基于流形正则化极限学习机的语种识别系统

徐嘉明 [1,2] 张卫强 [3] 杨登舟 [1,2] 刘 加 [3] 夏善红 [2]

摘 要 支持向量机 (Support vector machine, SVM) 在语种识别中已经起到了重要的作用. 近些年来, 极限学习机 (Extreme learning machine, ELM) 在很多领域取得了成功的应用. 相比于 SVM, ELM 最大的优点在于极易实现、训练速度快, 而且通常可以取得与 SVM 相近甚至优于 SVM 的识别性能. 鉴于 ELM 这些优异的特点, 本文将 ELM 引入到语种识别中, 并针对 ELM 由于随机初始化模型参数所带来的潜在问题, 提出了流形正则化极限学习机 (Manifold regularized extreme learning machine, MRELM) 算法. 实验结果表明, 在高斯超矢量 (Gaussian supervector, GSV) 特征空间上, 相对于 SVM 基线系统, 该算法对 30 秒语音的识别性能有明显的提升. 同时该算法也可以成功地应用到 i-vector 特征空间中, 取得与当前主流的打分算法相近的识别性能.

关键词 语种识别, 极限学习机, 流形学习, 支持向量机

引用格式 徐嘉明, 张卫强, 杨登舟, 刘加, 夏善红. 基于流形正则化极限学习机的语种识别系统. 自动化学报, 2015, **41**(9): 1680−1685

DOI 10.16383/j.aas.2015.c140916

Manifold Regularized Extreme Learning Machine for Language Recognition

XU Jia-Ming[1,2] ZHANG Wei-Qiang[3]
YANG Deng-Zhou[1,2] LIU Jia[3] XIA Shan-Hong[2]

Abstract Support vector machines (SVMs) have played an important role in the state-of-the-art language recognition systems. The recently developed extreme learning machine (ELM) which has been successfully applied to many areas tends to achieve much better generalization performance than the traditional SVM. Inspired by the excellent features of ELM, we introduce it into language recognition and propose a manifold regularized extreme learning machine (MRELM) to overcome the potential problem of ELM due to random initialization of model parameters. Experimental results show that the proposed algorithm can achieve much better performance than SVM at 30 s durations in the Gaussian supervector (GSV) feature space. In addition, MRELM can be applied to the i-vector space and get comparable results to the existing scoring methods.

Key words Language recognition, extreme learning machine (ELM), manifold learning, support vector machine (SVM)

Citation Xu Jia-Ming, Zhang Wei-Qiang, Yang Deng-Zhou, Liu Jia, Xia Shan-Hong. Manifold regularized extreme learning machine for language recognition. *Acta Automatica Sinica*, 2015, **41**(9): 1680−1685

语种识别的任务是自动识别出一段语音所属语言的种类[1]. 在语音识别系统前端[2] 以及多语言自动翻译系统前端[3−4], 语种识别起到了至关重要的作用, 因此, 语种识别已经引起了相关领域研究人员的广泛关注.

在语种识别中, 支持向量机 (Support vector machine, SVM) 已经被证明是一种非常有效的分类器[5−6], 并得到了广泛的应用. SVM 通过核方法将数据从原始特征空间映射到一个更高维甚至无穷维的特征空间中, 在该空间中通过标准的优化算法寻找一个可以最大化分类间隔的分类超平面. 在语种识别中, 我们通常采用一对其他 (One-vs-the-rest) 的训练方法来处理多类问题, 即将目标语种的所有训练数据设为正样本, 并将其他语种的所有训练数据设为负样本. 采用这种方法, 正样本和负样本间的数据量往往不平衡, 会造成性能上的损失. 此外, SVM 的训练过程需要计算核矩阵, 所以 SVM 训练的计算复杂度通常和数据样本数的平方成正比.

近些年来, 极限学习机 (Extreme learning machine, ELM)[7−8] 的出现引起了研究人员越来越多的关注. ELM 是一个单隐含层的前馈神经网络, 它最大的特点是输入层和隐含层之间的连接权值矩阵以及隐含层节点的偏置只需要一次性随机初始化, 而不需要通过传统的反向传播算法进行更新. 唯一需要计算的就是隐含层和输出层之间的权值矩阵, 该矩阵可以通过求广义逆矩阵的方法求得. 相比于 SVM, ELM 具有一些显著的优点. 从计算复杂度的角度来看, 由于隐含节点数通常少于训练样本数, 所以 ELM 的求解更加快速. 从分类性能的角度来看, 近些年来的文献报道表明, 通常情况下, ELM 会取得与 SVM 相近甚至优于 SVM 的分类性能[9−10]. 此外, ELM 还可以进行在线训练, 即可以对训练数据进行分块训练, 每次训练都会在上次训练的基础上更新模型参数[11]. 另外, ELM 还可以和传统的核方法相结合[10]. 正因为具有这些优异的特性, ELM 已经在许多领域取得了成功的应用, 并引起了人们极大的研究兴趣[12−14]. 许多来自不同领域的研究者提出了很多方法来改进 ELM 算法, 例如利用 ELM 解决带噪声数据的问题[15−16] 以及数据量不平衡的问题[17]、基于 Fisher 准则的 ELM 算法[18] 等.

