基于一种改进的监督流形学习算法的语音情感识别

张石清[①③] 李乐民[①] 赵知劲[②]
[①](电子科技大学通信与信息工程学院 成都 610054)
[②](杭州电子科技大学通信工程学院 杭州 310018)
[③](台州学院物理与电子工程学院 台州 318000)

摘 要：为了有效提高语音情感识别的性能，需要对嵌入在高维声学特征空间的非线性流形上的语音特征数据作非线性降维处理。监督局部线性嵌入(SLLE)是一种典型的用于非线性降维的监督流形学习算法。该文针对 SLLE 存在的缺陷，提出一种能够增强低维嵌入数据的判别力，具备最优泛化能力的改进 SLLE 算法。利用该算法对包含韵律和音质特征的 48 维语音情感特征数据进行非线性降维，提取低维嵌入判别特征用于生气、高兴、悲伤和中性 4 类情感的识别。在自然情感语音数据库的实验结果表明，该算法仅利用较少的 9 维嵌入特征就取得了 90.78%的最高正确识别率，比 SLLE 提高了 15.65%。可见，该算法用于语音情感特征数据的非线性降维，可以较好地改善语音情感识别结果。

关键词：语音情感识别；非线性降维；流形学习；监督局部线性嵌入
中图分类号：TN912.34 **文献标识码**： A **文章编号**：1009-5896(2010)11-2724-06
DOI: 10.3724/SP.J.1146.2009.01430

Speech Emotion Recognition Based on an Improved Supervised Manifold Learning Algorithm

Zhang Shi-qing[①③] Li Le-min[①] Zhao Zhi-jin[②]
[①](*School of Communication and Information Engineering*, *University of Electronic Science and Technology of China*, *Chengdu* 610054, *China*)
[②](*School of Telecommunication*, *Hangzhou Dianzi University*, *Hangzhou* 310018, *China*)
[③](*School of Physics and Electronic Engineering, Taizhou University*, *Taizhou* 318000, *China*)

Abstract: To improve effectively the performance on speech emotion recognition, it is needed to perform nonlinear dimensionality reduction for speech feature data lying on a nonlinear manifold embedded in high-dimensional acoustic space. Supervised Locally Linear Embedding (SLLE) is a typical supervised manifold learning algorithm for nonlinear dimensionality reduction. Considering the existing drawbacks of SLLE, this paper proposes an improved version of SLLE, which enhances the discriminating power of low-dimensional embedded data and possesses the optimal generalization ability. The proposed algorithm is used to conduct nonlinear dimensionality reduction for 48-dimensional speech emotional feature data including prosody and voice quality features, and extract low-dimensional embedded discriminating features so as to recognize four emotions including anger, joy, sadness and neutral. Experimental results on the natural speech emotional database demonstrate that the proposed algorithm obtains the highest accuracy of 90.78% with only less 9 embedded features, making 15.65% improvement over SLLE. Therefore, the proposed algorithm can significantly improve speech emotion recognition results when applied for reducing dimensionality of speech emotional feature data.

Key words: Speech emotion recognition; Nonlinear dimensionality reduction; Manifold learning; Supervised locally linear embedding

1 引言

情感计算，作为当前人工智能、信号处理等领域研究的一个新的热点课题，目的就是要赋予计算机类似于人一样的观察、理解和生成各种情感特征的能力，最终使计算机像人一样能进行自然、亲切和生动的交互[1]。语音作为人类最重要的交流媒介之一，携带着说话者丰富的情感信息。因此，从语音信号中分析和提取情感特征，让计算机自动识别出

