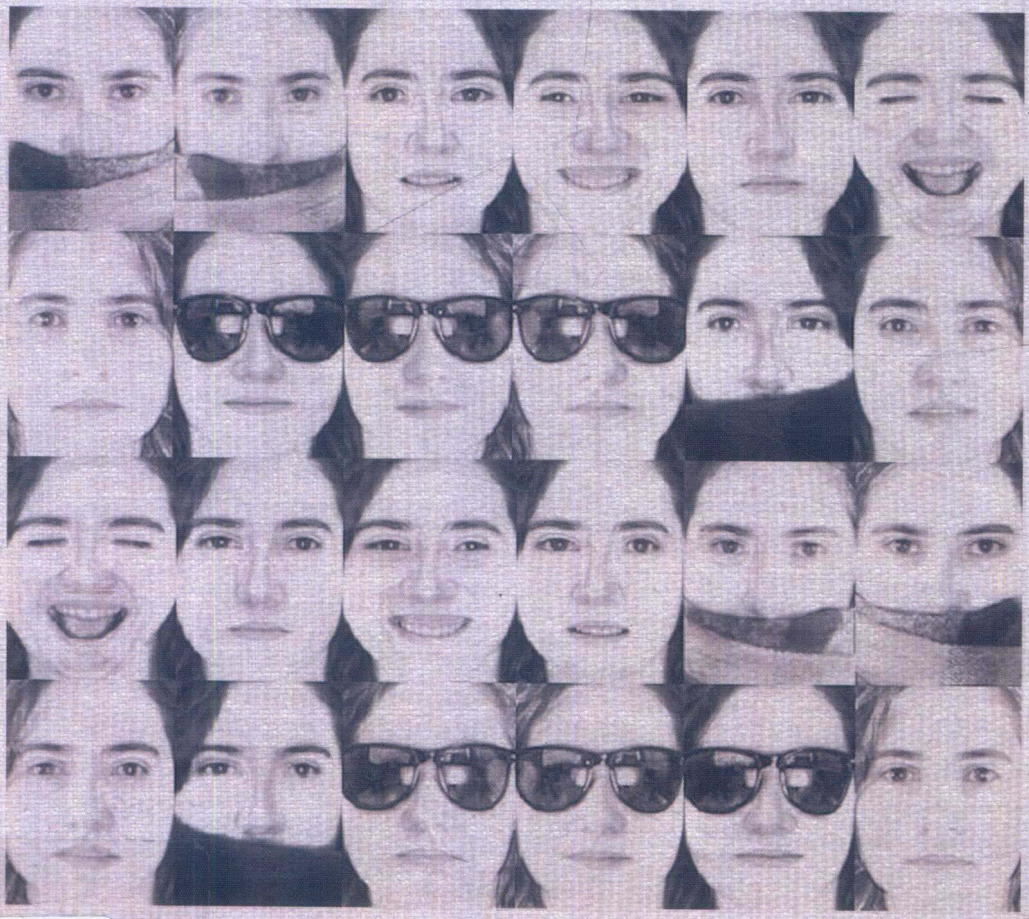
新型特征抽取算法研究

XINXING TEZHENG
CHOUQU SUANFA YANJIU

范自柱 著

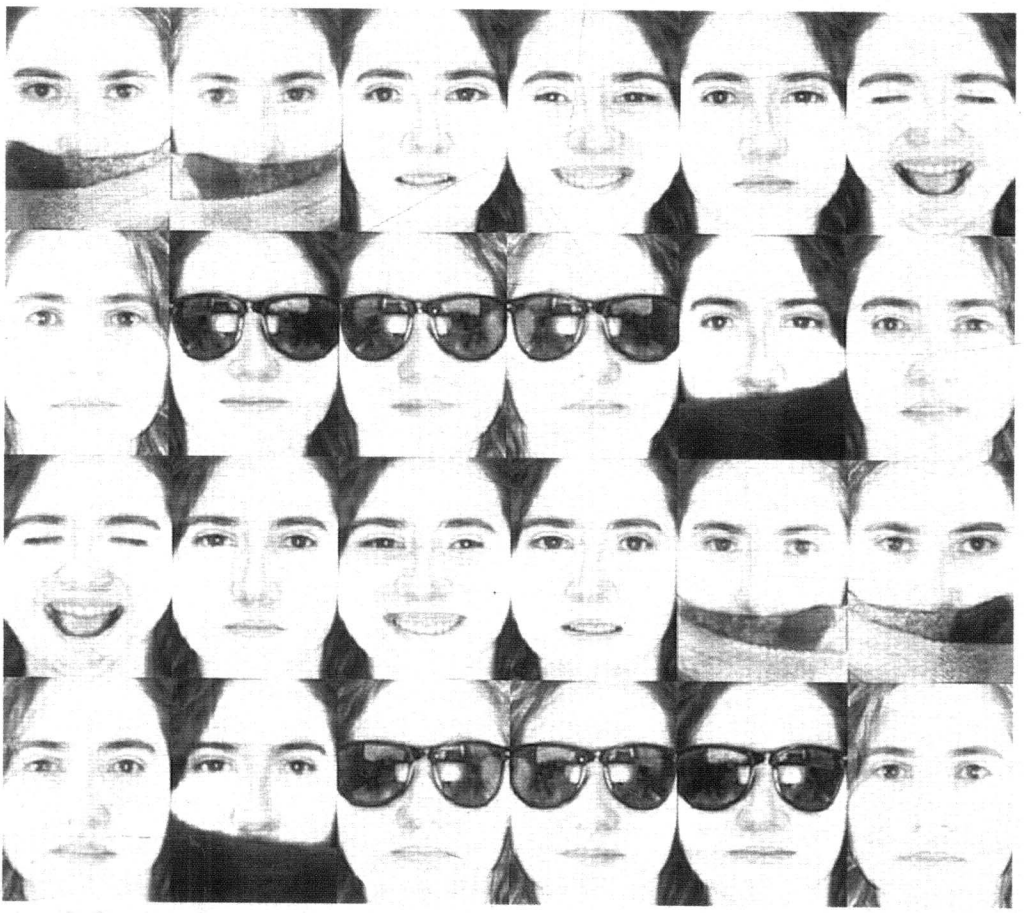


中国科学技术大学出版社

新型特征抽取算法研究

XINXING TEZHENG
CHOUQU SUANFA YANJIU

范自柱 著



中国科学技术大学出版社

内 容 简 介

本书的主要内容是特征抽取方法在人脸识别和其他分类任务中的应用。首先介绍了改进的特征抽取方法以提高经典特征抽取方法的分类精度。接着介绍了几种特征抽取方法，它们的目的是提高特征抽取算法的计算效率。最后从一个新颖的角度去描述特征抽取方法，即从样本表示的角度来阐述特征抽取，这源自目前备受关注的压缩感知理论。

本书既可作为自动化、计算机、电子工程和信息管理等专业本科生、研究生和研究人员的科研用书，又可作为从事模式识别、机器学习、计算机视觉和图像处理等工作的开发人员的参考资料。

图书在版编目(CIP)数据

新型特征抽取算法研究/范自柱著. —合肥：中国科学技术大学出版社，2016.12
ISBN 978-7-312-04049-8

Ⅰ.新… Ⅱ.范… Ⅲ.图像识别—研究 Ⅳ.TP391.41

中国版本图书馆 CIP 数据核字(2016)第 263065 号