鉴于 ELM 在分类问题中的成功应用, 本文将 ELM 算法引入到语种识别中, 但是, 直接使用 ELM 算法存在潜在的问题. ELM 的输出权值矩阵的求解采用的是广义逆矩阵的求解方法, 其本质上是最小二乘算法. 也就是说, 如果我们将隐含层输出看作一个新的特征空间 (这里称为 ELM 特征空间), ELM 算法是在 ELM 特征空间中求解一组线性分类超平面. 但是, 由于采用随机初始化输入权值矩阵及隐含层偏置, ELM 相当于将数据从原始特征空间随机映射到 ELM 特征空间中, 这种随机映射往往会使数据在 ELM 特征空间中呈现出某种难以预测的非线性分布, 而简单地采用最小二乘的方法无法挖掘数据在 ELM 特征空间中的非线性的几何结构.

为了改善这一潜在问题, 本文尝试将流行学习[19−20] 的思想引入到 ELM 算法中, 在流形正则化的理论框架下, 提出了流形正则化极限学习机算法 (Manifold regularized extreme learning machine, MRELM), 并将其成功应用到语种识别中. 值得一提的是, 文献 [21−22] 提出将流形正则化的理论框架与 ELM 相结合用于解决半监督学习或无监督学

收稿日期 2015-01-05 录用日期 2015-06-07
Manuscript received January 5, 2015; accepted June 7, 2015
国家自然科学基金 (61273268, 61370034, 61403224) 资助
Supported by National Natural Science Foundation of China (61273268, 61370034, 61403224)
本文责任编委 吴玺宏
Recommended by Associate Editor WU Xi-Hong
1. 中国科学院大学 北京 100190 2. 中国科学院电子学研究所传感技术国家重点实验室 北京 100190 3. 清华大学电子工程系 清华信息与科学技术国家实验室 北京 100084
1. University of Chinese Academy of Sciences, Beijing 100190 2. State Key Laboratory of Transducer Technology, Institute of Electronics, Chinese Academy of Sciences, Beijing 100190 3. Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084

习的问题, 而本文研究的是有监督学习的问题, 所以, 本文在训练方法上与之有所不同. 另外, 文献 [23] 中提出将结构风险引入到 ELM 算法中以提高算法的泛化能力, 本文也将遵照这一做法, 将该结构风险视作 ELM 算法优化目标的一部分.

为了验证 MRELM 算法的有效性, 本文采用了当前语种识别系统最常用的两种声学特征, 即高斯超矢量 (Gaussian supervector, GSV)[24] 和 i-vector[25]. GSV 是采用一种基于 KL (Kullback-Leibler) 距离的核函数的方法, 专门针对 SVM 训练而提取的一种特征. i-vector 是基于因子分析的方法提取的一种特征, 近些年来, 由于其优异的性能, 已经成为语种识别和说话人识别领域里广泛采用的特征之一.

1　ELM 算法简介

作为一种极简单且有效的学习算法, ELM 最早被提出用来训练单隐含层前馈神经网络[7], 该网络的输入层和隐含层间的权值矩阵以及隐含层节点的偏置可以随机初始化而不需要再次更新. 所需计算的只有隐含层和输出层间的权值矩阵, 该矩阵可以通过求广义逆矩阵的方法求得.

这里我们设定训练数据为 $\{\boldsymbol{x}_i, \boldsymbol{t}_i\}$, 其中, $\boldsymbol{x}_i$ 是 d 维的训练样本, 训练数据的总样本数为 N. $\boldsymbol{t}_i$ 是样本 $\boldsymbol{x}_i$ 的 m 维标签向量, m 是类别总数, 如果 $\boldsymbol{x}_i$ 属于第 j 类, 则将 $\boldsymbol{t}_i$ 的第 j 维设为 1, 其他维均设为 0. ELM 的整体结构如图 1 所示. 其中, L 是隐含层节点的数目, $\boldsymbol{w}_i$ 是隐含层第 i 个节点与输入层之间的权重向量, b_i 是隐含层第 i 个节点的偏置. $\boldsymbol{w}_i$ 和 b_i 一旦随机初始化后就不再改变. $\boldsymbol{\beta}_i$ 是隐含层第 i 个节点与输出层之间的权值向量. $\boldsymbol{y}$ 是 ELM 的决策层输出向量. 对于第 j 个样本点, 隐含层的输出表示为 $\boldsymbol{z}_j$.

$$\boldsymbol{z}_j = h(\boldsymbol{x}_j) = \begin{bmatrix} g(\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{x}_j + b_1) & \cdots & g(\boldsymbol{w}_L^{\mathrm{T}}\boldsymbol{x}_j + b_L) \end{bmatrix} \tag{1}$$

