doi:10.11959/j.issn.1000-436x.2015014

基于变分贝叶斯学习的音频水印盲检测方法

唐鑫[1]，马兆丰[1,2]，钮心忻[1]，杨义先[1]
(1. 北京邮电大学 信息安全中心，北京 100876；2. 北京国泰信安科技有限公司，北京 100086)

摘 要：为了提高音频水印的检测性能，基于音频帧 MFCC 特征的统计特性，提出了一种音频水印盲检测方法。在音频帧的 DCT 系数上嵌入扩频水印，对嵌入水印的音频帧和原始音频帧分别提取 MFCC 特征进行训练，分别建立高斯混合模型，并通过变分贝叶斯学习方法估计出高斯混合模型的参数，检测时依据最大似然的原则。实验结果显示提出的方法在音频信号受到噪声干扰和恶意攻击的情况下，相对基于 EM 算法的方法在误检率上有明显降低，在小样本训练情况下具有更好的效果并且可以有效避免过拟合的问题。

关键词：高斯混合模型；音频水印；盲检测；过拟合
中图分类号：TP309 **文献标识码**：A

Blind audio watermarking mechanism based on variational Bayesian learning

TANG Xin[1], MA Zhao-feng[1,2], NIU Xin-xin[1], YANG Yi-xian[1]
(1. Information Security Center, Beijing University of Posts and Telecommunications, Beijing 100876, China;
2. Beijing National Security Science and Technology Co., Ltd., Beijing 100086, China)

Abstract: In order to improve the performance of audio watermarking detection, a blind audio watermarking mechanism using the statistical characteristics based on MFCC features of audio frames was proposed. The spread spectrum watermarking was embedded in the DCT coefficients of audio frames. MFCC features extracted from watermarked audio frames as well as un-watermarked ones were trained to establish their Gaussian mixture models and to estimate the parameters by vatiational Bayesian learning method respectively. The watermarking was detected according to the maximum likelihood principle. The experimental results show that our method can lower the false detection rate compared with the method using EM algorithm when the audio signal was under noise and malicious attacks. Also, the experiments show that the proposed method achieves better performance in handling insufficient training data as well as getting rid of over-fitting problem.

Key words: Gaussian mixture model; audio watermarking; blind detection; over-fitting

1 引言

随着互联网技术的飞速发展，多媒体信息的传播变得迅捷而普遍，数字形式的多媒体信息为盗版提供了便利，恶意盗版者只需复制拷贝原始的数字文件，就可以通过非法传播获得利益，进而损坏版权所有者的权益。数字水印技术通过在数字内容中加入可见或不可见的水印，用来标识数字内容的版权归属，已经成为数字作品版权保护的一种主要手段[1,2]。

音频数字水印的盲检测方法可以分为基于线性相关的检测方法和基于高斯混合模型的检测方法，前者通过计算嵌入了数字水印的信号与伪随机序列的相关性，并与阈值相比较来提取数字水印。线性相关检测法没有考虑宿主信号和水印信号的相关性，也没有对信道噪声作相应考虑，因此，在有噪声的信道环境下，水印检测的性能并不理想。文献[3]通过对音频的左右声道进行相关检测来实现水印的提取，使用 MFCC 作为特征，给出了水印

收稿日期：2013-07-30；**修回日期**：2014-04-01
基金项目：国家自然科学基金资助项目（60803157, 90812001, 61170271, 61272519）
Foundation Item: The National Natural Science Foundation of China (60803157, 90812001, 61170271, 61272519)

提取的算法，然而算法在音频信号受到较强攻击的情况下性能不够理想。文献[4, 5]提出基于广义高期模型的水印检测方法，由于模型不够精细，所以检测性能不够理想。在此基础上，文献[6, 7]分别使用高斯混合模型对音频的 DFT 和 DCT 域建模，通过检测音频帧对于不同参数下高斯混合模型的似然度来实现水印的盲检测。由于高斯混合模型具有很好的局部自适应特性，在噪声环境下具有良好的性能，所以广泛地用于数字水印的盲检测。