出版 中国科学技术大学出版社
安徽省合肥市金寨路 96 号，230026
http://press.ustc.edu.cn
印刷 合肥市宏基印刷有限公司
发行 中国科学技术大学出版社
经销 全国新华书店
开本 710 mm×1000 mm 1/16
印张 10.25
字数 212 千
版次 2016 年 12 月第 1 版
印次 2016 年 12 月第 1 次印刷
定价 32.00 元

前　言

在模式识别与机器学习领域，人们在对原始数据进行分类或识别前，往往需要对这些数据进行预处理。特征抽取（也称数据降维）是非常重要的一种预处理方法。在很多模式识别方法中，如果不对原始数据进行特征抽取，而直接对数据进行分类或识别，计算量往往会比经特征抽取后进行分类的大，而且识别的效果常不理想。相反，若在对数据进行识别或分类前，利用特征抽取获得数据的关键信息，会有效提高数据分类的效果和效率。其中的重要原因是，特征抽取可以使数据的维数，尤其是高维数据（如图像或视频等）的维数大大降低，其计算量自然会减少。也就是说，特征抽取可以有效提高分类的效率，减少计算的时间。另一方面，一般的原始数据中常存在噪声或冗余信息，这些噪声或冗余信息很可能会干扰分类过程，使分类的效果变差。特征抽取的一个作用就是，可以有效去除噪声或冗余信息。因此，对原始数据抽取合适的特征，理论上是可以改善数据的识别效果的。

