

基于重组超矢量的 GMM-SVM 说话人辨认系统

欧国振,孙林慧,薛海双

(南京邮电大学 通信与信息工程学院,江苏 南京 210003)

摘要:在传统的高斯混合模型-支持向量机(Gaussian Mixture Model-Support Vector Machine, GMM-SVM)说话人辨认系统中,SVM 利用从 GMM 矢量空间中得到的超矢量(Super Vector)直接对说话人进行建模与识别,由于没有考虑到超矢量内各均值矢量之间的关联性,识别性能有限。为此,提出了基于重组超矢量构建文本无关的 GMM-SVM 说话人辨认系统。该系统充分利用各相邻高斯分量的均值矢量的高度关联性,保证了重组后的超矢量能充分反映说话人身份的内在细节,使得系统具有充分利用 SVM 处理高维小数据性能的优越特点。验证实验结果表明,与传统的 GMM-SVM 系统相比,重组超矢量 GMM-SVM 说话人辨认系统显著地缩短了系统建模的时间,同时有效地提高了说话人的辨别率。

关键词:说话人辨认;高斯混合模型-支持向量机;超矢量重组;辨别率;建模时间

中图分类号:TP302

文献标识码:A

文章编号:1673-629X(2017)07-0051-06

doi:10.3969/j.issn.1673-629X.2017.07.012

GMM-SVM Speaker Identification System with Recombination of GMM Super Vector

OU Guo-zhen, SUN Lin-hui, XUE Hai-shuang

(College of Telecommunications & Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: In the traditional speaker identification system with Gaussian Mixture Model-Support Vector Machine (GMM-SVM), SVM uses super vector derived from the vector space of GMM to model and identify the target speakers directly. Since the relationship between two of mean vectors among GMM super vectors has not been considered, the performance of GMM-SVM system is limited. Thus a new text-independent GMM-SVM speaker identification system with super vector has been proposed which has made full use of tremendous correlation of each mean vector of the adjacent Gaussian components. The recombination super vectors have presented more inner detail of speakers' identity and enable the new system to take the advantage of the characteristics of superior performance when SVM deals with the small and high dimensional data. The experimental results demonstrate that the GMM-SVM speaker identification system with recombination super vector has not only achieved a higher recognition rate than the traditional GMM-SVM system, but also significantly decreased identification time of speakers.

Key words: speaker identification; GMM-SVM; super vector recombination; identification rate; time of modeling

0 引言

语音信号中不仅包含文本内容的信息,也包含了说话人的身份信息,这使得自动说话人识别成为可能^[1-2]。说话人识别是指从语音信号中判断出说话者的身份信息,在司法、军事、经济等领域应用广泛。说话人识别可分为说话人辨认和说话人确认两个类别,前者是多对一的问题,而后者则是一对一的问题。文本无关的说话人辨认系统是指不用为测试语音指定特定的文本内容的说话人辨认系统^[3],与之相反的文本

相关说话人辨认系统需要为测试语音指定特定的文本内容^[4-5]。

在说话人识别中常用的概率统计模型是隐马尔可夫模型(Hidden Markov Model, HMM)和 GMM^[6-7]。在基于 GMM 的识别任务中,一般首先选择适当混合数的 GMM 来刻画说话人的个性特征,然后提取每个说话人的均值矢量,来代表说话人的身份信息。因为 GMM 能够有效擦除语音的文本信息,突显说话人的身份信息,所以更加适合于文本无关的识别任务。

收稿日期:2016-07-28

修回日期:2016-11-03

网络出版时间:2017-04-28

基金项目:国家自然科学基金资助项目(61271335);江苏省自然科学基金项目(BK20140891);南京邮电大学校科研基金项目(NY214038)

