取适当的特征提取方法是解决问题的关键。近年来为了提高特征提取效率研究者提出了大量的特征提取方法,但总体上可以分为3类:基于主成分分析(the principal component analysis, PCA)^[25]和线性判别分析(the linear discriminant analysis, LDA)^[26]的特征提取方法、流形学习方法和基于张量理论的特征提取方法。

2.1 基于PCA、LDA的特征提取方法

PCA作为一种无监督的特征提取方法,它将方差的大小作为衡量信息多少的标准,并实现特征降维。PCA旨在保持协方差结构不变的前提下,计算高维数据的低维表示。然而该方法在计算协方差时要做大量的浮点乘法运算,因此相对时间复杂度较高,特别是在处理高维大数据时尤为明显。文献[27]则使用整数条件来替换浮点乘法运算提出Binary PCA(B-PCA)方法,从而大大降低了PCA方法的时间复杂度,然而B-PCA方法在处理大规模高分辨率的图像数据时还是显得力不从心。和无监督的PCA方法正好相反,LDA是一种有监督的特征提取方法,该方法通过最小化类内散度同时最大化类间散度(即保证类内最紧、类间最大)得到优化的投影变换。然而,LDA方法具有一个严重的弱点就是在处理小样本数据时类内散度矩阵会发生奇异性问题。为了解决LDA方法这一弱点,现在开发大量基于LDA基本原理的新方法来解决这一问题。比如PCA+LDA^[28]、RDA(regularization LDA)^[29]、OLDA(orthogonal LDA)^[30]、最大散度差判别准则(maximum scatter difference discriminant criterion, MSD)^[31]、2D-LDA(two-dimensional LDA)^[32]等等。特别是文献[33]通过在MSD方法中引入模糊概念,提出基于模糊最大散度差判别准则的

聚类方法,该方法不但一定程度上可以克服MSD方法受到调节参数的影响,同时可以实现无监督的特征提取并实现模糊聚类。然而以上各种线性方法并不能发现数据隐含的非线性结构,为了克服这个问题,目前主要使用基于核的方法和流形学习方法(下面单独讨论)来解决这个问题。基于核的方法就是使用核技巧将输入空间映射到高维的特征空间,使得在输入空间中的非线性问题转化成线性可分问题。因此,在线性的PCA、线性的LDA方法中分别引入核技巧就形成了非线性核方法:KPCA^[34]、KLDA^[35]。以上两种方法虽然是较为有效的非线性降维方法,但如何在解决实际非线性降维过程中选择适当的核函数,以及如何较为科学地设定相应的参数目前还不得而知,同时以上两种非线性降维方法并不能发现隐藏在数据中的局部非线性流形结构。

2.2 基于流形学习的特征提取方法

流形学习方法不像基于核的特征提取方法,流形学习方法可以直接发现隐藏在样本中的非线性流形结构。在过去的几年中已有许多的流形学习方法,比如等距特征映射(isometric mapping, Isomap)^[36]、局部线性嵌入(locally linear embedding, LLE)^[37]、拉普拉斯特征映射(Laplacian eigenmap, LE)^[38]等等。以上各种方法都属于非参数方法,从而避免了学习过程中参数对学习效率的影响,同时以上方法都是通过关注样本的局部几何结构来发现样本内在的全局信息,并且在求解过程中一定程度上避免了局部极值问题。然而,以上几种流形学习方法所对应的映射只能被定义在训练样本上,而新的测试样本上很难获得低维的投影映射(即out-one-sample problem),同时还表现出抗躁性能不强等缺点。为了克服这些不足目前国内外学者提出了一些有效的方法,比如周志华等人提出的基于集成的流形学习方法:Ensemble-Isomap^[39]、Saul K L等人提出的参数方法以及非参数方法^[40]、Li B等人提出的LLDE(locally linear discriminant embedding)^[41]等等。特别是,He X F等人提出的局部保持投影(locally preserving projections, LPP)^[42]方法,该方法作为拉普拉斯特征函数优化的线性近似方法,不但可以保持样本间局部几何结构,而且又能在低维的表示空间提供显式的嵌入映射函数,同时容易被非线性嵌入,从而发现高维非线性流形结构。为了提高LPP方法的泛化能力,He X F等人