实际上，特征抽取的过程是利用一种或多种变换方法将数据从原始输入空间变换或映射到新的特征空间中。如前所述，特征抽取往往有两个目的：一是改善数据的分类或识别效果，二是提高计算效率，或者两者兼而有之。本书正是基于以上目的提出了几种特征抽取算法。

为了改善经典特征抽取算法如主成分分析（principal component analysis，PCA）等方法的识别效果，给出了改进的主成分分析（modified PCA，MPCA）、局部线性鉴别分析（local linear discriminant analysis，LLDA）、改进的核 Fisher 鉴别分析（improved kernel Fisher discriminant analysis，IKFDA）和局部最小均方误差（local minimum squared erro，LMSE）方法，这些方法都较明显地改善了原特征抽取方法的分类效果。

经典的基于核的特征抽取方法如核主成分分析（kernel PCA，KPCA）和核最小均方误差（kernel MSE，KMSE）分类方法虽然识别效果好，但它们的计算效率低，尤其是特征抽取的效率不高，从而影响了此类方法整体的识别效率。为此，本书介绍了对这两种算法进行改造后的方法，即快速的 KPCA 和 KMSE 算法，分别记为 FKPCA（fast KPCA）和 FKMSE（fast KMSE）。基于核的特征抽取方法，往往离不开核函数的参数选择问题，而此问题常常比较费时。为实现快速的核方法，本书专门利用一章来介绍核函数参数的快速选择算法，它可以使本书提出的快速

核方法进一步提高计算速度。当然，该方法也适用于任何一种需要选择核函数参数的核方法。

近几年，基于稀疏表示理论分类方法（sparse representation based classification，SRC）在模式识别与机器学习中受到广泛关注。该方法在对高维数据如人脸图像数据等进行学习分类时比较有优势，特别地，当这些高维数据被腐蚀或受噪声污染时，SRC 方法的效果比其他经典的分类识别方法更优。在有些场合，SRC 被看作一种分类器。其实，SRC 也可以被当作一种特征抽取方法。在 SRC 中，先用全体训练样本数据表示一个新的样本，经表示得到的系数就可以看作被表示样本的特征。所以，SRC 中的表示过程也是特征抽取过程。在处理高维数据时，若数据的维数远大于训练样本的个数，通过稀疏表示仍会达到数据降维的目的。稀疏表示是基于 L_1 范数的，同样，利用其他范数如 L_2 范数也可以通过对一样本进行表示来抽取其特征。本书最后将从这一角度出发，介绍几种基于表示的特征抽取方法，进而对特征抽取方法做一个新的诠释。

本书的工作得到国家自然科学基金项目（编号：61472138，61263032）和江西省自然科学基金项目（编号：20161BAB202066）的资助，特此感谢。另外，感谢硕士研究生康利攀、邹妍、胡玉林和熊磊等同学为本书收集整理了部分资料。

由于作者水平有限，书中难免存在不妥之处，敬请读者批评指正。

范自柱

2016 年 6 月于南昌

目　录

第1章 引 论

1.1 背 景

近年来，随着计算机和网络等技术的快速发展，人们需要处理的数据越来越多，也越来越复杂。目前，很多学科和应用领域都会遇到大规模复杂数据的处理与分类问题，如模式识别与机器学习领域的生物特征识别，视频内容理解和标注，海量网页和社交网络节点数据的检索或分类，生物工程的基因数据分析，医学图像检索和分类，环境动态监测等所需的遥感图像处理，天气预报中的卫星云图自动分析，月球等其他天体的地貌识别和大规模网络数据的实时处理等。其中，社交网络中数据节点规模庞大，有的社交网络中的节点多达上亿个，如著名的 Facebook 和 Twitter 等。而且，网络中的节点大都是非结构化的，其类型多种多样，包括订阅消息的节点、发布消息的节点和这两种节点的混合等。网页这一被搜索引擎处理的重要数据对象，包含的数据类型也不是唯一的。一个网页中常包含图像、视频、文本和超链接等对象。同样，生物工程中的蛋白质数据和网络入侵检测需要处理的网络攻击数据等都是非结构的复杂数据。一般地，这些要处理的数据大都数量巨大、维数高、结构复杂、分布没有明显的规律。在此，我们称这些数据为复杂数据[1]。