2009-11-06 收到，2010-04-13 改回
国家自然科学基金(60872092)资助课题
通信作者：张石清 tzczsq@163.com

说话人的情感状态方面的研究就显得尤为重要。该研究在新型人机交互[2]、电话客服中心[3]、智能机器人[4]等领域具有重要的应用价值。

近年来，文献[5,6]发现语音信号中的特征数据位于一个嵌入在高维声学特征空间的非线性流形上。这使得以寻求蕴含在高维数据集中的内在结构信息为目标的流形学习算法开始得以应用于语音特征数据的非线性降维处理，如语音低维可视化[6]和语音识别[7]。降维的主要目标是获取最优的低维嵌入判别特征，丢弃无关或次要的信息，减小数据的维数。其中，用于非线性降维的两种代表性流形学习算法是局部线性嵌入[8](Local Linear Embedding, LLE)和等距映射[9](Isometric Mapping, Isomap)。尽管这两种流形学习算法能够有效实现语音数据的低维可视化，但用于语音识别时表现不佳，甚至不如传统的线性主成分分析法[10](Principal Component Analysis, PCA)。主要原因是这两种流形学习算法都属于非监督方式的降维，没有考虑对分类有帮助的已有数据点之间的类别信息。

为了克服非监督流形学习算法模式识别方面的不足，Ridder 等人[11]通过使用考虑数据类别信息的监督距离修改 LLE 算法中的邻域点搜索，提出了一种代表性的监督式的局部线性嵌入(Supervised Locally Linear Embedding, SLLE)算法。SLLE 算法具有较好的模式识别性质，已经广泛应用表情识别[12]、人脸识别[13]等领域。然而，这种 SLLE 算法仍然有 3 个缺陷。(1)SLLE 采用的监督距离是线性的。这会导致所希望的数据点之间的类间距增大时，其类内距也同样保持同步增大，从而削弱了 SLLE 产生的低维嵌入数据的判别力，不利于数据的分类。(2)SLLE 算法以批处理方式运行处理已有的训练样本数据，不能有效解决新测试样本数据的泛化问题，因为 SLLE 从已有的训练样本提取的低维嵌入判别数据对新测试样本数据的输入不能直接给出合理的嵌入输出，即所谓的泛化能力的缺失。(3)在构成 SLLE 算法中的监督距离中的常数因子对 SLLE 的泛化性能有着极其重要的影响。然而，对如何最优化 SLLE 距离中的常数因子，已有的研究[11−13]大多采用在某一特定目标维度(如本征维度)上执行繁琐而重复的人工搜索试验而取得该常数因子的经验最优值，然后固定其值不变，在不同目标维度上都进行降维使用，但实际上该常数因子的最优值很容易受到不同的降维目标维度的影响。因此在某一目标维度上取得的常数因子的最优值对于其它降维的目标维度并不是最优的。为此，本文提出采用一种能增强低维嵌入数据的判别力的非线性监督距离替代 SLLE 中的线性监督距离，并发展一种在不同维度上能够自动最优化常数因子的算法，进而构造出具有最优泛化能力的改进 SLLE 算法，简称 Improved-SLLE(Improved Supervised Locally Linear Embedding, Improved-SLLE)。利用 Improved-SLLE 对较高维度的语音情感特征参数进行非线性降维，提取判别力增强的低维嵌入特征，从而在低维嵌入特征空间实现语音情感识别结果的改善。在建立的自然情感语音数据库的试验结果表明了该算法的有效性。

2　Improved-SLLE 算法

2.1　算法步骤

设 D 维的 N 个输入数据点为 X_i ($X_i \in R^D$, $i \in [1,N]$)，类别号为 L_i，嵌入输出的低维($d, d \le D$) N 个数据点为 Y_i ($Y_i \in R^d, i \in [1,N]$)。

Improved-SLLE 算法可以分为 3 个步骤：

(1)通过计算数据点之间的非线性监督距离，寻找每个数据点的 k 个邻域点。

(2)由每个数据点的邻域点计算出该数据点的局部重建权值矩阵。

计算最优的重构权值矩阵 $\boldsymbol{W}_{ij}$ 时，需要最小化代价函数：

$$\varepsilon(\boldsymbol{W}) = \sum_{i=1}^{N} \left\| X_i - \sum_{j=1}^{N} \boldsymbol{W}_{ij} X_j \right\|^2 \tag{1}$$