通过定义特殊的拉普拉斯相似矩阵将 PCA、LDA 融入到 LPP 方法下作为一种特殊的情形。文献[43]则基于图嵌入(graph embedding)框架将 PCA、LDA、Isomap、LLE、LE、LPP 基于同一框架下,并从理论上分析了各种方法处于该框架下的特殊性和不可替代性。然而,目前具有的流形学习方法都有一个共同的缺陷,即容易引起所谓的“局部过学习(overlearning locality)”^[44]问题。为此,目前已有部分文献[44-46]进行了初步探讨,但如何选取更有效的方法来保证在进行有效的局部学习的同时,更好地保持全局信息,以及如何有效解决子空间流形学习问题和多簇流形学习问题是该类方法面临的挑战。

2.3 基于张量理论的特征提取方法

传统的特征提取方法在处理具有张量形式的数据(比如图像数据(second-order tensor)、视频数据(third-order tensor))时,一般会将张量形式转换成矢量表示的形式。然而这样处理会在一定程度上破坏张量数据固有的空间信息,同时会引起所谓的小样本问题和较高计算复杂度。因此,近来今年提出了许多基于张量(特别是矩阵模式)的特征降维方法,比如基于 PCA 的方法:2D-PCA(two-directional PCA)^[47]、matPCA(matrix-PCA)^[48]、(2D)²PCA(two-directional two-dimensional PCA)^[49]、Diagonal PCA^[50]和 B-2DPCA(binary two-directional PCA)^[51];基于 LDA 的方法:2D-LDA^[52]、matLDA^[48]、(2D)²LDA^[53]、2D-DWLDA(two-directional direct and weighted LDA)^[54]、2D-LTDA(second-order linear tensor discriminant analysis)^[55];基于 LPP 的方法:TSA(tensor subspace analysis)^[56]等等。以上各种方法不管基于行的(2D-PCA、2D-LDA、TSA)还是同时基于行和列的((2D)²PCA、(2D)²

LDA)在一定程度上都能保持张量数据原有的空间结构信息,提高模式分类精度,同时可以较大幅度地降低原有方法的时间和空间复杂度,特别是基于 LDA 方法的矩阵模式的方法还可以较好地解决 LDA 中类内散度奇异性问题,提高算法的鲁棒性。诚然,有些方法还具有其它一些特殊功能,比如 B-2DPCA 方法在执行效率上远远优于 2D-PCA,2D-LTDA 方法在降维是可以自动选取适当的维数,2D-DWLDA 方法则通过加权的方法来弱化相邻类别之间的层叠现象等等。特别是文献[57]首次实现了张量模式的非线性方法,对于提取张量模式的非线性关系具有一定的指导意义。通过以上分析得知,基于张量理论的特征提取方法目前主要致力于处理图像数据上,即使有些方法在理论上可以处理任意阶张量数据上^[43],但在实际测试使用时还都是处理的 2 阶图像数据,因此该类方法还存在着如下的挑战:一是如何使用有效的表示方法来处理高阶张量数据(比如视频数据);二是由于该类方法对应特征提取效果一定程度上依赖数据的具体矩阵模式,因此如何根据实际处理的问题选择相应的矩阵模式也很重要;三是由于该类方法可以在一定程度上解决小样本问题,那么是否可以将该类方法推广到处理高维矢量表示的数据上,以及如何在降低模式分类效果的情况下将矢量形式更有效转换成具体的矩阵模式等等。