如果用人工方式对这些大规模复杂数据进行处理，哪怕是给每个数据样本编号，都将是难以顺利完成的。对这些数据的进一步分析和处理，比如对数据的分类或识别等，人工就更无法快速有效完成了。因此，必须要借助计算机等工具，使用合适的方法，根据不同的需求或应用场合，从不同方面来自动处理分析这些复杂数据。

复杂数据处理的一种常见目的就是对这些数据的分类学习或识别。以生物特征识别为例，它包括很多分支，如人脸识别、指纹识别、掌纹识别、虹膜识别和步态识别等。相对于其他生物特征识别，人脸识别技术最不具“侵犯性”，它不像指纹和

掌纹等识别技术那样，需要被识别者的紧密配合，甚至在有些情况下，被识别者在人脸识别时没有觉察。因此，人脸识别是最方便的一种生物特征识别技术，它可应用于考勤、门禁、民航、军事安全、银行金融系统和反恐等。近年来，人脸识别受到人们越来越多的关注，并付诸应用。例如，“9·11”恐怖事件之后，美国率先在波士顿等机场开始使用人脸识别系统，用来自动搜寻恐怖分子目标。在2008年北京奥运会开幕式期间，人脸识别系统被成功用于参加者的身份验证，该系统需要处理大规模的身份识别工作。以上这些系统都需要对数据量巨大的复杂人脸图像进行识别。

在上述人脸识别等应用中，数据维数高，结构复杂，分布也没有明显的规律。在对人脸图像等高维复杂数据进行分类学习或识别前，需要先抽取这些复杂数据的特征，然后再利用某个分类器如最近邻分类器对数据进行分类或识别。对数据特征抽取[2]的成功与否直接关系到分类的效果。在很多应用场合，特征抽取都是数据分类前不可或缺的环节。特别地，对于高维复杂数据，使用合适的特征抽取方法不仅会提高数据的分类精度，还可以大大缩短分类的时间。特征抽取对于复杂数据分类意义重大。因此，非常有必要对特征抽取在数据分类中的应用作全面系统的研究。

1.2 研究目的和意义

本书研究的一个主要目的就是针对不同的应用场合，提出新颖的特征抽取算法，来改善复杂数据的分类学习效果。在很多情况下，分类学习的效率是与数据集的规模成反比的。对于规模较大的数据或海量数据，分类学习效率往往不高，尤其是在高维的数据空间，提高特征抽取的计算效率进而提高分类效率将是本书的又一个重要目的。总之，本书将从上述两点出发，提出一些比较理想的特征抽取算法来提高复杂数据的分类学习效果或计算效率。

至于研究的意义，主要体现在两方面。理论上，将特征抽取应用于复杂数据的分类学习是一个前沿交叉课题，它涉及很多学科，如模式识别、机器学习、图像处理、数值优化和统计理论等。对此课题持续深入的研究将给这些相关学科提供一些新的研究思路，定会带动这些学科的进一步发展。在应用方面，复杂数据的分类学习应用范围非常广泛，它可用于多个应用领域，如生物特征识别、网页文档数据与分类、视频内容理解和标注、医学图像检索、天气预报中的卫星云图自动识别和大规模网络数据的实时处理等。如何从上述复杂数据中抽取有效特征越来越受到人们的关注。不仅在学术界，复杂数据的分类学习在工业界也受到高度关注。近

年来，Google、IBM 和微软等国际巨头都纷纷把目光投向这一领域，着手研究从复杂数据中抽取有效特征来对这些数据进行分类或做其他后续处理分析。

1.3 特征抽取方法概述

一般地，特征抽取是指基于某一优化准则，将高维数据变换到低维特征空间的过程[3]，或者是寻找面向某种应用的对数据最佳表示的过程，其目的是通过变换来发现数据的内在分布特征，或者寻找针对数据模式的最具鉴别性的描述。