作者简介:欧国振(1991-),男,硕士,研究方向为信号处理及其应用技术;孙林慧,副教授,研究方向为语音处理与现代语音通信技术。

网络出版地址:http://www.cnki.net/kcms/detail/61.1450.TP.20170428.1702.004.html

SVM 是一种有效的说话人识别模型^[8-12], SVM 技术通过将特征向量进行非线性映射来提高模型的区分性能,但当要处理的数据量较大时, SVM 系统的分类性能明显下降,并且算法收敛慢。GMM-SVM 系统的提出很好地解决了 SVM 模型的缺点^[13-14]。首先, GMM supervector 可以看成是 GMM 特征输入的提炼^[15-16],更好地刻画说话人的身份信息;其次,提取 GMM supervector 可以看成是特征输入的一个降维过程,从而降低 SVM 系统处理数据的复杂度,充分利用 SVM 处理小数据量时的优越性能。前期研究表明, GMM-SVM 系统在说话人识别领域取得了优越的成果。但是原始的 GMM-SVM 系统没有考虑到各个高斯分量的均值矢量间的关联性,也没有充分利用 SVM 处理高维小数据的优越性能。

为了更好地利用各相邻高斯分量间均值矢量的高相关性,提出了重组超矢量说话人辨认系统,对 GMM 组成超矢量的各个均值向量进行重新组合,使其更加契合 SVM 对数据处理的机制。首先依据各相邻高斯分量间均值矢量的相关性,选择适当数量的高斯分量的均值矢量,然后把这些均值矢量首尾连接形成新的均值矢量,按这个方法依次遍历整个超矢量形成重组超矢量,随后将所得到的重组超矢量运用于 SVM 系统的训练和识别。

1 基于重组的超矢量的 GMM-SVM 系统

提出的基于重组的超矢量的 GMM-SVM 系统框图如图 1 所示。

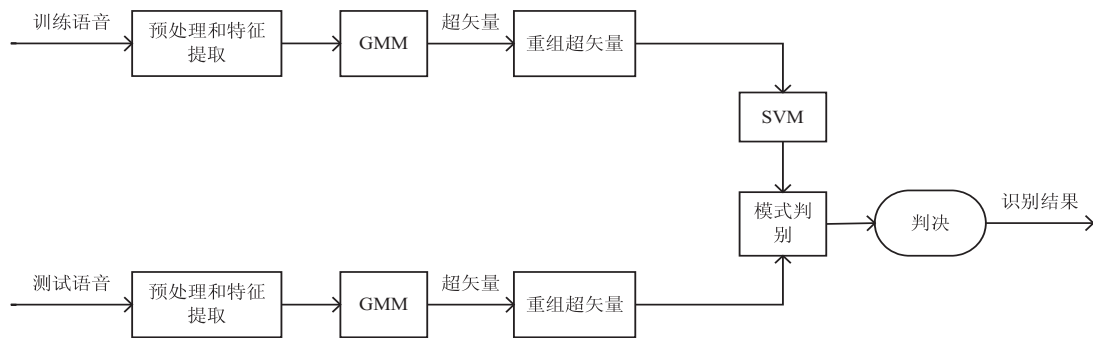


图 1 基于重组的超矢量的 GMM-SVM 系统框图

语音样本首先经过前端的预处理和特征提取,预处理主要包括预加重、分帧加窗等,语音信号的预处理对系统性能的提高作用明显。然后进行语音信号的特征参数提取,所提取的参数为梅尔频率倒谱系数(Mel-Frequency Cepstral Coefficients, MFCC)和 MFCC 的一阶差分。MFCC 是语音特征领域常用的特征参数^[17-18],符合人耳的听觉特性。随后将特征参数用于 GMM 训练。理论上,只要选择合适的 GMM 的混合数 M , GMM 可以拟合任何数据。超矢量是 GMM 的主要参数,在说话人识别领域,超矢量充分包含了说话人的个性信息,因此很多研究者将超矢量用于 SVM 的训练和识别。在此也不例外,但是将从 GMM 得到的超矢量进行重组,因为相邻的各均值矢量具有一定的关联性,并且 SVM 处理高维小数据的性能更好。提出的重组超矢量更好地利用了各个均值矢量之间的关联性,可以平滑均值矢量间的过渡,并且更加符合 SVM 对数据处理的机制,明显提高了系统的识别性能,大大缩短了系统建模所需的时间。