其中, $g(\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{x}_j + b_1)$ 是激活函数, 最常用的激活函数是 sigmod 函数. 从 ELM 的角度来看, $h(\boldsymbol{x}_j)$ 是一个特征映射, 它将样本 $\boldsymbol{x}_j$ 映射到 ELM 特征空间中, 所以, 这里 $\boldsymbol{z}_j$ 表示的是 ELM 特征空间中的第 j 个样本. 这样, 对所有训练数据而言, 隐含层的输出矩阵可以表示为

$$H = \begin{bmatrix} g(\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{x}_1 + b_1) & \cdots & g(\boldsymbol{w}_L^{\mathrm{T}}\boldsymbol{x}_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{x}_N + b_1) & \cdots & g(\boldsymbol{w}_L^{\mathrm{T}}\boldsymbol{x}_N + b_L) \end{bmatrix}_{N\times L} \tag{2}$$

ELM 的优化目标就是同时最小化训练误差 (经验风险) 以及输出矩阵的二次范数 (结构风险), 即

$$\min: \ \| H\beta - T \|_2^2 + \frac{C}{2} \| \beta \|_2^2 \tag{3}$$

其中, C 是加权系数,

$$\beta = \begin{bmatrix} \boldsymbol{\beta}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{\beta}_L^{\mathrm{T}} \end{bmatrix}_{L\times m}, \quad T = \begin{bmatrix} \boldsymbol{t}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{t}_N^{\mathrm{T}} \end{bmatrix}_{N\times m} \tag{4}$$

上述系统可通过如下求解.

$$\beta = (H^{\mathrm{T}}H + CI)^{-1}H^{\mathrm{T}}T \tag{5}$$

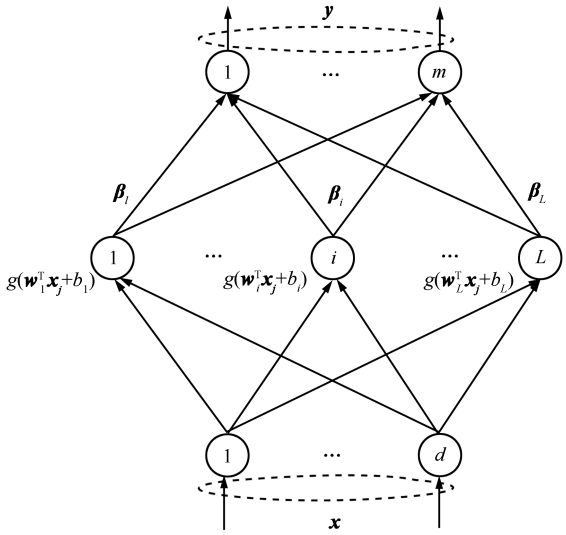

图 1　ELM 结构图

Fig. 1　ELM structure

2　流形正则化 ELM 算法

ELM 算法由于采用一次性随机初始化输入层和隐含层之间的权值以及隐含层节点的偏置, 这往往会使得数据在 ELM 特征空间中的分布具有一定的随机性, 这种随机性往往会使数据的分布出现某种非线性的几何结构. 简单地采用最小化训练误差和最小化输出权值的二次范数作为优化目标无法挖掘数据的这种非线性几何结构. 近些年来, 流形学习作为一种新的学习理论, 在机器学习领域取得了成功的应用[19−20]. 流形学习假设数据落在某个高维空间中的一个低维流形上, 它的目标就是要挖掘数据的这种非线性几何结构. 无论是在降维还是在分类问题上, 流形学习均取得了较好的效果.

正因为如此, 本文借鉴了流形学习的思想, 在流形正则化的理论框架下, 改进了 ELM 算法, 利用流形学习的方法挖掘数据在 ELM 特征空间中的非线性几何结构, 以进一步提高 ELM 算法的性能.

2.1　流形正则化理论框架

流形正则化的本质思想是使数据在新的投影空间 (这里指决策空间) 中能够保持数据在原特征空间 (这里指 ELM 特征空间) 中的局部几何结构, 即, 如果某两个样本在原特征空间中的相似度很大, 则它们在新的投影空间中的距离应该很近[26−27]. 该目标函数可以表示如下.

$$\min: \frac{1}{2}\sum_{ij} \omega_{ij} \| \boldsymbol{y}_i - \boldsymbol{y}_j \|_2^2 \tag{6}$$

其中, $\boldsymbol{y}_i$ 和 $\boldsymbol{y}_j$ 分别是样本 $\boldsymbol{z}_i$ 和 $\boldsymbol{z}_j$ 的输出预测向量 (在 ELM 结构中, $\boldsymbol{y}_i$ 是决策层输出向量, $\boldsymbol{z}_i$ 是隐含层输出向量), ω_{ij} 表示 $\boldsymbol{z}_i$ 和 $\boldsymbol{z}_j$ 的相似度. 为了计算两个样本点间的相似度, 通常采用 K 近邻方法, 即找到与某样本点 $\boldsymbol{z}_i$ 距离最近的 K 个样本, 将这 K 个样本设定为 $\boldsymbol{z}_i$ 的邻居, 然后利用高斯核函数计