特征抽取有时也被称为特征提取，根据特征抽取中使用变换的线性与非线性特点，特征抽取方法主要可分为线性特征抽取与非线性特征抽取方法。

目前主流的特征抽取方法都是基于统计学习的，我们对此将做重点介绍。

另外，增量学习形式的特征抽取很受关注，在此做简单介绍。近年来，基于表示理论的分类方法在模式识别和机器学习中受到较多的关注，对其研究也方兴未艾。实际上，基于表示理论的方法也可以看作一种新型的特征抽取方法，这里也做简要介绍。

1.3.1 线性特征抽取方法

由于应用上的简单方便，线性特征抽取是特征抽取中最常用的方法之一。它主要有下面几种形式：主成分分析（principal component analysis，PCA）[4]、线性鉴别分析（linear discriminant analysis，LDA）[5,6]、局部保持投影（local preserving projection，LPP）[7] 和独立成分分析（independent component analysis，ICA）[8]。这些方法也可称作线性子空间学习方法。

1.3.1.1 主成分分析

主成分分析也叫主分量分析，最早由 Pearson 于 1901 年提出，它的基本思想是对含有大量相关成分（变量）的数据集约减维数，同时尽可能保持数据间的差别[4]。它可以通过一个线性变换（K-L 变换），将数据变换到新的低维空间，在这个新空间中，数据的方差达到最大。在主成分分析中，把训练样本的协方差矩阵对应于前面最大特征值的那些特征向量称为主成分，这些主成分之间是互不相关的。使用 PCA 的一个主要目的是约减维数或称降维（在有噪声的情况下，降维的同时

可以去除一些或全部噪声),以发现数据样本中的本质分布特点。它已经被广泛应用于多个领域,尤其是高维复杂数据的处理,如人脸识别与重建等[9]。

将 PCA 用于人脸识别领域的经典方法是 Turk 和 Pentland 提出的特征脸(eigenfaces)方法[10]。在这一方法中,人脸图像样本被拉成列向量,对这些样本做中心化处理,即每个样本向量减去它们的均值向量,再计算它们的协方差矩阵。对此协方差矩阵做特征分解后,就可以选择部分特征向量作为变换轴,对原始输入样本实施变换,抽取样本特征。事实上,各个样本在主成分分析过程中所起的作用可以是不同的。为了能更好地去除噪声,Danijel Skočaj 和 Aleš Leonardis 等人提出了加权 PCA(weighted PCA, WPCA)[11]。在 WPCA 中,每个样本都被赋予一权值,再重新建立 PCA 模型,即 WPCA。经典 PCA 可以看作 WPCA 的一种特殊形式,即在 WPCA 中,数据样本的权值全取 1。与 PCA 相比,WPCA 对数据噪声或图像中的遮挡更具鲁棒性,这种情况下,用它抽取的特征要优于经典 PCA。

在上述 PCA 中,需要将一个图像样本的矩阵形式变成列向量或者行向量形式,因此,它又可以被称为一维 PCA(one-dimensional PCA, 1DPCA)。1DPCA 在处理图像等矩阵形式的数据时,由于需要将矩阵拉成向量,矩阵中包含的一些空间信息可能会丢失,从而会影响特征抽取的效果。为解决这个问题,杨健等人提出了二维 PCA(two-dimensional PCA, 2DPCA)[12]。2DPCA 不需要将矩阵形式的数据样本拉成向量,它可以直接对矩阵进行运算。与 1DPCA 相比,在对图像等矩阵数据进行特征抽取的过程中,2DPCA 涉及的协方差矩阵规模小很多,既避免了某些信息的丢失,提高了特征抽取效果,又大大加快了计算的速度。

如果数据是彩色图像或视频等张量形式,直接用 1DPCA 和 2DPCA 对其抽取特征会比较困难。为有效抽取张量数据的特征,H. Lu 等人提出了多线性主成分分析(multilinear principal component analysis, MPCA)[13,14]。MPCA 推广了 1DPCA 和 2DPCA,可以处理更多类型的数据。