3 总结与展望

通过以上分析,对于目前已有的各种特征降维方法都存在着这样那样不足,特别是如何根据新的实际去开发相应的特征降维新方法显得更为重要。

参考文献:

- [1] Huang J Z X, Ng M K, Rong H Q, et al. Automated variable weighting in k-Means type clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(5): 657-668.
- [2] Kira K, Rendell L A. A practical approach to feature selection[J]. In Proceedings of the 9th International Workshop on machine Learning. San Francisco, CA: Morgan Kaufmann, 1992: 249-256.
- [3] Sun Y J. Iterative Relief for features weighting: algorithms, theories, and applications[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1035-1051.
- [4] 皋军, 王士同. 具有特征排序功能的鲁棒性模糊聚类方法[J]. 自动化学报, 2009, 35(2): 145-153.
- [5] Lee K E, Sha N, Dougherty E R, Vannucci M, Mallick B K. Gene selection: A Bayesian variable selection approach[J]. Bioinformatics, 2003, 19(1): 90-97.

- [6] Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: Class discovery and class prediction gene expression monitoring[J]. Science, 1999, 286: 531 – 537.
- [7] Varma S, Simon R. Iterative class discovery and feature selection using minimal spanning trees[J]. BMC Bioinformatics, 2004, 5: 118 – 126.
- [8] TH B, Jonassen I. New feature subset selection procedures for classification of expression profiles[J]. Genome Biology, 2002, 3(4): 1 – 17.
- [9] Jacek B, Wlodzislaw D, Adam K, et al. Feature ranking methods based on information entropy with parzen windows[C]. International Conference on Research in Electrotechnology and Applied Informatics, Katowice, Poland, 2005.
- [10] Deng Z H, Chung F L, Wang S T. FRSDE: Fast reduced set density estimator using minimal enclosing ball approximation[J]. Pattern Recognition, 2008, 41(4): 1 363 – 1 372.
- [11] Kwak N, Choi C H. Input feature selection by mutual information based on Parzen window[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(12): 1 667 – 1 671.
- [12] Jing L P, Ng M K, Huang J Z X. An entropy weighted k – means algorithm for subspace clustering of high – dimensional sparse data[J]. IEEE Transactions on Knowledge & Data Engineering, 2007, 19(8): 1 026 – 1 041.
- [13] Yu L, Liu H. Feature selection for high – dimensional data: A fast correlation – based filter solution[J]. In Proceedings of the International Conference on Machine Learning 2003(3): 856 – 863.
- [14] Torkkola K. Feature extraction by non – parametric mutual information maximization[J]. Journal of Machine Learning Research, 2003, 3: 1 415 – 1 438.
- [15] Fleuret F. Fast Binary Feature Selection with Conditional Mutual Information[J]. Journal of Machine Learning Research, 2004, 5: 1 531 – 1 555.
- [16] Battiti R. Using mutual information for selecting features in supervised neural net learning[J]. IEEE Trans on Neural Networks, 1994, 5(2): 537 – 550.
- [17] Cang S, Yu H N. A new approach for detecting the best feature set[C]. In Proceedings of Networking, Sensing and Control [S. 1.]: IEEE CNF, 2005: 74 – 79.
- [18] Kwak N, Choic H. Input feature selection for classification problems[J]. IEEE Trans on Neural Networks, 2002, 13(1): 143 – 159.
- [19] Guoyou I, Weston J, Barnhill S, et al. Gene selection for cancer classification using support vector machines[J]. Machine Learning, 2002(46): 389 – 422.
- [20] Rakotomamonjy A. Variable selection using SVM – based criteria[J]. Journal of Machine Learning Research, 2003, 3: 1 357 – 1 370.
- [21] Claeskens G, Croux C, Kerckhoven J V. An information criterion for variable selection in support vector machines[J]. Journal of Machine Learning Research, 2008, 9: 541 – 558.
- [22] Hochreiter S, Obermayer K. Support vector machines for dyadic data[J]. Neural Computation, 2006, 18(6): 1 472 – 1 510.
- [23] 皋军, 王士同, 邓赵红. 广义的势支撑特征选择方法: GPSFM[J]. 计算机研究与发展, 2009, 46(1): 41 – 50.
- [24] Jolliffe I T. Principal component analysis[M]. New York: Spring, 1986.
- [25] Belhumeur P N, Hefanha J P, Kriegman D J, et al. fisherface: recognition using class specific linear projection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711 – 720.
- [26] Tang F, Crabb R, Tao H. Representing image using nonorthogonal Haar – like bases[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(12): 2 120 – 2 134.
- [27] Yu H, Yang J. A direct LDA algorithm for high – dimensional data – with application to face recognition[J]. Pattern Recognition, 2001, 34(10): 2 067 – 2 070.
- [28] Friedman J H. Regularized discriminant analysis[J]. Journal of the American Statistical Association, 1989, 84(405): 165 – 175.
- [29] Ye J. Characterization of a family of algorithm for generalized discriminant analysis on undersampled problems[J]. Journal of Machine Learning Research, 2005, 6: 483 – 502.
- [30] 宋枫溪, 张大鹏, 杨静宇, 等. 基于最大散度差鉴别准则的自适应分类算法[J]. 自动化学报, 2006, 32(4): 541 – 549.
- [31] Li M, Yuan B Z. 2D – LDA: a statistical linear discriminant analysis for image matrix[J]. Pattern Recognition Letters,