1.1 GMM

GMM 是说话人识别领域的一个成功应用,一个 M 阶的高斯混合模型由 M 个高斯概率密度函数加权得到。假设某说话人 j 的某段语音经过预处理后分为

T 帧信号,并且对每一帧信号提取 D 维的特征参数,这段语音的特征参数记为 X , $X \in R^{T \times D}$ 。对 X 建立 GMM,可表示为:

$$P(X/\lambda) = \sum_{i=1}^M \omega_i b_i(X) \quad (1)$$

其中, $\omega_i (i=1, 2, \dots, M)$ 是混合权重, M 是混合数,并且 $\sum_{i=1}^M \omega_i = 1$; $b_i(X) (i=1, 2, \dots, M)$ 是高斯概率密度函数子分布,可表示为:

$$b_i(X) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp \left\{ -\frac{1}{2} (X - \mu_i)^T \Sigma_i^{-1} (X - \mu_i) \right\} \quad (2)$$

其中, μ_i 为均值矢量; Σ_i 为协方差矩阵。

可以看出, GMM 由参数均值矢量、混合权重和协方差矩阵组成,即:

$$\lambda = \{\omega_i, \mu_i, \Sigma_i\}, i=1, 2, \dots, M \quad (3)$$

对 GMM 的训练实际上就是对参数 λ 估计的过程,对序列 $X = \{X_1, X_2, \dots, X_T\}$, GMM 的似然表示为:

$$P(X/\lambda) = \prod_{i=1}^T P(X_i/\lambda) \quad (4)$$

常用的参数估计方法是最大似然(Maximum Likelihood, ML)估计,由于式(4)是 λ 的非线性函数,所以

采用 EM (Expectation Maximization, EM) 算法对参数进行估计。可得:

混合权重的迭代公式为:

$$\omega_i = \frac{1}{T} \sum_{t=1}^T P(i/X_t, \lambda) \quad (5)$$

均值的迭代公式为:

$$\mu_i = \frac{\sum_{t=1}^T P(i/X_t, \lambda) X_t}{\sum_{t=1}^T P(i/X_t, \lambda)} \quad (6)$$

方差的迭代公式为:

$$\sigma_i^2 = \frac{\sum_{t=1}^T P(i/X_t, \lambda) (X_t - \mu_i)^2}{\sum_{t=1}^T P(i/X_t, \lambda)} \quad (7)$$

其中 i 的后验概率为:

$$P(i/X_t, \lambda) = \frac{\omega_i b_i(X_t)}{\sum_{k=1}^M \omega_k b_k(X_t)} \quad (8)$$

在 EM 算法开始前要先给定混合数 M 。

1.2 重组的 GMM 超矢量

选择合适的 GMM 混合数 M , 通过对参数协方差矩阵、混合权重和均值矢量的估计, GMM 可以很好地描述说话人的身份信息。而每个高斯分量的均值矢量按照次序连接起来就构成了超矢量, 超矢量包含了绝大多数说话人的个性信息。若输入的特征参数的维数为 D , 每个高斯分量的均值矢量为 m_1, m_2, \dots, m_M , 其中 $m_i (i = 1, 2, \dots, M)$ 是维数为 D 的向量。超矢量 m 可以表示为:

$$m = (m_1, m_2, \dots, m_M)^T \quad (9)$$

一般的, 在训练阶段将代表话者身份信息的矢量 m 作为 SVM 的训练数据。

通过实验发现, 在一定范围内的高斯分量的均值矢量之间存在一定的关联性, 为此, 提出了重组超矢量这一概念。其基本思想是, 根据 GMM 的混合数 M , 关联一定范围内高斯分量的均值矢量, 构成一个新的均值向量; 按照这个方法依次重组, 直到遍历所有的均值向量。所关联的高斯分量均值矢量的个数被称为高斯关联数。