文献[6, 7]使用 EM 算法[8]对高斯混合模型进行求解，通过对大量的音频帧进行监督学习，估计出高斯混合模型的参数，然后再计算嵌入水印的音频帧对于各个高斯混合模型的似然度，以此来检测嵌入的数字水印信息。EM 算法是求解高斯混合模型的一个常用方法，然而，由于 EM 算法不能自适应地匹配高斯混合模型的混合度，而是需要预先设定，所以会出现欠学习和过学习的问题。针对这一问题，学术界提出了使用交叉认证 EM 算法(EM, cross-validation)[9]取代 EM 算法求解高斯混合模型，这种改进的 EM 算法将原始的训练数据分成了多个子集，除了一个子集外的所有子集都用作模型训练，而剩下这个子集的数据被划分到不同的训练集合中，重复这个训练过程，从而找到最合适的数据分类方式。交叉认证方法使用交叉认证的方式来决定何时应该停止训练，所以可以有效避免欠学习和过学习问题。然而，这种交叉认证的方式对小数据量的情况效果不够理想[10]，并且计算开销很高。虽然交叉认证 EM 算法可以自适应地匹配高斯混合模型的混合度，但是往往对于混合度的估计不够准确。同时由于计算代价的问题，该算法仅仅适用于估计1~2 个未知参数的情况[11]。文献[12]研究了扩频水印技术在多媒体文件中的应用，通过实验验证了扩频水印良好的抗攻击性，文献[13]使用音频信号作为载体，提出一种扩频水印方案，文献[14]探讨了扩频水印的盲检测机制，实验结果表明，对于音频水印载体来说，使用扩频水印方案可以显著地提高检测准确率。在以上工作的基础上，研究基于高分贝叶斯学习[15]的音频水印盲检测方法，在音频的 DCT 域嵌入扩频水印，对于嵌入水印后的 MFCC 系数建立高斯混合模型，检测水印时，计算检测样本对于各个高斯混合模型的似然度，依据最大似然原理，就可以检测出水印信息。使用变分贝叶斯方法估计高斯混合模型的参数[15]，克服了 EM 算法不能自适应地确定高斯混合模型的混合度的问题，实验结果表明，在小样本训练情况下能够取得更好的检测效果，并且能够有效避免过拟合问题。

2 水印嵌入

在水印嵌入时，首先对水印比特进行扩频调制，然后嵌入到音频信号的 DCT 域中[7]。将 0 bit 和 1 bit 分别调制成长度为 n 的比特串，0 bit 可以映射成 $RN_0=\{r_0(i)\,|\,i=0,1,\cdots,n\}$，1 bit 可以映射成 $RN_1=\{r_1(i)\,|\,i=0,1,\cdots,n\}$，在这 2 个序列中，$r_0(i),r_1(i)\in\{-1,1\}$。将一段固定长度的音频划分为长度为 L 的帧，然后对每一帧分别做 DCT 变换，在变换后每一帧中找到 n 个 DCT 中频系数，嵌入扩频序列 RN_0 或 RN_1，嵌入公式为

$$F(i)=F_0(i)(1+\alpha r_m(i)) \tag{1}$$

其中，$F(i)$ 是嵌入了扩频序列后的 DCT 系数，$F_0(i)$ 是嵌入前的 DCT 系数，在 $r_m(i)$ 中，$m\in\{0,1\}$，只对应 0 bit 的扩频序列和 1 bit 的扩频序列 2 种情况，$i\in\{1,2,\cdots,n\}$，α 表示嵌入强度，取值越大，嵌入的顽健性就越强，同时对原始音频造成的失真也会越大。对于一个扩频序列，使用 n 次式(1)，直到对应的 n 个 DCT 系数都做了对应的修改，则完成了一个扩频序列的嵌入。

3 水印检测

3.1 高斯混合模型的建立

本文采用由多个高斯分布线性组合而成的有限阶的高斯混合模型对音频帧的Mel频率倒谱系数(MFCC, mel frequency cepstrum coefficient)[16]建模，可以更好地描述音频帧的局部特征，从而在水印检测时有效地提高识别准确度。