要求 $\sum_{j=1}^{N} \boldsymbol{W}_{ij} = 1$；$X_j$ 不是 X_i 的邻域点时，$\boldsymbol{W}_{ij} = 0$。

(3)由该数据点的局部重建权值矩阵和其邻域点计算出该数据点的输出值。

求数据点在低维空间中的嵌入映像使得低维重构误差最小，需要最小化代价函数为

$$\phi(\boldsymbol{Y}) = \sum_{i=1}^{N} \left\| Y_i - \sum_{j=1}^{N} \boldsymbol{W}_{ij} Y_j \right\|^2 = \mathrm{tr}(\boldsymbol{Y}\boldsymbol{M}\boldsymbol{Y}^{\mathrm{T}}) \tag{2}$$

其中 $\boldsymbol{M} = (\boldsymbol{I} - \boldsymbol{W})^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{W})$，输出 Y 由矩阵 $\boldsymbol{M}$ 的 d 个最小的非零特征值所对应的特征向量组成。

在上述 Improved-SLLE 算法中的第(1)步，计算非线性监督距离的公式如下：

$$\Delta' = \begin{cases} \sqrt{1 - e^{-\Delta^2/\beta}}, & L_i = L_j \\ \sqrt{e^{\Delta^2/\beta}} - \alpha, & L_i \ne L_j \end{cases} \tag{3}$$

其中 Δ' 是结合数据点类别信息计算后的距离，Δ 是忽略数据点类别信息的原始欧氏距离。参数 β 用来防止指数函数中的 Δ 增长过快，尤其当 Δ 本身就相对比较大，β 作用更明显。因此，参数 β 与数据集中的数据密集程度密切相关，一般取所有成对数据

点的欧氏距离的平均值。而参数 $\alpha\,(0 \le \alpha \le 1)$ 是一个常数因子，用来控制不同类别数据点的距离，在某种概率上接近或者小于同种类别数据点距离的程度。

与 Improved-SLLE 相比，原始的 SLLE 算法在计算点与点之间的距离时，采用的线性监督距离公式如下：

$$\Delta' = \begin{cases} \Delta, & L_i = L_j \\ \Delta + \alpha \max(\Delta), & L_i \ne L_j \end{cases} \tag{4}$$

其中 $\max(\Delta)$ 是表示最大欧氏距离，而常数因子 $\alpha\,(0 \le \alpha \le 1)$ 也是用来控制距离计算时数据点类别信息的结合数量程度。

为了更好理解Improved-SLLE算法中采用的非线性监督距离的优越性，图 1 举例说明了，当原始欧氏距离 Δ 在指定区间[0,3]呈线性规律增长时，比较了 Improved-SLLE 和 SLLE 两种算法的距离曲线的不同变化特点。这两种算法的常数因子 α 都设为 0.3。

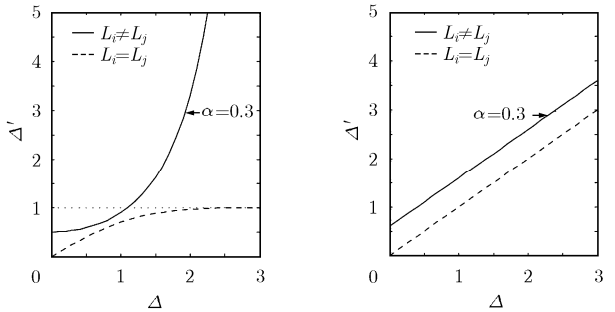

图 1 两种不同算法的距离曲线比较

由图 1 (a) 可见，Improved-SLLE 采用了非线性监督距离，不同类别数据点的类间距呈指数快速增长，而同种类别数据点的类内距被控制在[0,1]缓慢增长。这样会使得类间距尽可能增大，而类内距尽可能小，保持在区间[0,1]内。类间距和类内距之间的比值随着距离的增大而增大，使得低维嵌入数据的判别力也会随着距离的增大而增强。这种特点非常有利于嵌入数据的分类。相反，SLLE 算法由于采用线性监督距离，则没有这种好的性质。如图 1 (b)所示，当类内距快速增长时，类间距也同样保持同步增长，使得类间距和类内距之间的比值不变。这会削弱低维嵌入数据的判别力，不利于数据的分类。

