

基于流形学习 ISOP 算法的语音特征提取及应用研究

赵媛媛, 王力

(贵州大学 计算机科学与信息学院, 贵州 贵阳 550025)

摘要: 主要研究了基于流形学习 ISOP 算法的语音特征提取。将流形学习 ISOP 算法应用到语音识别特征提取模块中。仿真实验结果表明, 该算法与传统的特征提取算法 MFCC、LPCC 等相比, 可以取得较高的识别率。

关键词: ISOP 算法; 语音识别; 流形学习; 特征提取

中图分类号: TP912.3

文献标识码: A

文章编号: 1674-7720(2014)01-0043-03

Research of speech feature extraction and application based on ISOP manifold learning algorithm

Zhao Yuanyuan, Wang Li

(College of Computer Science and Information, Guizhou University, Guiyang 550025, China)

Abstract: The paper prefers to study speech feature extraction based on ISOP manifold learning algorithm. The research attempts to apply ISOP manifold learning algorithm to the model of speech recognition feature extraction. Simulation experiments results show that the proposed algorithm can get a higher recognition rate than the traditional feature extraction algorithm, such as MFCC, LPCC etc.

Key words: ISOP algorithm; speech recognition; manifold learning; feature extraction

通过语音传递信息是人类最重要、最有效、最常用且最方便的交换信息的形式。而信息时代的到来, 使得如何使计算机智能化地与人类进行通信, 成为现代计算机科学的重要研究课题之一。同时, 现实世界中的数据往往是高维的, 难以被理解、表示和处理, 因此对语音信号中的数据进行降维, 找到一组稳定的、能表征其本质特性的特征参数是其中的一个重要步骤。

流形学习^[1-2](Manifold Learning)方法是近十年才发展起来的一种非线性降维方法。流形是拓扑学中的概念, 其表示一个局部处为欧几里得的拓扑空间。局部欧几里得特性意味着对于空间上的任意点都有一个邻域, 在这个邻域中的拓扑与 R^m 空间中的开放单位圆相同 (R^m 表示 m 维欧式空间)。也就是说, 流形是一个局部可坐标化的拓扑空间。基于流形的定义, 可以得到流形学习的本质是, 当数据均匀采样于一个高维欧式空间中的低维流形时, 要从高维采样数据中恢复低维流形的内在几何结构或者内在规律, 并求出相应的嵌入映射, 以实现维数约减或者数据可视化。这就意味着流形学习比传统的

维数约减方法更能体现事物的本质。20 世纪 80 年代末, 在 PAMI 上就已经有流形模式识别的说法。2000 年《Science》杂志上发表的 3 篇论文从认知上讨论了流形学习, 并使用了 Manifold Learning 术语, 强调认知过程的整体性。几年来, 流形学习领域产生了大量的研究成果。LLE 和 ISOMAP^[2]是两种有代表性的非线性降维方法。LLE 算法认为, 在局部意义下数据结构为线性, 即点在一个超平面上。任取一点, 可以用它的邻近点的线性组合来表示。其主要思想是在样本点和它的邻域点之间构造一个重构权向量, 并在低维空间中保持每个邻域中的权值不变, 在嵌入映射是局部线性的条件下, 最小化重构误差。ISOMAP 建立在多维尺度变换 (MDS) 的基础上, 力求保持数据点的内在几何性质, 即保持两点间的测地距离。它用流形上的两点间的测地距离来取代经典 MDS 方法中的欧氏距离, 能够准确地发现数据流形潜在的参数空间, 是一种全局优化算法。

本文将流形学习的降维方式引入语音识别的特征提取模块, 提出一种新的基于流形学习 ISOP 算法的语

音信号特征提取方式。在仿真实验室中,针对0~9的特定人的语音信号进行流形学习的特征提取并进行识别,得到了不错的识别率。

1 流形学习 ISOP 算法

判别等度规映射(Isometric Projection)^[3]是对ISOMAP算法的线性推广,其目的是寻求一个投影矩阵 A ,将高维数据样本映射到低维特征空间中,且在此映射过程中保持样本数据点间的局部结构关系,由此解决了ISOMAP算法在面对新样本的加入无能为力的窘境。

Isometric Projection作为一种维数约减的方法^[4],直接在原始数据中进行构图和多维尺度分析,不仅能渐变地将训练样本和测试样本映射到低维子空间,而且克服了以往算法的缺点,保留了数据间的几何结构。假设有 m 个样本,每个样本的维数是 n ,存在一个数据集 $X=\{x_1, x_2, \dots, x_m\}$,其中 $x_i \in R^n (i=1, \dots, m)$,找到一个投影矩阵 A ,将高维数据样本映射到低维特征空间中,即变换为 $y_i=A^T x_i$,得到 X 的低维嵌入 Y 。