水印的嵌入分为 3 种情况，即嵌入 0 bit 对应的扩频序列，嵌入 1 bit 对应的扩频序列和未嵌入。本文对这 3 种情况音频帧的 MFCC 系数分别建立 3 个高斯混合模型，在检测一个新的音频帧嵌入的水印情况时，只需计算该音频帧对于 3 个高斯混合模型的似然度，最大似然对应的那种情况即为嵌入的情况。3 种情况得到的高斯混合模型如式(2)所示[6,7]

$$p(x\mid\pi_i,\mu_i,\Lambda_i)=\sum_{k=1}^{K_i}\pi_{i,k}N(x\mid\mu_{i,k},\Lambda_{i,k}^{-1}) \tag{2}$$

其中，$i\in\{0,1,2\}$，i 取 0 对应嵌入 0 的情况，i 取 1 时对应嵌入 1 的情况，而 i 取 2 时，对应未嵌入的情况。这 3 种情况各对应一组参数$\{K_i,\boldsymbol{\pi}_i,\boldsymbol{\mu}_i,\boldsymbol{\Lambda}_i\}$，其中，$K_i$ 表示高斯混合模型的混合度，$\boldsymbol{\pi}_i$ 表示每一个高斯分量在高斯混合模型中的权重，称为混合系数，$\boldsymbol{\mu}_i$ 和 $\boldsymbol{\Lambda}_i$ 分别表示高斯分量的均值和精度矩阵[15,17]，$\boldsymbol{\Lambda}_i^{-1}=\sum_i$，假定高斯混合模型由 K 个高斯分量线性组合而成，即模型的混合度为 K，则有

$$\begin{cases}\boldsymbol{\pi}_i=\boldsymbol{\pi}_{i,k}\\ \boldsymbol{\mu}_i=\boldsymbol{\mu}_{i,k}\\ \boldsymbol{\Lambda}_i=\Lambda_{i,k}\end{cases} \tag{3}$$

对式(3)有 $k\in\{1,2,\cdots,K\}$，混合系数必须满足

$$\sum_{k=1}^{K}\pi_{i,k}=1,i\in\{0,1,2\} \tag{4}$$

至此，3 种情况对应的高斯混合模型已经建立起来，3 个模型各自的混合度可能相同，也可能不同，这里没有必然联系，具体要视 3 种情况下高斯混合模型的求解而定。值得注意的是，采用音频帧的 MFCC 系数建模是因为其能够很好地表示音频帧的特性，并且，在样本训练的过程中，能够有效地避免高维样本的降维问题。

3.2　高斯混合模型的求解

针对 3.1 节建立起来的高斯混合模型，需要通过求解估计出参数$\{K,\boldsymbol{\pi},\boldsymbol{\mu},\boldsymbol{\Lambda}\}$。变分贝叶斯期望最大化(VBEM, variational Bayesian expectation maximization)算法求解高斯混合模型[15]，从而自适应地确定模型的混合度，同时有效地改进了传统 EM 算法存在的欠学习和过拟合问题，进而提高音频水印检测时的识别率。

定义数据集 $X=\{x_1,x_2,\cdots,x_N\}$ 表示音频样本每一帧的 MFCC 系数，隐变量 $Z=\{z_1,z_2,\cdots,z_N\}$ 表示在高斯混合模型中，每一帧音频样本对于对应高斯分量的归属情况，集合 Z 中的每一个元素 $z_k\in\{0,1\}$ 并且 $\sum_k z_k=1$。考虑一个音频训练样本 MFCC 特征的集合 X，其条件概率密度可以写成[15]

$$p(X\mid Z,\boldsymbol{\mu},\boldsymbol{\Lambda})=\prod_{n=1}^{N}\prod_{k=1}^{K}N(x_n\mid\mu_k,\Lambda_k^{-1})^{z_{nk}} \tag{5}$$

其中，$\boldsymbol{\mu}=\{\mu_k,k=1,2,\cdots,K\}$ 和 $\boldsymbol{\Lambda}=\{\Lambda_k,k=1,2,\cdots,K\}$ 分别表示均值向量和精度矩阵。隐变量 z_{nk} 表示训练样本中每一帧的 MFCC 特征对于混合模型中 k 个高斯分量的归属情况，并且有 $\sum_{k=1}^{K}z_{nk}=1,(n=1,2,\cdots,N)$。