算各点与 $\boldsymbol{z}_i$ 的相似度, ω_{ij} 的计算方法如下式所示.

$$\omega_{ij}=\begin{cases}\exp\left(\dfrac{-\parallel \boldsymbol{z}_i-\boldsymbol{z}_j\parallel_2^2}{\rho}\right), & e(\boldsymbol{z}_i,\boldsymbol{z}_j)=1 \text{ 并且} \\ & C(\boldsymbol{z}_i)=C(\boldsymbol{z}_j) \\ 0, & \text{其他}\end{cases} \tag{7}$$

其中, $e(\boldsymbol{z}_i,\boldsymbol{z}_j)=1$ 表示 $\boldsymbol{z}_i$ 是 $\boldsymbol{z}_j$ 的邻居或者 $\boldsymbol{z}_j$ 是 $\boldsymbol{z}_i$ 的邻居. $C(\boldsymbol{z}_i)=C(\boldsymbol{z}_j)$ 表示 $\boldsymbol{z}_i$ 和 $\boldsymbol{z}_j$ 属于同一个类别, 因为本文研究的是有监督学习算法, 所以这里只对来自同一类别的互为邻居的数据计算相似度, 其他互为邻居的数据不进行计算, 这一点不同于半监督的流形学习算法[28] (不考虑类别信息, 对所有互为邻居的数据计算相似度). 另外, 文献 [26, 29] 将类间信息也引入到流形正则化理论框架中, 其本质思想是, 若来自不同类别的数据在原特征空间中是邻接的, 则拉大其在决策空间中的距离, 两篇文献采用了不同的优化方法. 然而, 文献 [26] 的报道表明, 并不能确定加入类间信息是否可以提高系统性能. 而文献 [29] 采用的优化方法由于存在一些潜在问题[30], 有时会使系统性能下降. 所以, 本文只考虑类内信息, 并没有利用类间信息.

通过上述方法, 可以得到相似度矩阵 W.

$$W=\begin{bmatrix}\omega_{11} & \cdots & \omega_{1N}\\ \vdots & \ddots & \vdots\\ \omega_{N1} & \cdots & \omega_{NN}\end{bmatrix}_{N\times N} \tag{8}$$

最小化式 (6) 中的目标函数相当于最小化如下目标函数[28].

$$\min:\mathrm{tr}(Y^{\mathrm{T}}LY) \tag{9}$$

其中, $\mathrm{tr}(\cdot)$ 表示矩阵的迹, Y 是所有训练样本的输出预测向量拼接而成的矩阵, 即

$$Y=\begin{bmatrix}\boldsymbol{y}_1^{\mathrm{T}}\\ \vdots\\ \boldsymbol{y}_N^{\mathrm{T}}\end{bmatrix}_{N\times m} \tag{10}$$

矩阵 L 是拉普拉斯矩阵, 即 $L=D-W$, D 是一个对角矩阵.

$$D_{ii}=\sum_{j=1}^{N}W_{ij} \tag{11}$$

2.2 流形正则化 ELM 算法

基于上述分析和推导, 在这一部分, 我们将流形学习理论引入到 ELM 算法中, 以克服 ELM 算法存在的潜在问题. 这里, 我们以 ELM 特征空间作为出发点, 假定由于随机性因素, 数据分布在该高维特征空间中的一个低维流形上, 这种数据的几何结构需要通过上述流形学习的方法加以挖掘, 所以, 这里将式 (9) 引入到 ELM 目标函数中, 则 MRELM 的目标函数如下所示:

$$\begin{aligned}\min:\ & \frac{1}{2}\sum_{i=1}^{N}\parallel\boldsymbol{\xi}_i\parallel_2^2+\frac{C_1}{2}\parallel\beta\parallel_2^2+\frac{C_2}{2}\mathrm{tr}(Y^{\mathrm{T}}LY)\\ \text{s.t.}\ & h(\boldsymbol{x}_i)^{\mathrm{T}}\beta=\boldsymbol{t}_i^{\mathrm{T}}-\boldsymbol{\xi}_i^{\mathrm{T}},\quad i=1,\cdots,N\\ & \boldsymbol{y}_i^{\mathrm{T}}=h(\boldsymbol{x}_i)^{\mathrm{T}}\beta\end{aligned} \tag{12}$$

这里, 拉普拉斯矩阵 L 是由在 ELM 特征空间中的样本求得. 上式目标函数中第三项的意义就是使得在 ELM 特征空间中相似度较大的数据在决策空间中的距离较小, 即在决策空间中保持数据在原 ELM 特征空间中的几何性质. 将约束项代入目标函数, 上式可以表示为如下形式:

$$\min:\frac{1}{2}\parallel H\beta-T\parallel_2^2+\frac{C_1}{2}\parallel\beta\parallel_2^2+\frac{C_2}{2}\mathrm{tr}(\beta^{\mathrm{T}}H^{\mathrm{T}}LH\beta) \tag{13}$$