1.3.1.2 线性鉴别分析

线性鉴别分析(LDA)是一种有监督的特征抽取和降维方法[15],它的基本思想最初由 Fisher 于 1936 年提出。经典 LDA 算法的主要目的就是寻找一些变换轴(在 LDA 中称为最优鉴别向量),使得变换后数据的类间散度与类内散度之比最大[6]。一般地,如果数据集中有 c 个类别,LDA 在抽取数据特征时,将把数据样本降到 $c-1$ 维。

在对高维复杂数据如人脸图像抽取特征时,经典 LDA 中的类内散度矩阵往往是奇异的,也就是会遇到通常所说的"小样本"问题。这种情况下,就不能直接使用经典 LDA。一种解决办法是利用正则化技术,即在类内散度矩阵上加一个很小的同阶单位矩阵,使其非奇异,这样就可以直接应用经典 LDA。另一种解决办法的

代表是 Chen 等人提出的零空间 LDA(null LDA, NLDA)方法[16],它较好地解决了"小样本"问题。随后,H. Yu 和 J. Yang 提出了直接 LDA(direct LDA, DLDA)[17]来解决此问题,取得了较好的效果。LDA 的另一种实现方式是基于最大间隔思想,寻找鉴别向量,使得变换后数据的类间散度与类内散度之差最大化,即最大间隔准则(maximum margin criterion, MMC)[18]。由于在求解过程中,不需要计算类内散度矩阵的逆,所以也就避免了"小样本"问题。

同样,在前述的 LDA 方法中,数据都是列向量或行向量形式,因此,它们都可以称为 1DLDA。与 1DPCA 类似,1DLDA 对向量形式的数据处理起来比较方便,但不适合直接处理矩阵形式的数据,如图像等。为能直接处理矩阵形式的数据,H. Xiong 等人提出了二维 Fisher 线性鉴别(two-dimensional FLD, 2DFLD)[19],W. H. Yang 和 D. Q. Dai 等人提出了二维最大间隔(two-dimensional maximum margin criterion)特征抽取方法[20],简称为 2DMMC。2DFLD 和 2DMMC 可以看作二维 LDA(two-dimensional LDA, 2DLDA)[21]的不同形式。在处理图像等高维复杂数据时,1DLDA 往往需要先用 PCA 对数据做降维等预处理,而 2DLDA 可以直接使用,比 1DLDA 方便快捷。为对更复杂的数据如视频等做分析,S. Yan 等人提出了一种多线性鉴别分析[22],它可以对张量形式的数据进行鉴别分析。

1.3.1.3 局部保持投影

局部保持投影(LPP)是由 X. He 等人提出的,最先用于人脸识别[7],它是流形学习的一种线性化方式。与 PCA 试图保持数据的全局结构不同,LPP 寻求一种嵌入模型,来保持数据的局部信息。在实际的分类问题中,数据的局部流形结构往往比全局的欧氏结构重要,特别是用最近邻分类器来分类。虽然 LPP 是一种无监督的特征抽取或降维方法,但它拥有一定程度的鉴别能力。在 LPP 中,需要用一个局部参数来确定一个数据点的局部范围。当需要保持数据的全局结构时,可以把局部参数取得很大甚至是无穷大,再取 LPP 特征方程最大特征值对应的那些特征向量作为变换轴。这时,LPP 实际上就变成了经典 PCA 了。如果要保持数据的局部结构信息,可以把此局部参数取得充分小,取 LPP 特征方程中最小特征值对应的那些特征向量作为变换轴,则经典 LDA 就被包含在 LPP 框架内了。与 LDA 只能用于有监督学习分类不同的是,LPP 既可以用于有监督学习,又可以用于无监督学习分类。

上述 LPP 在处理图像等数据时,还是要先把图像变成向量形式再做处理。因此,与前面的 1DPCA 和 1DLDA 一样,经典 LPP 也是基于向量形式的,可以称为 1DLPP。在处理图像等矩阵形式的数据时,1DLPP 需将图像矩阵拉成向量后,再用 PCA 等方法对数据样本降维。为能用 LPP 方法直接处理图像等矩阵数据,S. Chen 等人提出了二维 LPP(two-dimensional LPP, 2DLPP)[23],它可以直接处