给定混合系数，隐变量的条件分布可以写成[11]

$$p(Z\mid\pi)=\prod_{n=1}^{N}\prod_{k=1}^{K}\pi_k^{z_{nk}} \tag{6}$$

高斯混合模型中的混合系数 $\boldsymbol{\pi}$ 的先验分布服从 Dirichlet 分布[18]

$$p(\pi)=\mathrm{Dir}(\pi\mid\alpha_0)=C(\alpha_0)\prod_{k=1}^{K}\pi_k^{\alpha_0-1} \tag{7}$$

均值 μ 和精度矩阵 $\boldsymbol{\Lambda}$ 的共轭先验分布服从 Gaussian-Wishart 分布[15, 19]

$$\begin{aligned}p(\mu,\boldsymbol{\Lambda})&=p(\mu|\boldsymbol{\Lambda})p(\boldsymbol{\Lambda})\\&=\prod_{k=1}^{K}N(\mu_k\mid\mathrm{m}_0,(\beta_0V_k)^{-1})W(\Lambda_k\mid W_0,v_0)\end{aligned} \tag{8}$$

其中，目标是在给定训练样本 MFCC 特征的情况下估计出高斯混合模型的参数，因此需要计算后验分布 $p(Z,\mu,\boldsymbol{\Lambda}\mid X,\pi)$ $p(Z,\mu,\Lambda\mid X,\pi)$。VBEM 算法引入分布 $q(Z,\mu,\boldsymbol{\Lambda})$ 来逼近真实的后验分布[15]。考虑对数边缘似然 $\ln p(X\mid\pi)$。

$$\begin{aligned}&\ln p(X\mid\pi)=\ln\iiint p(X,Z,\mu,\boldsymbol{\Lambda}\mid\pi)\mathrm{d}Z\mathrm{d}\mu\mathrm{d}\boldsymbol{\Lambda}\\&=\ln\iiint q(Z,\mu,\boldsymbol{\Lambda})\frac{p(X,Z,\mu,\boldsymbol{\Lambda}\mid\pi)}{q(Z,\mu,\boldsymbol{\Lambda})}\mathrm{d}Z\mathrm{d}\mu\mathrm{d}\boldsymbol{\Lambda}\\&=\iiint q(Z,\mu,\boldsymbol{\Lambda})\ln\frac{p(X,Z,\mu,\boldsymbol{\Lambda}\mid\pi)}{q(Z,\mu,\boldsymbol{\Lambda})}\mathrm{d}Z\mathrm{d}\mu\mathrm{d}\boldsymbol{\Lambda}+\\&\quad(-\iiint q(Z,\mu,\boldsymbol{\Lambda})\ln\{\frac{p(Z,\mu,\boldsymbol{\Lambda}\mid X,\pi)}{q(Z,\mu,\boldsymbol{\Lambda})}\}\mathrm{d}Z\mathrm{d}\mu\mathrm{d}\boldsymbol{\Lambda})\\&=L(q)+KL(q\parallel p)\end{aligned} \tag{9}$$

在式(9)中，$L(q)$ 是对数边缘似然的下界，又称为变分下限。$KL(q\parallel p)$ 称为 $q(Z,\mu,\boldsymbol{\Lambda})$ 和真实的后验分布 $p(Z,\mu,\boldsymbol{\Lambda}\mid X,\pi)$ 之间的 Kullback-Leibler 散度[15,20]。

由于 $KL(q\parallel p)\geqslant 0$，所以要逼近 $\ln p(X\mid\pi)$，必须选定合适的分布 $q(Z,\mu,\boldsymbol{\Lambda})$ 使变分下限 $L(q)$ 最大化。根据式(9)，最大化 $L(q)$ 等价于最小化 $KL(q\parallel p)$，$KL(q\parallel p)$ 的最小值是 0，在式(9)中可以看出，当 $q(Z,\mu,\boldsymbol{\Lambda})=p(Z,\mu,\boldsymbol{\Lambda}\mid X,\pi)$ 时，$KL(q\parallel p)=0$，所以从理论上说，当变分下限 $L(q)$ 取到最大值时，$q(Z,\mu,\boldsymbol{\Lambda})$ 取到真实后验分布。所以