对矩阵 β 求导可得:

$$\nabla=H^{\mathrm{T}}(H\beta-T)+C_1\beta+C_2H^{\mathrm{T}}LH\beta \tag{14}$$

令 $\nabla=0$, 可求得:

$$\beta=(H^{\mathrm{T}}H+C_1I+C_2H^{\mathrm{T}}LH)^{-1}H^{\mathrm{T}}T \tag{15}$$

至此, MRELM 的训练算法已经实现. 现将 MRELM 算法总结如下:

步骤 1. 设定隐含节点数和激活函数, 随机初始化输入层与隐含层之间的权值矩阵以及隐含层节点偏置.

步骤 2. 将全体训练数据映射到 ELM 特征空间中.

步骤 3. 在该空间中, 构建拉普拉斯矩阵 L.

步骤 4. 根据式 (15) 计算输出矩阵 β.

在测试阶段, 对于测试样本 $\boldsymbol{x}$, 直接按如下公式计算其分数:

$$\boldsymbol{y}=h(\boldsymbol{x})^{\mathrm{T}}\beta \tag{16}$$

3 实验设置

3.1 数据库

为了验证算法的有效性, 本文的实验基于美国国家技术和标准署 (National Institute of Standards and Technology, NIST) 2009 年语种识别评测[31]. NIST 2009 语种识别评测包含 23 个语种, 分别是阿姆哈拉语、波斯尼亚语、汉语广东话、海地克里奥尔语、克罗地亚语、达里语、美国英语、印度英语、波斯语、法语、格鲁吉亚语、豪撒语、印地语、韩语、汉语普通话、普什图语、葡萄牙语、俄语、西班牙语、土耳其语、乌克兰语、乌尔都语、越南语.

训练数据包含电话信道数据 (Conversational telephone speech, CTS) 和广播信道数据 (Voice of america radio broadcasts, VOA), CTS 数据来源于 NIST 2007 语种识别评测提供的数据以及 CallFriend、CallHome、OGI-22 数据集, 训练数据一共有将近 30 000 条.

后端校准基于开发集数据. 开发集数据来源于 NIST 2003、2005、2007 提供的数据以及 VOA 数据集, 只有上述 23 个语种的数据被采用, 这部分数据一共有将近 10 000 条. 这里, 后端校准算法采用的是 LDA+MMI 算法[32].

测试集为 NIST 2009 标准测试库, 这里采用闭集测试, 即只对上述 23 个目标语种进行测试. 评价指标等错率 (Equal error rate, EER) 和平均代价 (Average cost, Cavg), 这两种指标由 NIST 定义, 是目前国际上广泛采用的评价指标.

3.2 特征提取

特征提取阶段, 我们首先提取了 13 维的梅尔频率倒谱系数 (Mel frequency cepstral coefficient, MFCC) 特征, 经过相对谱 (Relative spectral, RASTA) 技术滤波后, 进一步提取了移位差分倒谱系数 (Shifted dalta cepstral coefficient, SDCC) 特征, 特征参数为 7-1-3-7. 之后, 用所有的训练数

据训练出一个性别无关的通用背景模型 (Universal background model, UBM), 混合度为 512.

在此基础上, 接下来对每段语音提取了两种不同的特征, 分别为 GSV[24] 和 i-vector[25]. GSV 是采用一种基于 KL 距离的核函数的方法, 专门针对 SVM 训练而提取的一种特征. 这里, 通过均值自适应, 为每段语音提取了一条 28 672 维的 GSV 特征, 每条 GSV 特征都做了 NAP 补偿[33]. i-vector 是基于因子分析的方法提取的一种特征, 近些年来, 由于其优异的性能, 已经成为语种识别和说话人识别领域里广泛采用的特征之一. 这里采用了标准的 i-vector 提取方法, 对每段语音, 提取了 400 维的 i-vector 特征.

基于 MRELM 算法的语种识别系统流程如图 2 所示.

4　实验结果

4.1　基于 GSV 的系统

为了验证算法的有效性, 首先我们在 GSV 特征上将 MRELM 与 SVM 以及传统的 ELM 作对比.

对于 SVM 系统, 这里采用一对其他的训练方法, 该方法是目前基于 SVM 的语种识别系统广泛采用的训练方法.

对于 MRELM 系统, 这里设定隐含节点数为 20 000, 输入权值以及隐含层偏置采用均匀分布的随机初始化, 初始化区间为 $[-0.5, +0.5]$. 这里, 我们采用交叉验证的方法来确定平衡参数 $C1$ 和 $C2$, 即将全体训练数据平均分成 3 份, 每次用其中的 2 份进行训练, 剩下的 1 份用于测试, 以此来确定和的最优值. 这里, 通过实验确定 $C1 = 10\,000, C2 = 0.001$.

对于传统的 ELM 算法, 隐含节点数以及随机初始化参数与 MRELM 相同, $C1 = 10\,000$.