为了使得Improved-SLLE对新测试样本具有较好的泛化能力，本文采用 Nystrom 方法[14]构建 Improved-SLLE 的泛化算法。

2.2 Improved-SLLE 常数因子的自动最优化算法

Improved-SLLE 距离中常数因子 α 的最优值，应该由Improved-SLLE对测试样本获得的最优泛化性能来决定。训练样本一般是预先已有的，而测试样本是未知的，如实时性处理。因此，为了获取常数因子 α 的最优值，可以把已有的训练样本拆分成两部分：一部分用于训练，一部分用于测试。这样就可以根据从训练样本拆分出来的测试样本的最低识别错误率来选择最优的常数因子 α，具体算法步骤如表 1 所示。对 SLLE 常数因子 α 的最优化，也可采用表 1 的算法实现。

表 1 Improved-SLLE 常数因子的自动最优化算法

(1)输入：训练样本数据 X，降维维度最大值 $d_{\max}$ ($d_{\max} \le D$)，初始值 $\alpha = 0$，初始错误率 error-rate=1,初始降维维度 $d = 2$，最优值 α 记为 bestα。

(2)拆分训练样本数据 X：50%用于训练，训练数据记为 X_1；50%用于测试，测试数据记为 X_2。

(3)执行以下循环程序：

for d to $d_{\max}$, $d = 2, 3, \cdots, d_{\max}$

for α to 1, $\alpha = 0, 0.1, 0.2, \cdots, 1$

(a)利用 Improved-SLLE 计算出 X_1 的 d 维嵌入数据 X_{1d}；

(b)利用 Improved-SLLE 泛化算法计算出 X_2 的 d 维嵌入数据 X_{2d}；

(c)采用最简单的 K 最近邻分类器 KNN(K-Nearest-Neighbor, KNN)[15](K=1)对 X_{1d} 训练，对 X_{2d} 测试，计算出 d 维嵌入数据的识别错误率 new-error，并将 new-error 与 error-rate 比较。

if new-error < error-rate

error-rate = new-error, bestα = α

(d)输出每一维 d 对应的最优常数因子 bestα，以及 error-rate。

end for

end for

3 语音情感识别实验研究

为了检验 Improved-SLLE 的语音情感识别性能，将采用 PCA，LLE，Isomap，SLLE 和 Improved-SLLE 分别应用于提取的语音情感特征数据的降维，然后比较这 5 种方法在不同维度上降维后的语音情感识别结果。

3.1 自然情感语音数据库

目前，国内外研究者大多采用人工模仿的模拟情感语音数据库进行语音情感识别的研究，但这种模拟数据库中的语音的情感自然度跟现实真实情感还有差距，备受质疑。因而，对人类现实生活中真实情感语音的识别研究更接近实际，更有意义。为此，本文建立了一个包含说话人自然情感的汉语语