- 2004(26):527-532.
- [32] Scholkopf B, Smola A, Muller K R. Nonlinear component analysis as a kernel eigenvalue problem[J]. *Neural Computation*, 1998,10(5):1 299-1 319.
- [33] Mika S, Ratsch G, Weston J, et al. Constructing descriptive and discriminative nonlinear features: Rayleigh coefficients in kernel feature spaces[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003,25(3):623-628.
- [34] Tenenbaum J B, Silva V D, Langford J C. A Global Geometric Framework for Nonlinear Dimensionality Reduction[J]. *Science*, 2000,290:2 319-2 323.
- [35] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding[J]. *Science*, 2000,290:2 323-2 326.
- [36] Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. *Neural Computation*, 2003,15(6):1 373-1 369.
- [37] 詹德川,周志华. 基于集成的流形学习可视化[J]. *计算机研究与发展*, 2005,42(9):1 533-1 537.
- [38] Saul K L, Roweis S. Think globally, fit locally: unsupervised learning of low dimensional manifolds[J]. *Journal of Machine Learning Research*, 2003,4:119-155.
- [39] Li B, Zheng C H, Huang D S. Locally linear discriminant embedding: an efficient method for face recognition[J]. *Pattern Recognition*, 2001,41(12):3 813-3 821.
- [40] He X F, Niyogi P. Locality Preserving Projections. *Advances in Neural Information Processing Systems (NIPS)* [J]. Canada: Vancouver, 2003.
- [41] Yan S C, Zhang B Y, Zhang H J, et al. Graph embedding and extensions: a general framework for dimensionality reduction[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007,29(1):40-51.
- [42] Wang H X, Chen S B, Hu Z L, et al. Locality - preserved maximum information projection[J]. *IEEE Trans on Neural Networks*, 2008,19(4):571-585.
- [43] Cai D, He X F, Zhou K, et al. Locality Sensitive Discriminant Analysis[C]. In *Procession 2007 International Joint Conference on Artificial Intelligence (IJCAI07)*, Hyderabad, India, Jan. 2007.
- [44] Yang J, Zhang D, Frangi A F, et al. Two - dimensional PCA: a new approach to appearance - based face representation and recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004,26(1):131-137.
- [45] Chen S C, Zhu Y L, Zhang D Q, et al. Feature extraction approaches based on matrix pattern: MatPCA and MatFLDA [J]. *Pattern Recognition Letters*, 2005,26(8):1 157-1 167.
- [46] Zhang D Q, Zhou Z H. (2D)²PCA: Two - directional two - dimensional PCA for efficient face representation and recognition[J]. *Neurocomputing*, 2005,69(1-3):224-231.
- [47] Zhang D Q, Zhou Z H, Chen S C. Diagonal principal component analysis for face recognition[J]. *Pattern Recognition*, 2006,39(1):140-142.
- [48] Pang Y W, Tao D C, Yuan Y, et al. Binary two - dimensional PCA[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 2008,38(4):1 176-1 180.
- [49] Li M, Yuan B Z. 2D - LDA: A statistical linear discriminant analysis for image matrix[J]. *Pattern Recognition Letters*, 2005,26(5):527-532.
- [50] Nagabhushan P, Guru D S, Shekar B H. (2D)²FLD: An efficient approach for appearance based object recognition[J]. *Neurocomputing*, 2006,69(7-9):934-940.
- [51] Zhi H Q, Ruan Q Q. Facial expression recognition based on two - dimensional discriminant locality preserving projections [J]. *Neurocomputing*, 2008,71(7-9):1 730-1 734.
- [52] Nie F P, Xiang S M, Song Y Q, Zhang C S. Extracting the optimal dimensionality for local tensor discriminant analysis [J]. *Pattern Recognition*, 2009,42(1):105-114.
- [53] He X F, Cai D, Niyogi P. Tensor Subspace Analysis[C]. *Advances in Neural Information Processing Systems (NIPS)*, 2005.