具体步骤如下:

(1)选取邻域。对于样本点 x_i ,选取包括其自身在内的 k 个邻近点作为邻域。确定邻近点的方法主要有 ε 近邻法和 k 近邻法两种。在 ε 近邻法中, $\varepsilon \in R$,如果 x_i 和 x_j 满足 $|x_i - x_j| < \varepsilon$,即两点是相近邻的。在 k 近邻法中,计算每个样本 x_i 点与其他样本点的距离,当 x_j 是 x_i 最近的 k 个样本点中的一个时,认为它们是相邻的。

(2)进行映射。设嵌入函数满足 $\min_f \sum_{i,j} (d_M(x_i, x_j) - d(f(x_i), f(x_j)))^2$,对目标函数进行优化^[5],借助MDS技术把测地线距离矩阵转化为原始数据的内积形式。即定义矩阵 $S_y = D_y^2$ 和 $H = I - \frac{1}{M} ee^T$ (其中 D 为距离矩阵, I 为单位矩阵, e 为全一矩阵),矩阵 H 通常被称为中心化矩阵, $\tau(D) = -HSH/2$ 就是转化后的内积形式。令 D_x 为降维后的子空间的欧氏距离矩阵, $\tau(D_y)$ 为对应的内积矩阵,则目标函数转为 $\min_f \|\tau(D_G) - \tau(D_Y)\|_F$,令线性映射函数 $f(x) = a^T x$,其中 $y_i = f(x_i)$, $Y = (y_1, \dots, y_m) = a^T X$,即有 $\tau(D_Y) = Y^T Y = X^T a a^T X$ 。所以映射函数 $f(x)$ 的求解转化为此最优解:

$$a^* = \min_a \|\tau(D_G) - X^T a a^T X\|^2$$

其中, $\|\tau(D_G) - X^T a a^T X\|^2 = \text{tr}(\tau(D_G)\tau(D_G)^T - X^T a a^T X \tau(D_G)^T - \tau(D_G) X^T a a^T X + X^T a a^T X X^T a a^T X)$ 。

这里, a 只是投影方向,而大小没有什么实际含义,因此添加归一化限制条件: $a^T X X^T a = 1$,从而得到:

$$\|\tau(D_G) - X^T a a^T X\|^2 = \arg \max_{a^T X X^T a = 1} a^T X \tau(D_G) X^T a$$

目标向量 $a_i (i=1, \dots, l)$ 是使目标函数达到最小值的一组向量,等价于求如下广义特征解问题:

$$X \tau(D_G) X^T a = \lambda a X X^T a$$

令 $A = [a_1, \dots, a_l]$,最终线性嵌入表示为:

$$x \rightarrow y = A^T x$$

其中, y 是 l 维向量,用来表示高维数据点 x , A 是映射矩阵。

2 基于流形学习 ISOP 算法的语音特征提取

语音识别系统^[6]主要包括预处理、特征提取、模型训练和模式匹配几个部分,其中特征提取过程如图1所示。

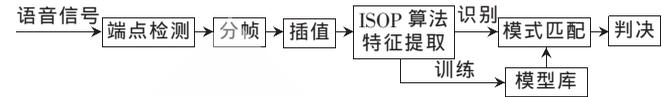


图1 基于LPP算法的语音特征提取

具体步骤如下:

(1)端点检测^[7]。利用双门限检测法对原始语音信号进行检测,正确判定每一个语音信号的起点和终点。

(2)插值。由于流形学习算法ISOP要求数据等长,因此对检测的数据进行插值。具体的插值数量根据实验需求设定,本实验中取值为3072。

(3)加窗分帧后将数据导入ISOP算法,进行特征提取,利用 $X \tau(D_G) X^T a = \lambda X X^T a$,得到特征映射矩阵 A 。

(4)当新的语音信号进入系统进行识别时,按照以上步骤依次处理,通过特征映射矩阵 A 进行特征提取。

3 仿真实验及结果分析

实验在特定人条件下采用相同识别方式,所用的语音信号是在实验室安静环境下录制的0~9共10个数字的语音段,每段语音采样频率为8kHz,16bit量化,PCM编码。具体为采集特定人的每个数字100个。

表1为ISOP算法与传统MFCC/LPCC算法以及其他流形算法在不同训练样本背景下的识别效果的对比。

表1 本文算法与传统特征提取算法识别率对比

训练样本	20	50	90
	识别率/%		
MFCC	92.8	92.6	92.9
LPCC	89.1	88.7	89
ISOP	94	96	96
ISOMAP	22.2	21	20
LTSA	31	30.8	30