音数据库。我们通过从 20 个电视访谈对话节目视频素材中建立一个与说话者无关的自然度较高的 800 句大小的汉语情感语音数据库。每一个访谈对话节目中，有至少 2 个人自然和随意性地讨论一些当代的典型社会现象，家庭冲突或感人事迹等话题。这些人物谈论话题时，一般预先没有讲稿，自发地进行即兴讨论。因此，这些人物讨论时情感的表达是非常真实的。由于讨论话题范围的限制，生气、高兴、悲伤和中性 4 类常见的情感类型的视频片段数量较多。采用专业音频编辑软件 CoolEdit(http://www.mp3-converter.com/cool_edit_2000.htm)从整个视频中提取语音文件，剔除其中人物情感表达不明显，背景音乐等杂音较多的语音片段，然后从较理想的语音片段中切分出不同人物在某一短时间内完整的一小句情感语音,作为情感分析语料。最终提取到 53 人(女性 37，男性 16)的采样率为 16 kHz，16 bit 的单声道 WAV 格式的 800 句数据库。其中，生气、高兴、悲伤和中性各 200 句。最后请 4 个听众随机听取测试，对于情感特征不是很明显的语句进行了删除和重新提取。

3.2 语音情感特征参数提取

目前，研究者普遍发现与说话人发音时密切相关的情感特征参数，主要包括发音语调和轻重相关的基音频率、振幅(或能量)、发音持续时间等韵律特征[3,16]，以及发声方式相关的共振峰、频谱能量分布，谐波噪声比等音质特征[17,18]。因此，本文对自然情感语音数据库的每一句语音，提取能够表达情感信息的韵律特征和音质特征参数，共 48 个，如表 2 所示。

3.3 实验测试及结果分析

实验时，全部特征参数数据归一化到[0,1]，分类器采用 KNN 分类器。KNN 是一种基于样本学习的传统无参数分类器，运算快，采用一个最近邻训练模式(K=1)时用于语音情感识别的性能比较好[3]。

为了提高识别结果的可信度，识别中采用 10 次交叉检验技术。即所有语句被平分为 10 份，每次使用其中的 9 份数据用于训练，剩下的 1 份数据用于测试。这样的识别实验相应重复 10 次，最后取 10 次的平均值作为识别结果。每次交叉检验时， 对 Improved-SLLE 和 SLLE 的常数因子 α 的自动最优化算法也相应执行一次。

实验 1　对提取的原始 48 维语音特征数据不作任何降维处理，直接进行情感识别实验，识别结果如表 3 所示。

由表 3 可得，生气和中性的识别结果较为令人满意，正确识别率分别达到了 83%和 86%。4 类情感的总体平均正确识别率为 75.13%。但高兴和悲伤的正确识别率略低，其中高兴为 70%，悲伤为 61.5%。主要原因是，高兴和生气发音时的韵律特征相似，而悲伤和生气发音时的音质特征相似，从而导致高兴与生气、悲伤与生气这两对情感相互之间较易混淆[17]。

实验 2　采用 PCA，LLE，Isomap，SLLE 和 Impoved-SLLE 及相应的泛化算法分别应用于原始 48 维语音特征数据的降维，然后对降维后的低维判别特征数据进行情感识别测试，并比较识别结果。LLE，Isomap 和 SLLE 的泛化算法的实现类似于 Impoved-SLLE。而 PCA 的泛化算法则可以直接通