一个混合数为 M 的 GMM 模型, 即原始的超矢量有 M 个均值矢量, 若选择高斯关联数的个数是 $q (1 \leq q \leq M)$, 则得到的第一个新的均值向量为: $m'_1 = (m_1, m_2, \dots, m_q)$, 依次遍历整个超矢量, 得到重组超矢量。重组超矢量的均值向量个数为 p , 并且满足以下关系:

$$M = q \times p \quad (10)$$

则重构超矢量 m' 为:

$$m' = (m'_1, m'_2, \dots, m'_p)^T \quad (11)$$

对任意的 $i \in [1, p]$, 重组超矢量 m'_i 的高斯关联数相等。重组超矢量充分利用了高斯分量均值向量之间的关联性, 平滑数据间的过渡, 重组出更能代表说话人身份信息的超矢量, 更加契合 SVM 分类器对数据实现分类的要求。实验结果表明, 提出的重组超矢量有效提高了系统的识别性能, 大大缩短了系统建模的时间。

1.3 支持向量机

SVM 是基于超平面的一个二分类模型, 通过最大化间隔边界到超平面的间隔实现数据的分类。假设线性可分的分类任务特征空间训练数据集 $T = \{x_i, y_i\}$, 其中, $x_i \in R^n$, 是特征向量, $y_i \in \{+1, -1\}$, 是分类标签。数据集由以下表示的超平面分开。

$$w \cdot x + b = 0 \quad (12)$$

其中, w 为超平面的法向量; b 为截距。

支持向量机的学习可以通过求解以下最优化问题得到:

$$\min_{w, b} \frac{1}{2} \|w\|^2 \quad (13)$$

$$\text{s. t. } y_i(w \cdot x_i + b) \geq 1, i = 1, 2, \dots, n$$

这是一个凸二次规划 (convex quadratic programming) 问题, 并且将 $y_i(w \cdot x_i + b)$ 称为函数间隔。

当某些特征参数样本点不能满足式 (13) 的约束条件时, 上述最优化问题解决这个分类问题不是很理想。

为此引入一个松弛变量 $\xi_i \geq 0$, 放低函数间隔的要求, 做法是使函数间隔加上这个松弛变量之后大于等于 1。即

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i \quad (14)$$

同时, 对每一个松弛变量相应地带上一个惩罚参数, 原先的最优化问题变为:

$$\min_{w, b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (15)$$

$$\text{s. t. } y_i(w \cdot x_i + b) \geq 1 - \xi_i, i = 1, 2, \dots, n$$

$$\xi_i \geq 0, i = 1, 2, \dots, n$$

当数据必须要通过利用非线性模型才能很好地进行分类时, 以上方法就不能有效地解决问题。为此, 引入了核函数的方法, 通过把输入数据映射到一个新的特征空间, 一般的, 这个新的特征空间的维数比原始空间的维数更高, 也有可能是无穷维。在这个新的空间中, 就可以用线性分类的方法来解决这个问题。

常用的核函数有径向基核函数、多项式核函数、sigmoid 核函数和线性核函数。采用径向基函数。核函数方法的模型学习是隐式进行的, 不需要显式地定义新的特征空间, 也不需要显式定义映射函数。在此采用一对一的方法实现多分类问题。

1.4 重组超矢量作为 SVM 的输入

通过对偶算法求解最优化问题(见式(13))可以得到 w 的解为:

$$w^* = \sum_i \alpha_i^* y_i x_i$$

(16)

其中, α^* 是对式(13)构建拉格朗日函数而引入的拉格朗日乘子的解。

而 b 的解为:

$$b^* = y_j - \sum_i \alpha_i^* y_i (x_i \cdot x_j)$$

(17)

则分离超平面可以写成:

$$\sum_i \alpha_i^* y_i (x_i \cdot x_j) + b^* = 0$$

(18)