VBEM 算法考虑找到一个分布函数 $q(Z,\mu,\boldsymbol{\Lambda})$ 的族，通过最大化变分下限，让这一族函数来逼近真实后验分布。

考虑变分分布 $q(Z,\mu,\boldsymbol{\Lambda})$，可以写成下面的分解形式[15]

$$q(Z,\mu,\boldsymbol{\Lambda})=q(Z)q(\mu,\boldsymbol{\Lambda})=\prod_{k=1}^{M}q(Z_k)q(\mu_k,\Lambda_k) \tag{10}$$

VBEM 算法是一种迭代算法，每一次的迭代都分为 2 个步骤，即变分贝叶斯期望(VB-E，variational Bayesian-expectation)步骤和变分贝叶斯最大化(VB-M,variational Bayesian-maximization)步骤，具体步骤的执行如下。

VB-E：将式(10)中分解的变分分布形式代入式(9)变分下限的表达式，可得

$$L(q)=\iiint q(Z)q(\mu,\boldsymbol{\Lambda})\ln\frac{p(X,Z,\mu,\boldsymbol{\Lambda}|\pi)}{q(Z)q(\mu,\boldsymbol{\Lambda})}\mathrm{d}Z\mathrm{d}\mu\mathrm{d}\boldsymbol{\Lambda} \tag{11}$$

求解 $\partial L(q)/\partial q(Z)$，并令偏导的取值为 0，可以得到

$$q^{(t+1)}(Z)=\frac{1}{NC_Z}\exp[\iint q^{(t)}(\mu,\boldsymbol{\Lambda})\ln p(X,Z|\mu,\boldsymbol{\Lambda},\pi)\mathrm{d}\mu\mathrm{d}\boldsymbol{\Lambda}] \tag{12}$$

其中，NC_Z 是标准化常量。

VB-M：由式(11)求解 $\partial L(q)/\partial q(\mu,\boldsymbol{\Lambda})$，并令偏导的取值为 0，可以得到[10,15]

$$q^{(t+1)}(\mu,\boldsymbol{\Lambda})=\frac{1}{NC_{\mu,\boldsymbol{\Lambda}}}p(\mu,\boldsymbol{\Lambda}|\pi)\cdot \exp[\int q^{(t+1)}(Z)\ln p(X,Z|\mu,\boldsymbol{\Lambda},\pi)\mathrm{d}Z] \tag{13}$$

其中，$NC_{\mu,\boldsymbol{\Lambda}}$ 是标准化常量。

VB-E 和 VB-M 步骤交替进行，随着迭代的不断重复，变分下限 $L(q)$ 逐渐增大，直到 $|L^{(t+1)}(q)-L^{(t)}(q)|<\varepsilon$，迭代终止，其中 ε 是误差限。

在 VB-E 步骤中，得到的隐变量变分分布的对数形式如下[15]

$$\ln q^{(t+1)}(Z)=\sum_{n=1}^{N}\sum_{k=1}^{K}z_{nk}\ln\rho^{(t)}{}_{nk}+\mathrm{const} \tag{14}$$

其中

$$\rho_{nk}=\frac{1}{2}E[\ln|\Lambda_k|]-\frac{D}{2}\ln 2\pi-\frac{1}{2}E_{\mu,\boldsymbol{\Lambda}}(x_n-\mu_k)^{\mathrm{T}}\Lambda_k(x_n-\mu_k) \tag{15}$$

定义 $r_{nk}=\rho_{nk}/\sum_{j=1}^{K}\rho_{nk}$，$N_k=\sum_{n=1}^{N}r_{nk}$，则 VB-E 步骤中参数按下式进行更新[10]

$$\mu_k=\frac{1}{N_k}\sum_{n=1}^{N}r_{nk}x_n$$
$$\Lambda_k^{-1}=\frac{1}{N_k}\sum_{n=1}^{N}r_{nk}(x_n-\mu_k)(x_n-\mu_k)^{\mathrm{T}} \tag{16}$$

在 VB-M 步骤中，将 $q(\mu_k,\Lambda_k)$ 写成如下形式

$$q(\mu_k,\Lambda_k)=q(\mu_k|\Lambda_k)q(\Lambda_k) \tag{17}$$

$q(\mu_k,\Lambda_k)$ 服从 Gaussian-Wishart 分布[21]，即

$$q(\mu_k,\Lambda_k)=N(\mu_k|m_k,(\beta_k\Lambda_k)^{-1})W(\Lambda_k|W_k,\nu_k) \tag{18}$$