表 2 语音情感特征参数

特征类型	特征组	统计
韵律特征	基因频率	(1)最大值 $f_{\max}$ (2)最小值 $f_{\min}$ (3)变化范围 f_d=$f_{\max}$−$f_{\min}$ (4)上四分位数 $f_{0.75}$ (5)中位数 $f_{0.5}$ (6)下四分位数 $f_{0.25}$ (7)内四分极值 f_i= $f_{0.75}$−$f_{0.25}$ (8)平均值 f_m (9)标准差 σ_p (10)平均绝对斜度 M_s
	振幅	(11)平均值 A_m (12)标准差 σ_p (13)最大值 $A_{\max}$ (14)最小值 $A_{\min}$ (15)变化范围 A_d=$A_{\max}$−$A_{\min}$ (16)上四分位数 $A_{0.75}$ (17)中位数 $A_{0.5}$ (18)下四分位数 $A_{0.25}$ (19)内四分极值 A_i
	发音持续时间	(20)发音持续总时间 T_s (21)有声发音持续时间 T_v (22)无声发音持续时间 T_u (23)有声与无声时间的比值 T_{vu}=T_v/T_u (24)有声与发音总时间的比值 T_{vs}=T_v/T_s (25)无声与发音总时间的比值 T_{us}=T_u/T_s
音质特征	共振峰 F_1-F_3	(26) F_1平均值 (27)F_2平均值 (28)F_3平均值 (29)F_1标准差 (30)F_2标准差 (31)F_3标准差 (32) F_1中位数(33) F_2中位数 (34)F_3中位数 (35)F_1中位数所占带宽 (36)F_2中位数所占带宽 (37)F_3中位数所占带宽
	频带能量分布	(38)0−500 Hz 的频带能量平均值 SED_{500} (39)500−1000 Hz 的频带能量平均值 SED_{1000} (40)2500−4000 Hz 的频带能量平均值 SED_{4000} (41)4000−5000 Hz 的频带能量平均值 SED_{5000}
	谐波噪声比	(42)最大值 $H_{\max}$ (43)最小值 $H_{\min}$ (44)变化范围 H_d=$H_{\max}$−$H_{\min}$ (45)平均值 H_m (46)标准差 σ_H
	短时抖动参数	(47)短时基频抖动参数 Jitter (48)短时能量抖动参数 Shimmer

表 3 原始特征数据不降维时的语音情感识别结果

情感类型	生气	高兴	悲伤	中性
生气	166	24	10	0
高兴	36	140	20	4
悲伤	16	40	123	21
中性	0	20	8	172

过从训练样本得到的线性映射矩阵与新测试样本相乘得到。降维的目标维度范围取 $2 \le d \le 20$。LLE，Isomap，SLLE 和 Impoved-SLLE 的近邻数取 $k = 12$ 时的性能较好[5]。每次交叉检验时，每一维对应的 Improved-SLLE 和 SLLE 常数因子 α 的最优值，采用常数因子 α 的自动最优化算法取得。表 4 和表 5 分别列出了 10 次交叉检验中所取得的每一维对应的 Improved-SLLE 和 SLLE 常数因子 α 最优化的平均值。由表 4 和表 5 可见，不同维度的 Improved-SLLE 常数因子 α 的最优化平均值大小一般不超过 0.5，比较稳定，而 SLLE 常数因子 α 的最优化平均值大小变化明显。主要原因是 Improved-SLLE 采用的非线性监督距离中存在另一参数 β。该 β 的取值是所有成对数据点欧氏距离的平均值，因而 β 能对 α 的变化起到一定的平衡作用。

图 2 给出了 5 种不同降维方法取得的每一维的语音情感识别结果。表 6 列出了在不同维度上各种方法取得的最好性能的比较。其中，“Original”方法表示对原始 48 维特征数据不作任何降维所取得的识别结果(见表 3)。

由图 2 和表 6 的结果，可以得知：(1)与其它方法相比，Improved-SLLE 经过泛化和常数因子 α 最优化后，取得了最好的情感识别性能。Improved-SLLE 仅利用较少的 9 维嵌入特征就取得了 90.78% 的最高正确识别率，比 Original，PCA，LLE，Isomap 和 SLLE 5 种方法分别高出了 15.65%，18.28%，26.13%，22.01%和 10.03%。原因是 Improved-SLLE 算法中采用了非常有利于嵌入数据分类的非线性监督距离，使得产生的低维嵌入特征数据具有最好的泛化能力和判别力。(2)SLLE 经过泛化和常数因子 α 最优化后，取得的识别性能高于 PCA，LLE 和 Isomap。作为一种监督降维方法，所以 SLLE 能够比非监督的 PCA，LLE 和 Isomap 3 种降维方法性能更好。值得指出的是，当直接利用没有经过泛化的 SLLE 算法进行情感识别时，识别效果很差，只能取得 22.45%的最高正确识别率。(3)与 LLE 和 Isomap 相比，PCA 取得了更好的识别结果。这说明位于非线性流形上的语音特征数据的非线性程度并不是很高，使得线性 PCA 方法仍然可以提取到比非线性方法 LLE 和 Isomap 具有更强判别力的低维嵌入特征数据。另外一个原因是，LLE 和 Isomap 都属于非监督方式的降维，不能有效发挥出它们的最佳性能。(4)对于 LLE 和 Isomap 两种方法，Isomap 比 LLE 表现更好。Isomap 是一种全局降维方法，嵌入时保留数据点的全局结构信息。而 LLE 是一种局部降维方法，嵌入时只保留数据点的局部结构信息。实验表明嵌入时保留全局信息比局部信息更有效。(5)所有降维算法的识别性能刚开始随着维度的增加而显著提高，但当维度更高时，他们的性能反