从表1可以看到,在特定人语音识别中,本文提出的基于流形学习ISOP算法的特征提取方法^[8]比传统的LPCC/MFCC方法的

识别率都略高;与其他流形算法相比,识别率方面有明显优势。MFCC算法的提出基于人耳的听觉特性,人耳听到的声音的高低与声音的频率并不是线性的关系,用MEL频率尺度更符合人耳的听觉特性。相比之下,LPCC算法基于合成的参数,没有充分利用人耳的听觉特性。大量的研究表明,MFCC性能优于LPCC。

流形算法ISOMAP^[9]作为一种非线性的全局优化学习方法,在构建测地线距离时有两个问题:(1)对样本点的噪声比较敏感,且不能处理存在多个聚类的数据集;(2)测地线距离矩阵的计算复杂度大,且距离矩阵为稠密矩阵,本征分解需要的计算复杂度也比较高,因此在大容量的语音识别中,并不能取得很好的识别率。LTSA^[10]算法虽然能很好地探测出低维流形的空间结构,并且不

要求原始数据是凸分布的,但是该算法所反映的局部结构是它的局部 d 维坐标系,由于噪声等因素的影响,数据集的局部低维特征不明显时,它的局部邻域到局部切空间的投影距离往往并不小,由此构造的重建误差也不会小,这样的情况下,LTSA 就无法得到理想的嵌入结果。此外,LTSA 算法对样本点的密度和曲率变化比较敏感,样本点的密度及曲率^[11]的变化会使得样本点到流形局部切空间的投影产生偏差。当样本量较大时,算法会失效,并且 LTSA 算法对新样本无法进行有效处理。

表 2 是在相同的训练样本量的背景下,不同的邻近值对 ISOP 算法的识别率的影响,从数据分析得出,不同的 k 对识别率有一定影响,但并没有使识别率产生大的偏差。

表 2 参数取值对算法识别率的影响对比

邻近值 k	识别率/%
5	96
7	95.6
13	93

综合以上分析,在实验室条件下,本文提出的基于流形学习 ISOP^[12]算法在语音识别的特征提取模块得到了应用,并取得了一定成果。

本文提出了一种基于流形学习的语音特征提取方法,实验结果表明,在数字 0~9 的识别中,与传统的特征提取算法相比,该方法取得了较高的识别效果。由于 Isometric Projection 是直接原始数据中进行构图和多维尺度分析,因此牵涉大量高位矩阵运算而耗费了较多的时间和资源,为进一步提高性能,下一步将考虑引入主成分分析,对该算法进行改进。

流形学习^[13-14]作为一种新的机器学习,尤其在人脸识别中已取得一定成就的背景下,在语音识别领域中必将成为一个重要的发展方向。

参考文献

- [1] 鲁春元. 流形学习的统一框架及其在模式识别中的应用[D]. 广州:中山大学,2009.
- [2] 李春光. 流形学习及其在模式识别中的应用[D]. 北京:北京邮电大学,2007.
- [3] 邵艳玲,葛玻,宋书中. 基于判别等度规映射的人脸识

别[D]. 洛阳:河南科技大学,2012.

- [4] 王泽杰. 两类非线性降维流形学习算法的比较分析[D]. 上海:上海工程技术大学,2008.
- [5] 曾宪华,罗四维. 全局保持的流形学习算法对比研究[D]. 重庆:重庆邮电大学,北京:北京交通大学,2010.
- [6] 车士伟,吾守尔·斯拉木. 浅谈连续语音识别中的关键技术[D]. 乌鲁木齐:新疆大学,2010.
- [7] 蔡莲红,黄德智,蔡锐. 现代语音技术基础与应用[M]. 北京:清华大学出版社,2003.
- [8] 谭璐. 高维数据的降维理论及应用[D]. 长沙:国防科学技术大学,2005.
- [9] He Xiaofe, Yan Shuncheng, Hu Yuxia, et al. Face recognition using Laplacianfaces[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005,27(3): 328-340.
- [10] BREGLER C, OMOHUNDRO S M. Nonlinear manifold learning for visual speech recognition[C]. International Conference of Computer Vision, 1995:20-23.
- [11] SEUNG H S, LEE D D. The manifold ways of perception[J]. Science, 2000, 290(5500): 2268-2269.
- [12] AGRAFIOTIS D K, XU H. A self-recognizing principle for learning nonlinear manifolds[C]. Proceedings of Natl. Acad. Sci., 1999:15869-15872.
- [13] TENENBAUM J B, SILVA V, LANGFORD J C. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000, 290(5500): 2319-2323.
- [14] He Xiaofe, NIYOGI P. Locality Preserving Projections[C]. Proceedings of 16th Conference on Neural Information Processing Systems, 2003.

(收稿日期:2013-09-06)

作者简介:

赵媛媛,女,1988年生,硕士研究生,主要研究方向:模式识别和语音图像处理。