理矩阵形式数据而不需要用 PCA 给数据样本降维。

1.3.1.4 独立成分分析

独立成分分析(ICA),也称为独立分量分析。它的目的是寻找对非高斯分布数据的一种线性表达,使得数据的各成分之间在统计上独立或尽可能独立[24]。ICA 是来自信号处理领域的一种方法,最初提出的目的是用来解决盲信号分离的,也就是著名的"鸡尾酒会问题":假设在一房间内,有两个人同时说话,用两个处在不同位置的麦克风收集到两个信号,每个信号都是两个人说话的加权和,如何仅利用收集到的两个信号来估计出每个人说话的信号?从数学上看,假设有一组独立的原始信号,经某一线性系统混合后,得到一组新的数据,虽然我们不知道具体的混合过程,但仍然可以找到一个分离矩阵将原始信号从混合数据中分离出来。

近年来,ICA 不仅在信号处理领域使用广泛,在模式识别与机器学习领域也备受关注。Bartlett[25] 和 Liu[26] 将 ICA 用于人脸表示和识别,发现当把余弦距离用于相似度量方法时,ICA 的效果优于 PCA。我们知道,PCA 是基于二阶统计信息的,而 ICA 不仅基于二阶统计信息,还利用了其他高阶统计信息。因此,ICA 可以看作 PCA 的一个推广。ICA 的一个常用实现方法是快速 ICA 方法(fast ICA[8,27])。

在这几种线性特征抽取方法中,除了 LDA 在建模过程中用到样本的类别信息,其他三种方法都不使用类别信息建立学习模型。也就是说,LDA 是有监督的学习方法,侧重于对样本的分类学习。而 PCA、LPP 和 ICA 都是无监督的,它们侧重于对数据的表示与重建。

除了上述四种线性特征抽取方法,还有非负矩阵分解(non-negative matrix factorization, NMF)[28,29]等线性方法,NMF 常用于数据聚类等应用中。

1.3.2 非线性特征抽取方法

除了线性特征抽取方法,近几年,人们又提出了非线性特征抽取方法,它主要包括两类:第一类是核方法,第二类以流形学习为代表。

1.3.2.1 基于核方法的非线性特征抽取

核方法的历史可追溯到 20 世纪五六十年代。Aronszajn 于 1950 年给出了再生核理论[30],之后,1964 年特征空间中的核视作内积这一思想被引入到模式识别与机器学习领域[31]。1992 年,Boser、Guyon 和 Vapnik 讨论了高斯和多项式核函

数，并在机器学习领域把核函数与大间隔超平面相结合，由此产生了支持向量机(support vector machine，SVM)[32]。SVM是一种有效利用有序风险最小化或结构风险最小化(structural risk minimization，SRM)原则的统计学习方法，它常用一非线性映射，将原始输入空间的数据样本映射到某一特征空间中，基于最大间隔寻找最优分类面，处于分类边界上的点被称为支持向量。核方法中的特征空间往往是一个高维甚至无穷维希尔伯特空间(Hibert space)，它是一个具有完备性和可分性的内积空间。特征空间中的数据或向量不需显示表示，可以用向量间的内积来表示它们之间的关系，而向量间的内积又可以在不明确给出非线性映射的情况下，用核函数取代，这种思想就是所谓的“核技巧”(kernel trick)。使用“核技巧”后，就可以在原始输入空间中计算，达到了利用线性手段来求解非线性问题的目的。

使用核方法的一个优点是，在原始输入空间线性不可分的样本，经非线性映射后，在高维特征空间变得线性可分。所以，核方法更有利于对数据的分类。SVM提出之后，很多线性特征抽取方法被扩展为基于核的非线性方法。事实上，任何一个涉及内积运算的线性特征抽取方法都可以经核扩展变成非线性方法。其中，经典代表就是核主成分分析(kernel principal component analysis，KPCA)[33]、核鉴别分析(kernel discriminant analysis，KDA)[34-37]、核最小均方误差(kernel minimum squared error，KMSE)[38]以及多核学习方法[39]等。

1998年，B. Schölkopf首次将核方法思想应用到PCA中，提出了著名的KPCA算法[33]。在KPCA中，用一非线性映射将数据样本映射到高维特征空间，在此特征空间中施行PCA，其中的非线性映射可以通过核函数如高斯核函数来定义。KPCA的主要计算步骤是：利用核函数计算核矩阵，再对此矩阵做特征分解，最后抽取样本的特征。虽然KPCA在特征空间中施行的是线性变换，但对于原始输入空间，它依然可以看作非线性的。与PCA相比，KPCA可以有效提取数据中的非线性信息。一般地，KPCA的特征抽取效果优于PCA，它比较适合对复杂数据进行特征抽取。