表 4 不同维度的 Improved-SLLE 常数因子 α 的最优化平均值

d	2	3	4	5	6	7	8	9	10	11
α	0.16	0.27	0.33	0.42	0.25	0.36	0.21	0.59	0.12	0.24
错误率(%)	52.95	48.18	44.32	35.23	35.68	33.86	31.23	30.85	27.50	28.06
d	12	13	14	15	16	17	18	19	20	
α	0.26	0.23	0.14	0.38	0.22	0.12	0.36	0.41	0.23	
错误率(%)	29.32	30.36	32.59	34.32	35.45	35.77	36.52	36.09	37.82	

表 5 不同维度的 SLLE 常数因子 α 的最优化平均值

d	2	3	4	5	6	7	8	9	10	11
α	0.62	0.35	0.41	0.72	0.44	0.76	0.53	0.35	0.78	0.24
错误率(%)	54.09	46.04	43.86	38.63	35.82	34.14	33.50	29.54	30.90	28.61
d	12	13	14	15	16	17	18	19	20	
α	0.55	0.14	0.28	0.67	0.21	0.37	0.45	0.62	0.42	
错误率(%)	31.80	32.90	34.09	38.60	40.90	38.33	39.71	38.60	42.04	

表 6 不同方法取得的最好性能比较

方法	Original	PCA	LLE	Isomap	SLLE	Improved-SLLE
d	48	12	19	12	11	9
正确识别率(%)	75.13	72.50	64.65	68.77	80.75	90.78

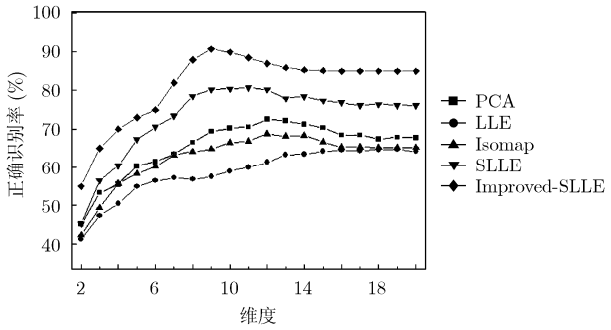

图 2 不同降维算法取得的语音情感识别结果

而会有所下降，最终趋向于稳定。这归咎于嵌入在 48 维声学特征空间数据的内在结构信息的本征维数刚好位于维度范围[2, 20]之间。这也说明实验时降维的维度范围取[2, 20]是合理的。

4 结束语

本文在克服 SLLE 算法的不足基础上提出了一种改进的监督局部线性嵌入算法 Improved-SLLE，同时发展了 Improved-SLLE 的泛化算法和距离中常数因子的自动最优化算法。利用 Improved-SLLE 实现对 48 维语音情感特征参数数据的非线性降维，提取相应的低维嵌入判别特征进行语音情感识别，取得了 90.78%的正确识别率。与其它使用方法相比，Improved-SLLE 的识别性能最好。当前，语音情感识别的研究还处于初级阶段，发展比 Improved-SLLE 更强的新型监督流形算法，对以后的语音情感识别的研究具有重要意义。

参 考 文 献

[1] Picard R. Affective Computing[M]. MIT Press, Cambridge, MA, 1997: 1–24.