分类决策函数可以写成:

$$f(x) = \text{sign}(\sum_i \alpha_i^* y_i (x \cdot x_i + b^*))$$

(19)

由此可看出,分类决策函数只依赖于输入 x 和训练样本输入的内积。当引入核函数后,式(19)变成:

$$f(x) = \text{sign}(\sum_i \alpha_i^* y_i (x \cdot x_i) + b^*)$$

(20)

从式(19)和式(20)可以看出,SVM 的计算复杂度与输入 x 的个数和训练样本的个数有很大关系。在提出的重组超矢量中,当选择的高斯关联数 $1 < q \leq M$, 则重组的超矢量均值矢量样本数比原始的超矢量的均值矢量样本数少,计算复杂度有所减少。文献[19]指出,SVM 的计算复杂度在 $O(N_s^3 + N_s^2 l + N_s d_l l)$ 和 $O(d_l l^2)$ 之间,其中 N_s 是支持向量的个数, l 是训练样本的个数,而 d_l 是每个训练样本的维数。虽然重组超矢量的每个样本的维数增加了,但是样本的个数却减少了。根据文献[19]的观点,样本维数对 SVM 计算复

杂度的影响最高是线性的,而样本个数有可能是平方的,因此使用重组超矢量作为 SVM 的输入可以有效减少 SVM 的计算复杂度。并且通过实验发现,使用重组超矢量训练 SVM 的支持向量的个数也得到了有效减少,从而可以有效降低系统的计算复杂度。

通过关联一定范围内的均值矢量,首尾连接形成新的均值矢量,新的均值矢量携带更丰富的说话人的身份信息,数据之间的过渡更加平滑,因而使用重组超矢量来表示说话人的个性特征可以提高系统的识别率。

2 实验结果和分析

实验数据采用笔者所在团队在消音室录制的语料,该数据库包含 210 个说话人的语音,每个人 180 条,语音信号的采样频率为 16 kHz。实验选择 10 个说话人,每人 30 条语句,其中训练语句 20 条,平均时长 100 s,测试语句 10 条,平均时长 30 s。

2.1 不同的超矢量重组方式和识别率的关系

在对语音进行分帧处理时,为了使数据的过渡更为平滑,帧长取 256 点,帧移取 128 点。每帧信号提取 24 维的 MFCC 参数和 24 维的一阶差分 MFCC 参数。如果不做特别声明,所采用的核函数为径向基核函数。固定高斯混合数为 128 的前提下,可选的高斯关联数有 $2^i, i \in [1,7]$ 。训练样本和测试样本的重组超矢量可选择不同的高斯关联数,用 q 表示。不同的关联数的组合方式对识别率的影响如表 1 所示。

表 1 不同高斯关联数组合方式对识别率的影响

训练样本 q	测试样本 q	正确识别率/%	训练样本 q	测试样本 q	正确识别率/%	训练样本 q	测试样本 q	正确识别率/%
16	128	10	8	128	10	4	128	10
16	64	10	8	64	10	4	64	10
16	32	96.9	8	32	10	4	32	10
16	16	97.5	8	16	83.75	4	16	10
16	8	73.12	8	8	93.75	4	8	80.63
16	4	33.75	8	4	54.06	4	4	94.68
16	2	15.63	8	2	17.65	4	2	48.44
16	1	14.063	8	1	17.26	4	1	27.10

表 1 为部分组合方式。由表 1 可看出,不同的组合方式对识别率具有巨大的影响。当训练样本的高斯关联数和测试样本的高斯关联数都一样时,系统的识别率达到最高。当两者所选的高斯关联数不相等时,识别率急剧下降,因为当两者的高斯关联数不相等时,也即 SVM 输入数据的维数不相等,从而造成维数不匹配,引入了大量的干扰噪声。值得一提的是,若训练样