继KPCA提出之后，S. Mika[35]和G. Baudat[34]几乎同时提出了利用Fisher准则的一种KDA算法，即KFDA算法。KFDA的基本思想同KPCA类似，先将原始输入空间中的样本，经由某个核函数确定的非线性映射，变换到高维特征空间，再利用Fisher准则进行线性鉴别分析。杨健等人指出，KFDA实际上是经KPCA处理后再施行LDA[40]，这大大方便了KFDA的实现。KDA的另一种实现形式是核散度差鉴别分析(kernel scatter-difference-based discriminant analysis，KSDA)[36]，利用了最大散度差准则，此准则来源于MMC思想。与KPCA不同，KDA是有监督的，侧重于分类，而KPCA主要侧重于对数据的表示。与经典LDA相比，只要能够指定合适的核函数并能选择好此函数的最佳参数，KDA往往会取得较好的鉴别分类效果。

面向分类的最小均方误差(minimum squared error, MSE)也是一种线性鉴别方法[41]。2001 年,Xu 等人借助核思想将它拓展成核最小均方误差(kernel minimum squared error, KMSE[38]或 kernel least mean square, KLMS[42])。我们知道,在两类问题中,当样本类别标签满足一定条件时,MSE 的特征抽取效果就与 Fisher 鉴别分析的相同。Xu 也证明了 KFDA 和最小均方支持向量机(least squares support vector machines, LS-SVM)都是 KMSE 的一种特殊形式[38]。当样本数趋于无穷时,KMSE 模型将以最小均方误差逼近特征空间中的贝叶斯判别函数。因此,理论上,KMSE 的应用范围比 KFDA 和 LS-SVM 更广。

为解决 LPP 很难抽取数据非线性特征的问题,G. Feng 等将 LPP 核化,提出了基于核的 LPP(kernel LPP, KLPP)[43]。该方法由两步构成,即 KPCA + LPP。对于 ICA,J. Yang 将核技巧和该方法结合,提出了核 ICA(kernel ICA, KICA)[44],将其用于抽取人脸特征。上述核方法的共同点是只使用一个核函数来获取数据的非线性信息。然而,在一些非常复杂的情况下,比如数据异构或不规则、样本分布不平坦[39],这时候,使用单一的核函数很难对这类数据做有效处理。一个解决办法就是,将若干核函数做线性组合,构成一个多核函数,再利用新的多核函数学习抽取数据特征并分类。多核学习的思想最早由 Lanckriet 等人于 2004 年引入[45],他们用半正定规划(semidefinite programming, SDP)来学习由多核函数定义的核矩阵。随后,多核学习引起了广泛关注,陆续提出了多种多核学习算法[46-49]。多核学习虽然能够在某些方面取得好的效果,但其计算效率往往比较低,这是它需要解决的一个问题。

1.3.2.2 基于流形学习的非线性特征抽取

流形学习是非线性特征抽取或降维的一种方法。在很多现实应用中,数据常以复杂的高维形式呈现,如果这些数据的结构由某些内在规律确定,那我们就可以称其为流形。从几何上直观地看,一个 d 维流形其实就是一个在任一点处局部同胚于一个 d 维的欧氏空间。假设高维数据近似地位于一个可能与人的认知流形一致的低维空间流形上,通过保持流形的某些特征而获得这些离散数据点的低维表示,这一过程就是流形学习。换句话说,流形学习就是从高维采样数据中恢复低维流形结构,找出相应的嵌入映射来实现降维。它不需要计算流形的具体表示形式,而是从观测到的数据中寻找数据产生的内在规律[50]。

流形学习的主要目的是寻找蕴含在高维数据空间中的低维非线性的数据结构。它常从数据的局部特征入手,再扩展到整体,来发现高维数据的内蕴结构特征。流形学习方法的经典代表有等距映射(isometric mapping, ISOMAP)[51]、局部线性嵌入(locally linear embedding, LLE)[52]和拉普拉斯特征映射(Laplacian eigenmap, LE)[53]等。2000 年,ISOMAP 和 LLE 被《Science》同期发表,引领了流