VB-M 步骤中的参数按照下式进行更新[10,15]。

$$\beta_k=\beta_0+N_k$$
$$m_k=\frac{1}{\beta_k}(\beta_0 m_0+N_k\mu_k)$$
$$\nu_k=\nu_0+N_k$$
$$W_k^{-1}=W_0^{-1}+N_k\Lambda_k^{-1}+\frac{\beta_0 N_k}{\beta_0+N_k}(\mu_k-m_0)(\mu_k-m_0)^{\mathrm{T}} \tag{19}$$

根据式(19)计算均值

$$\mu_k=m_k=\frac{1}{\beta_k}(\beta_0 m_0+N_k\mu_k) \tag{20}$$

根据 Wishart 分布的性质[18]，可知

$$\mathrm{E}[\boldsymbol{\Lambda}]=\nu W \tag{21}$$

由式(7)及 Dirichlet 分布的性质，混合系数的期望表示为

$$\mathrm{E}[\pi_k]=\frac{\alpha_k+N_k}{K\alpha_0+N} \tag{22}$$

其中，集合 $\boldsymbol{\pi}$ 中的非零元素个数即为高斯混合模型的混合度 K。

3.3 基于最大似然的水印检测

基于式(2)建立的高斯混合模型，对嵌入水印的情况进行检测。对于嵌入 0、嵌入 1 以及未嵌入水印的 3 种情况的音频数据，分别提取音频帧的 MFCC 系数，作为输入，计算输入的 MFCC 系数对于 3 个高斯混合模型的似然度，最大似然对应的即为相应的水印嵌入情况。采用式(23)的准则[6,7]来判定水印的嵌入情况

$$i_m = \arg \max_{i \in \{0,1,2\}} p(x \mid \pi_i, \mu_i, \Lambda_i) \tag{23}$$

在式(23)中，i_m 的取值为 0,1,2，取 0,1 时分别对应当前音频帧被检测出嵌入水印比特 0 和 1，取 2 时对应检测出当前音频帧没有嵌入水印的情况。