本高斯关联数为 q_1 ,测试样本高斯关联数为 q_2 ,当 $q_1 < q_2$ 时,系统识别率要高于 $q_1 > q_2$ 时的系统识别率。当 $q_1 < q_2$ 时,即训练样本维数小于测试样本维数,相反,当 $q_1 > q_2$ 时,训练样本维数大于测试样本维数。样本维数的减少,意味着样本属性缺失,相比测试样本维数增多而言,使得分类器更加难以识别。

在训练样本高斯关联数和测试样本的高斯关联数

相等的前提下,选择不同的高斯关联数,系统识别率的变化如图 2 所示。

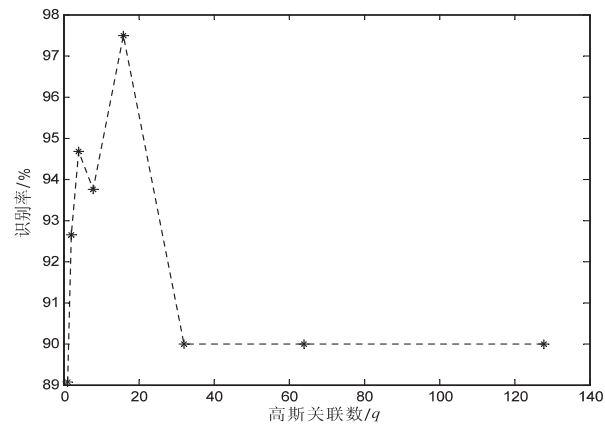


图 2 不同高斯关联数对识别率的影响

从图 2 可看出,当高斯关联数 q 为 1 时,即重组超矢量等于原始超矢量,这时系统识别率最低,为 89.06%。随着高斯关联数递增至 8,在这区间内系统的识别率大体上为一个上升的趋势,当高斯关联数为 16 时,系统的识别率达到最高,为 97.75%。此后,随着高斯关联数的增加系统识别率骤降,并很快趋于平稳。在高斯混合数 M 不变的前提下,高斯关联数 q 和均值矢量数 p 互为反比,所以随着高斯关联数的增加,均值矢量数随之减少,但是每一个均值矢量的样本维数随之增加。当高斯关联数增加到一定程度时,会产生样本维数冗余的现象,噪声干扰增大,识别率下降。可以看出将重组超矢量应用于系统中,均比应用原始超矢量的系统识别率高。根据所提供的数据可看出,提出的重构超矢量能够提高系统的识别率。

2.2 使用不同核函数和系统识别率的关系

SVM 的性能很大程度上取决于核函数的选择,但是目前并没有很好的方法来针对具体问题指导选择特定的核函数。核函数的方法把原问题空间的训练样本映射到新的特征空间中,将原来的非线性问题变为线性问题。图 3 是不同核函数相对应的系统识别率的关系图。

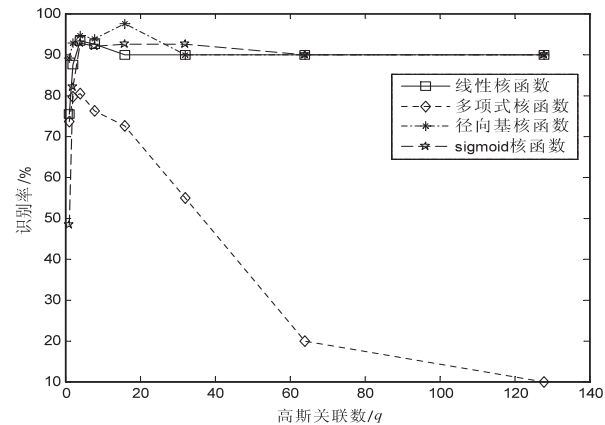


图 3 不同核函数类型对识别率的影响

从图 3 可以看出,使用多项式核函数时,系统的识别率最差,而使用径向基核函数时,系统的识别率最好,最高可达到 97.75%。在实际应用中,一般通过经验来选择合适的核函数。