实验结果如表 1 所示, 该结果对各系统做分数校准之前 (BB) 和之后 (AB) 的性能均进行了比较. 从实验结果可以看出: 相比与 SVM, MRELM 在不同段长的测试条件下、在做分数校准之前和之后的各项指标性能均优于 SVM, 尤其是在 30 秒的测试条件下, MRELM 相对 SVM 可以取得 22.69 % (BB, EER)、22.75 (BB, Cavg)、13.08 % (AB, EER) 和 16.67 % (AB, Cavg) 的提升. 此外, 在大多数测试条件下, MRELM 的各项性能也优于传统的 ELM 算法, 这说明了引入流形正则化的必要性, 同时也验证之前关于 ELM 特征空间可能存在潜在问题的假设.

4.2　基于 i-vector 的系统

此外, MRELM 也被应用到了 i-vector 特征中. 目前, 针对 i-vector 特征, 国际上最主流的方法是首先采用线性鉴别性分析 (Linear discriminative analysis, LDA) 对 i-vector 进行降维, 然后采用余弦距离打分 (Cosine distance scoring, CDS) 的方法进行打分[34]. 这里, 令 $\boldsymbol{l}$ 表示 i-vector 通过 LDA 降维后得到的单位长度的特征, 则语种 k 的模型 $\boldsymbol{m}_k$ 可由下式求得:

$$\boldsymbol{m}_k = \frac{\sum\limits_{i=1}^{N_k} \boldsymbol{l}_i}{\left\| \sum\limits_{i=1}^{N_k} \boldsymbol{l}_i \right\|_2} \tag{17}$$

其中, N_k 表示语种 k 中的训练样本数. 则测试段 $\boldsymbol{l}_{\text{test}}$ 对语种 k 的分数为

$$s = \boldsymbol{m}_k^{\mathrm{T}} \boldsymbol{l}_{\text{test}} \tag{18}$$

这里, 我们首先采用该标准方法进行识别. 然后, 为了对比, 将 SVM 应用到经过 LDA 降维的 i-vector 特征空间中. 最后, 直接将 MRELM 算法应用到 i-vector 原始特征空间中. 这里, 隐含节点数设为 6 000, $C1 = 200$, $C2 = 0.001$, 实验发现, 继续增大隐含节点数, 系统性能没有进一步提高.

表 1　不同系统在 GSV 特征空间中的性能

Table 1　Performance of different systems in GSV space

系统 (s)			SVM	ELM	MRELM
EER	30	BB	7.05	6.31	**5.45**
		AB	3.21	3.01	**2.79**
	10	BB	13.02	12.30	**11.98**
		AB	8.77	7.79	**7.37**
	3	BB	25.25	**23.92**	24.20
		AB	21.88	22.00	**19.50**
Cavg	30	BB	7.34	6.56	**5.67**
		AB	3.24	2.82	**2.70**
	10	BB	13.17	12.69	**12.48**
		AB	8.76	7.86	**7.66**
	3	BB	25.35	**24.43**	24.62
		AB	21.69	21.97	**19.71**

三种方法的实验结果如表 2 所示. 从实验结果可以看出, 在大多数测试条件下, MRELM 的性能优于 LDA+SVM, 并且与标准的 LDA+CDS 打分方法可以取得相近的分类性能, 尤其是在做完后端分数校准之后, 性能略优于 LDA+CDS 打分. 这说明 MRELM 算法可以成功地应用在 i-vector 特征空间中.

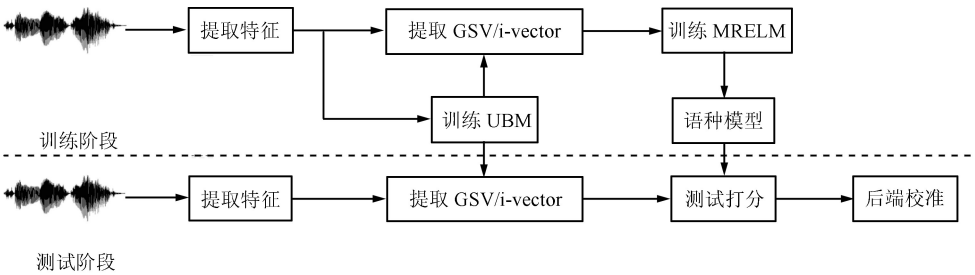

图 2　基于 MRELM 算法的语种识别系统框图

Fig. 2　Diagram of language recognition system with MRELM

表 2 不同系统在 i-vector 特征空间中的性能

Table 2 Performance of different systems in i-vector space

系统 (s)			LDA+SVM	LDA+CDS	MRELM
EER	30	BB	6.05	**4.52**	4.84
		AB	2.70	2.63	**2.49**
	10	BB	11.40	**9.56**	10.32
		AB	6.78	6.72	**6.43**
	3	BB	24.27	**21.70**	23.14
		AB	19.92	**20.22**	20.52
Cavg	30	BB	6.27	**4.79**	5.20
		AB	2.62	2.59	**2.46**
	10	BB	11.76	**9.71**	10.87
		AB	6.97	6.80	**6.49**
	3	BB	24.47	**21.90**	23.18
		AB	**19.80**	20.36	21.10