[2] Jones C and Deeming A. Affective human-robotic interaction[C]. Affect and Emotion in Human-Computer Interaction, Springer, 2008, Lecture Notes in Computer Science, 4868: 175–185.

[3] Morrison D, Wang R, and De Silva L C. Ensemble methods for spoken emotion recognition in call-centres[J]. *Speech Communication*, 2007, 49(2): 98–112.

[4] Picard R. Robots with emotional intelligence[C]. 4th ACM/IEEE international conference on Human robot interaction, California, 2009: 5–6.

[5] Errity A and McKenna J. An investigation of manifold learning for speech analysis[C]. 9th International Conference on Spoken Language Processing (ICSLP'06), Pittsburgh, PA, USA, 2006: 2506–2509.

[6] Goddard J, Schlotthauer G, and Torres M, *et al*.. Dimensionality reduction for visualization of normal and pathological speech data[J]. *Biomedical Signal Processing and Control*, 2009, 4(3): 194–201.

[7] Yu D. The application of manifold based visual speech units for visual speech recognition[D]. [Ph.D.dissertation], Dublin City University, 2008.

[8] Roweis S T and Saul L K. Nonlinear dimensionality reduction by locally linear embedding[J]. *Science*, 2000, 290(5500): 2323–2326.

[9] Tenenbaum J B, Silva Vd, and Langford J C. A global geometric framework for nonlinear dimensionality reduction[J]. *Science*, 2000, 290(5500): 2319–2323.

[10] Jolliffe I T. Principal Component Analysis[M]. New York: Springer, 2002: 150–165.

[11] De Ridder D, Kouropteva O, and Okun O, *et al*.. Supervised locally linear embedding[C]. Artificial Neural Networks and Neural Information Processing-ICANN/ICONIP-2003, Springer, 2003, Lecture Notes in Computer Science, 2714, 333–341.

[12] Liang D, Yang J, and Zheng Z, *et al*.. A facial expression recognition system based on supervised locally linear embedding[J]. *Pattern Recognition Letters*, 2005, 26(15): 2374–2389.

[13] Pang Y, Teoh A, and Wong E, *et al*.. Supervised Locally Linear Embedding in face recognition[C]. International Symposium on Biometrics and Security Technologies, Islamabad, 2008: 1–6.

[14] Platt J C. Fastmap, MetricMap, and Landmark MDS are all Nystrom algorithms[C]. 10th International Workshop on Artificial Intelligence and Statistics, Barbados, 2005: 261–268.

[15] Aha D, Kibler D, and Albert M. Instance-based learning algorithms[J]. *Machine Learning*, 1991, 6(1): 37–66.

[16] 赵力, 将春辉, 邹采荣等. 语音信号中的情感特征分析和识别的研究[J]. 电子学报, 2004, 32(4): 606–609.
Zhao Li, Jiang Chun-hui, and Zou Cai-rong, *et al*.. A study on emotional feature analysis and recognition in speech[J]. *Acta Electronica Sinica*, 2004, 32(4): 606–609.

[17] Zhang S. Emotion recognition in Chinese natural speech by combining prosody and voice quality features[C]. Advances in Neural Networks - ISNN 2008, Springer, 2008, Lecture Notes in Computer Science, 5264: 457–464.

[18] Zhao Y, Zhao L, and Zou C, *et al*.. Speech emotion recognition using modified quadratic discrimination function[J]. *Journal of Electronics* (*China*), 2008, 25(6): 840–844.

张石清：　男，1980 年生，博士生，研究方向为语音信号处理与情感计算.

李乐民：　男，1932 年生，教授，博士生导师，中国工程院院士，主要研究方向为信息与通信工程.

赵知劲：　女，1959 年生，教授，博士生导师，主要研究方向为信息与通信工程.