2.3 重组方式和系统建模时间的关系

不同高斯关联数的重组方式对应的系统建模时间也各不相同,其关系如图 4 所示。该图是在固定训练样本和测试样本数的高斯关联数都一样的前提下,高斯关联数和系统建模时间的关系图。

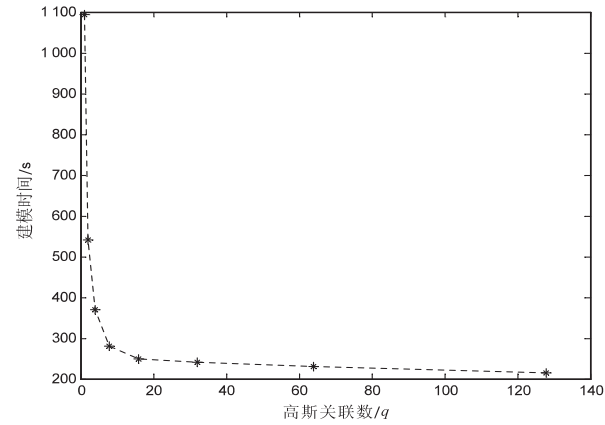


图 4 系统建模时间和高斯关联数的关系

从图 4 可看出,高斯关联数越大,系统建模所需的时间越短。实际应用中应综合考虑系统识别率和建模时间两个因素来选择高斯关联数,高斯关联数越大,建模时间越短,但有可能会降低系统的识别率,应折中选择。实验中当高斯关联数为 16 时,系统识别率最高,此时系统建模时间为 248.65 s;而在使用原始超矢量的系统中,系统识别率最低,建模所需时间最长,为 1 094.3 s。使用所提出的重构超矢量的系统,在达到最大识别率的前提下,建模时间比原始系统的建模时间缩短了 4.4 倍。

不同高斯关联数的重组超矢量具有的均值矢量样本个数不同,以此重组超矢量训练得到的 SVM 模型的支持向量个数也不同,图 5 是支持向量个数和高斯关联数的关系图。

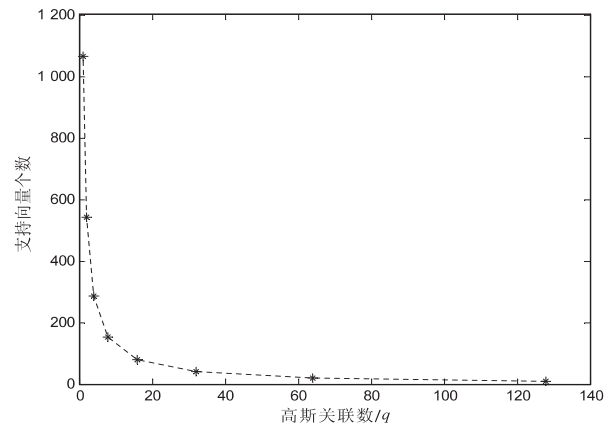


图 5 支持向量个数和高斯关联数的关系

从图 5 可以看出,随着高斯关联数的增加,SVM 模型的支持向量个数呈下降趋势。高斯关联数和重组超矢量的均值矢量的样本个数呈反比,高斯关联数增加,样本个数减少,即样本个数的减少,可以使得支持向量的个数减少。文献[19]指出,支持向量个数的减少,可以减少 SVM 的计算复杂度,即所提出的重组超矢量可以减少系统识别说话人身份所需的时间。

3 结束语

针对基本的 GMM-SVM 说话人辨认系统采用的超矢量没有考虑到各个高斯分量的均值矢量之间的关联性的问题,同时为了充分利用 SVM 处理高维小数据的优越性能,提出了超矢量的重组方法。根据高斯混合模型的混合数 M 关联一定范围内的高斯分量的均值矢量,使其首尾相连构成新的均值矢量,遍历整个超矢量就形成了重组超矢量,利用效果最好的径向基核函数将代表了说话人身份信息的重组超矢量映射到高维空间进行分类和识别。实验结果表明,选择合适的高斯关联数,可以有效提高系统的识别率,并极大地改善了系统建模的时间复杂度。