4.3 实验结果分析

从上述实验中可以看出, 在绝大多数情况下, MRELM 算法的性能优于 SVM. 相比与 SVM, MRELM 具有一些潜在的优点. 首先, 在文献 [8] 中, 作者将 ELM 和 SVM 在优化理论上进行了对比. ELM 和 SVM 具有相似对偶形式的目标函数, 但是 SVM 在求解拉格朗日算子 α 的过程中需要满足如下两项约束:

$$\sum_{i=1}^{N} \alpha_i t_i = 0 \tag{19}$$

$$0 < \alpha_i < \lambda \tag{20}$$

其中, λ 是一个大于 0 的平衡参数. 即 SVM 是在 N 维空间中的某个有限空间里的一个平面上寻求最优解. 然而, ELM 只需要满足该约束条件 (20), 即 ELM 是在 N 维空间中的某个有限空间里寻求最优解, 从这个角度上看, ELM 可以找到一个全局最优解, 而 SVM 只能找到局部最优解. MRELM 算法作为 ELM 算法的一个扩展, 继承了 ELM 算法的这个优点. 此外, MRELM 改进了 ELM 算法的不足, 能够挖掘数据在 ELM 特征空间中的非线性结构, 所以, 综上所述, 相对于 SVM, MRELM 往往可以取得更好的识别性能.

5 结论

本文将极限学习机算法引入到语种识别中, 并针对 ELM 算法存在的问题, 提出了 MRELM 算法, 相对于传统的 SVM 系统, 该算法使语种识别系统的性能有了显著提升, 同时在 i-vector 特征空间中, 该算法的性能与当前主流的打分算法相当. 因此, 该算法可以成功地应用到当前语种识别系统当中, 为语种识别开辟了一个新的研究方向.

References

1 Li H Z, Ma B, Lee K A. Spoken language recognition: from fundamentals to practice. *Proceedings of the IEEE*, 2013, **101**(5): 1136−1159

2 Biadsy F. Automatic dialect and accent recognition and its application to speech recognition [Ph. D. dissertation], Columbia University, USA, 2011.

3 Zissman M A, Berkling K M. Automatic language identification. *Speech Communication*, 2001, **35**(1−2): 115−124

4 Muthusamy Y K, Barnard E, Cole R A. Reviewing automatic language identification. *IEEE Signal Processing Magazine*, 1994, **11**(4): 33−41

5 Campbell W M, Singer E, Torres-Carrasquillo P A, Reynolds, D A. Language recognition with support vector machines. In: Proceedings of the 2004 ODYSSEY-The Speaker and Language Recognition Workshop. Toledo, Spain: ISCA, 2004. 285−288

6 Campbell W M, Campbell J P, Reynolds D A, Singer E, Torres-Carrasquillo P A. Support vector machines for speaker and language recognition. *Computer Speech & Language*, 2006, **20**(2−3): 210−229

7 Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: a new learning scheme of feedforward neural networks. In: Proceedings of the 2004 IEEE International Joint Conference on Neural Networks. Budapest, Hungary: IEEE, 2004. 985−990

8 Huang G B, Wang D H, Lan Y. Extreme learning machines: a survey. *International Journal of Machine Learning and Cybernetics*, 2011, **2**(2): 107−122

9 Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: theory and applications. *Neurocomputing*, 2006, **70**(1−3): 489−501

10 Huang G B, Zhou H M, Ding X J, Zhang R. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2012, **42**(2): 513−529

11 Liang N Y, Huang G B, Saratchandran P, Sundararajan N. A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Transactions on Neural Networks*, 2006, **17**(6): 1411−1423

12 Xu J T, Zhou H M, Huang G B. Extreme learning machine based fast object recognition. In: Proceedings of the 15th IEEE International Conference on Information Fusion. Singapore: IEEE, 2012. 1490−1496

13 Sole M M, Tsoeu M S. Sign language recognition using the extreme learning machine. In: Proceedings of the 2011 IEEE AFRICON Conference. Livingstone, Zambia: IEEE, 2011. 1−6

14 Suresh S, Babu V, Sundararajan N. Image quality measurement using sparse extreme learning machine classifier. In: Proceedings of the 9th IEEE International Conference on Control, Automation, Robotics and Vision. Singapore: IEEE, 2006. 1−6

15 Horata P, Chiewchanwattana S, Sunat K. Robust extreme learning machine. *Neurocomputing*, 2013, **102**: 31−44

16 Yu Q, Miche Y, Eirola E, Van Heeswijk M, Séverin E, Lendasse A. Regularized extreme learning machine for regression with missing data. *Neurocomputing*, 2013, **102**: 45−51