Analysis of Stress Distribution and Length on the Mass Spring without Load in Rotating System

YU Xiao-ming

(Department of Fundamental Sciences Teaching, Yancheng Institute of Technology, Jiangsu Yancheng 224051, China)

Abstract: When considering the mass of spring in rotating system the spring will not uniformly stretches, which makes the stress distribution point – to – point changes. In this article the author theoretically deduces the formulas for stress distribution and spring length without load in uniform rotating system when the angle which is made by the plane of rotating system and horizontal plane is arbitrary. According to the formulas the author writes out the programs and respectively maps out both the stress distribution under the circumstances of three angular velocity and three inclined angels and spring length in the conditions of three inclined angels. In engineering design one can both easily and quickly selects spring and accurately determine the stress distribution or forecast the length of the spring so as to protect springs.

Keywords: rotating system; mass spring without load; linear density of mass; stress distribution; spring length

(责任编辑:张英健;校对:沈建新)

(上接第20页)

- [54] Guyon I, Elisseeff A. An instruction to variable and feature selection[J]. Journal of Machine Learning Research, 2003 (3):1 157 – 1 182.
- [55] 皋军,王士同. 基于模糊最大散度差判别准则的聚类方法[J]. 软件学报,2009,20(11):2 939 – 2 949.
- [56] 皋军,王士同,邓赵红. 基于全局和局部保持的半监督支持向量机[J]. 电子学报,2010,38(7):1 626 – 1 633.
- [57] 皋军,王士同. 基于矩阵模式的类内散度支持向量机[J]. 电子学报,2009,37(5):1 051 – 1 057.

An Introduction to Dimensionality Reduction Algorithms in Intelligent Recognition

GAO Jun

(School of Information Engineering, Yancheng Institute of Technology, Jiangsu Yancheng 224051, China)

Abstract: Dimensionality reduction algorithms, as the key technologies of data preprocessing in intelligent recognition, have been used successfully recently. On the basis of the comparison of the classical dimensionality reduction algorithms, the characteristics and advantages of each algorithm are formulated from the perspectives of feature – selection and feature – extraction separately and the corresponding problems and challenges about each of them are pointed out to certain extent.

Keywords: feature dimensionality reduction; feature selection; feature extraction; intelligent recognition

(责任编辑:沈建新;校对:张英健)