参考文献:

- [1] Squires B, Sammut C. Automatic speaker recognition: an application of machine learning [C]//Proceedings of the twelfth international conference on machine learning. Tahoe City, California, USA: [s. n.], 2012: 515–521.
- [2] Evans N, Kinnunen T, Yamagishi J, et al. Speaker recognition anti-spoofing [M]//Handbook of biometric anti-spoofing. London: Springer, 2014: 125–146.
- [3] Narayanaswamy B, Gangadharaiah R. Extracting additional information from Gaussian mixture model probabilities for improved text-independent speaker identification [C]//Proceedings of international conference on acoustics, speech, and signal processing. Philadelphia, USA: [s. n.], 2005: 621–624.
- [4] Larcher A, Kong A L, Ma B, et al. Text-dependent speaker verification: classifiers, databases and RSR2015 [J]. Speech Communication, 2014, 60(3): 56–77.
- [5] Hébert M. Text-dependent speaker recognition [M]. Berlin: Springer, 2008.
- [6] Gajšek R, Mihelič F, Dobrišek S. Speaker state recognition using an HMM-based feature extraction method [J]. Computer

- Speech & Language, 2013, 27(1): 135–150.
- [7] Reynolds D A, Quatieri T F, Dunn R B. Speaker verification using adapted gaussian mixture models [J]. Digital Signal Processing, 2000, 10(1–3): 19–41.
 - [8] Campbell W M. Generalized linear discriminant sequence kernels for speaker recognition [C]//IEEE international conference on acoustics, speech, and signal processing. [s. l.]: IEEE, 2002.
 - [9] Solomonoff A, Quillen C, Campbell W M. Channel compensation for SVM speaker recognition [C]//IEEE international conference on acoustics, speech, and signal processing. [s. l.]: IEEE, 2010: 629–632.
 - [10] 王波, 徐毅琼, 李弼程. 基于 SVM 的多分类器融合算法在说话人识别中的应用 [J]. 计算机工程与设计, 2007, 28(12): 2909–2910.
 - [11] 张振领, 贾仰理. 考虑性别差异基于 SVM 的说话人识别研究 [J]. 计算机工程与设计, 2008, 29(6): 1516–1518.
 - [12] 邱政权, 尹俊勋, 薛丽萍. 基于因子分析信道失配补偿的 SVM 话者确认方法 [J]. 自动化学报, 2006, 32(5): 753–759.
 - [13] You C H, Lee K A, Li H. GMM-SVM kernel with a Bhattacharyya-based distance for speaker recognition [J]. IEEE Transactions on Audio Speech & Language Processing, 2010, 18(6): 1300–1312.
 - [14] Li Q, Liu Y H. SVM-GMM based speaker identification [J]. Advanced Materials Research, 2014, 1044–1045: 1370–1374.
 - [15] You C H, Kong A L, Li H. An SVM kernel with GMM-supervector based on the Bhattacharyya distance for speaker recognition [J]. IEEE Signal Processing Letters, 2009, 16(1): 49–52.
 - [16] Yessad D, Amrouche A. SVM based GMM supervector speaker recognition using LP residual signal [C]//International conference on image and signal processing. [s. l.]: [s. n.], 2012: 579–586.
 - [17] Reynolds D A. Experimental evaluation of features for robust speaker identification [J]. IEEE Transactions on Speech & Audio Processing, 1994, 2(4): 639–643.
 - [18] 曹洁, 余丽珍. 基于 MFCC 和运动强度聚类初始化的多说话人识别 [J]. 计算机应用研究, 2012, 29(9): 3295–3298.
 - [19] Burges C J C. A tutorial on support vector machines for pattern recognition [J]. Data Mining & Knowledge Discovery, 1998, 2(2): 121–167.