17 Zong W W, Huang G B, Chen Y Q. Weighted extreme learning machine for imbalance learning. *Neurocomputing*, 2013, **101**: 229−242

18 Iosifidis A, Tefas A, Pitas I. Minimum class variance extreme learning machine for human action recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 2013, **23**(11): 1968−1979

19 Tenenbaum J B, De Silva V, Langford J C. A global geometric framework for nonlinear dimensionality reduction. *Science*, 2000, **290**(5500): 2319−2323

20 Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000, **290**(5500): 2323−2326

21 Huang G, Song S J, Gupta J N D, Wu C. Semi-supervised and unsupervised extreme learning machines. *IEEE Transactions on Cybernetics*, 2014, **44**(12): 2405−2417

22 Liu B, Xia S X, Meng F R, Zhou Y. Manifold regularized extreme learning machine. *Neural Computing and Applications*, 2015, DOI: 10.1007/s00521-014-1777-8

23 Deng W Y, Zheng Q H, Chen L. Regularized extreme learning machine. In: Proceedings of the 2009 IEEE Symposium on Computational Intelligence and Data Mining. Nashville, USA: IEEE, 2009. 389−395

24 Campbell W M, Sturim D E, Reynolds D A. Support vector machines using GMM supervectors for speaker verification. *IEEE Signal Processing Letters*, 2006, **13**(5): 308−311

25 Dehak N, Kenny P, Dehak R, Dumouchel P, Ouellet P. Front-end factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011, **19**(4): 788−798

26 Tomar V S, Rose R C. Manifold regularized deep neural networks. In: Proceedings of the 2014 Annual Conference of the International Speech Communication Association. Singapore: ISCA, 2014. 348−352

27 Guan N Y, Tao D C, Luo Z G, Yuan B. Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent. *IEEE Transactions on Image Processing*, 2011, **20**(7): 2030−2048

28 Belkin M, Niyogi P, Sindhwani V. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. *The Journal of Machine Learning Research*, 2006, **7**: 2399−2434

29 Peng Y, Zhu J Y, Zheng W L, Lu B L. EEG-based emotion recognition with manifold regularized extreme learning machine. In: Proceedings of the 36th IEEE International Conference on Engineering in Medicine and Biology Society. San Diego, USA: IEEE, 2014. 974−977

30 Wang H, Yan S C, Xu D, Tang X A, Huang T. Trace ratio vs. ratio trace for dimensionality reduction. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA: IEEE, 2007. 1−8

31 Martin A F, Greenberg C S. The 2009 NIST language recognition evaluation. In: Proceedings of the 2010 ODYSSEY-The Speaker and Language Recognition Workshop. Brno, Czech Republic: ISCA, 2010. 165−171

32 Zhang W Q, Hou T, Liu J. Discriminative score fusion for language identification. *Chinese Journal of Electronics*, 2010, **19**(1): 124−128

33 Campbell W M, Sturim D E, Reynolds D A, Solomonoff A. SVM based speaker verification using a GMM supervector kernel and NAP variability compensation. In: Proceedings of the 2006 IEEE International Conference on Acoustics, Speech and Signal Processing. Toulouse, France: IEEE, 2006. 1−1

34 Singer E, Torres-Carrasquillo P, Reynolds D, McCree A, Richardson F, Dehak N, Sturim D. The MITLL NIST LRE 2011 language recognition system. In: Proceedings of the 2012 The Speaker and Language Recognition Workshop. Singapore: ISCA, 2012. 209−215

徐嘉明 中国科学院电子学研究所博士研究生. 主要研究方向为语音信号处理, 机器学习. 本文通信作者. E-mail: xujiaming09@sina.com
(**XU Jia-Ming** Ph. D. candidate at the Institute of electronics, Chinese Academy of Sciences. His research interest covers speech signal processing and machine learning. Corresponding author of this paper.)

张卫强 清华大学电子工程系副研究员. 主要研究方向为语音信号处理, 机器学习. E-mail: wqzhang@tsinghua.edu.cn
(**ZHANG Wei-Qiang** Associate professor in the Department of Electronic Engineering, Tsinghua University. His research interest covers speech signal processing and machine learning.)

杨登舟 中国科学院电子学研究所博士研究生. 主要研究方向为语音信号处理, 机器学习. E-mail: yangdengzhou@sina.com
(**YANG Deng-Zhou** Ph. D. candidate at the Institute of electronics, Chinese Academy of Sciences. His research interest covers speech signal processing and machine learning.)

刘 加 清华大学电子工程系教授. 主要研究方向为语音识别, 信号处理. E-mail: liuj@tsinghua.edu.cn
(**LIU Jia** Professor in the Department of Electronic Engineering, Tsinghua University. His research interest covers speech recognition and signal processing.)

夏善红 中国科学院电子学研究所研究员. 主要研究方向为信号处理, 传感技术. E-mail: shxia@mail.ie.ac.cn
(**XIA Shan-Hong** Professor at the Institute of Electronics, Chinese Academy of Sciences. Her research interest covers signal processing and transducer technology.)