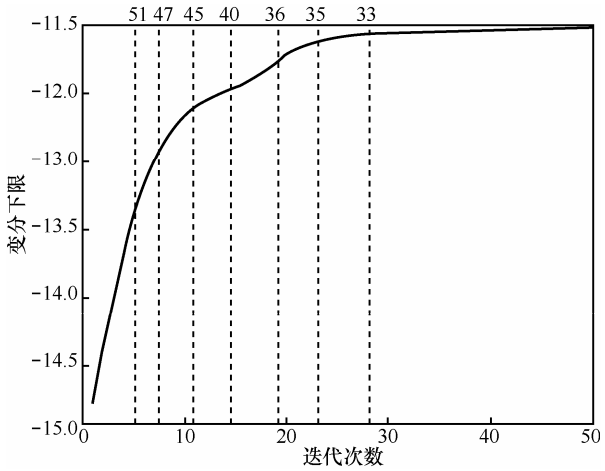
4　实验仿真与结果分析

4.1　实验背景与参数设置

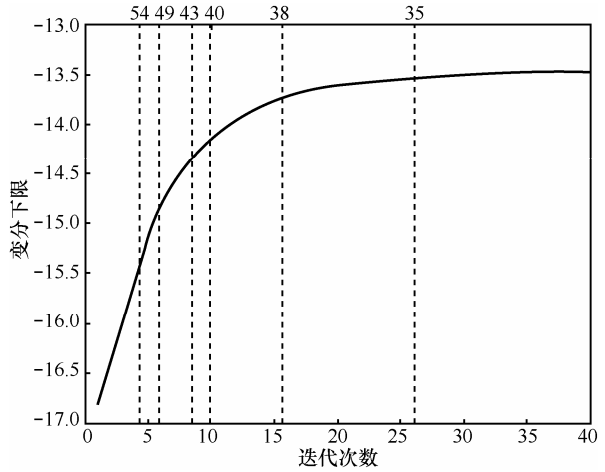
在上文对 3 种类型的音频帧分别建立高斯混合模型，使用 VBEM 算法求解高斯混合模型，进而得到标识水印嵌入情况的 3 组参数。在音频水印检测时，根据待测音频帧对 3 组参数下高斯混合模型的似然度，取最大似然对应的情况，判断水印嵌入的情况。在实验的过程中，选择采样率为 44.1 kHz，16 bit 量化，长度为 50 s 的单声道音频作为样本，同一个样本选取 3 份拷贝。首先对音频进行端点检测，去除静音段，接下来对音频进行分帧处理，取帧长为 256，考虑 3 种情况：第 1 种情况是不嵌入水印，即对音频样本不做处理；第 2 种情况是嵌入 0 的扩频序列，在嵌入时，首先对音频帧做 DCT 变换，变换后的音频帧取 DCT 中频系数嵌入，嵌入完成后进行 DCT 逆变换；第 3 种情况与第 2 种情况类似，不同之处在于此时嵌入的是 1 的扩频序列。其中，0 或 1 的扩频序列长度设定为 63，VBEM 算法的初始混合度设置为 55。音频特征提取阶段，采用 24 阶的 Mel 滤波器，提取出 12 维的 MFCC 作为特征，进行训练。如果某一帧检测出的水印情况与实际情况不符，则认为是出现了误检，在实验中，在训练样本中选出 1 000 帧作为检测样本，发生的误检的总次数除以检测样本的总数即得到误检率。

4.2　变分下限与混合度的确定

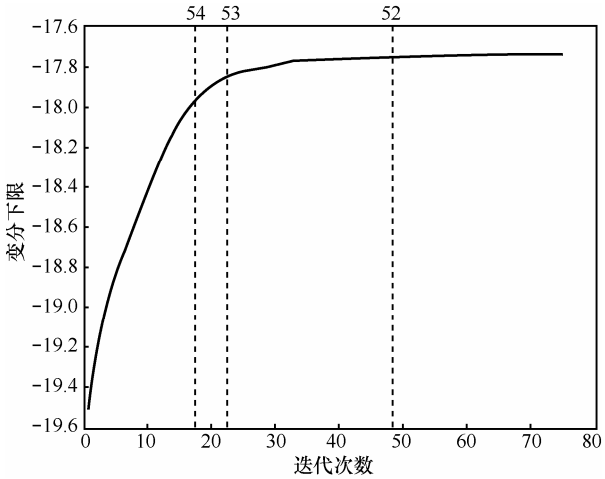
图 1 给出了嵌入 0、嵌入 1 和未嵌入水印 3 种情况下变分下限与迭代次数的关系比较，可以看到，3 种情况下，随着迭代次数的增加，变分下限都是呈上升趋势，并且高斯混合模型的混合度随着变分下限的不断增大，逐渐减小，直到减小到某个值之后，变分下限也趋于稳定，不再增加。由式(10)~式(12)可以看出，随着变分下限的逐渐增加，Kullback- Leibler 散度趋于 0，从而使变分分布逼近真实的后验分布。此时，高斯混合模型的混合度也自适应地调整到固定的值，避免了使用 EM 算法来实现音频水印盲检测时人为给定高斯混合模型的混合度引入的误差[7]。



(a) 嵌入 0 的情况



(b) 嵌入 1 的情况



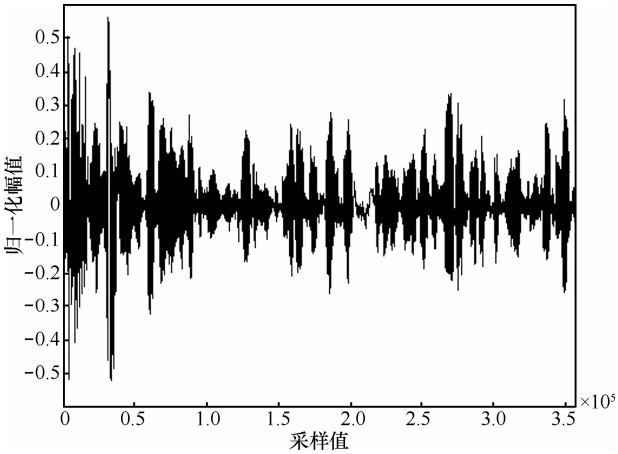
(c) 未嵌入的情况

图 1　3 种情况下变分下限与迭代次数的关系比较

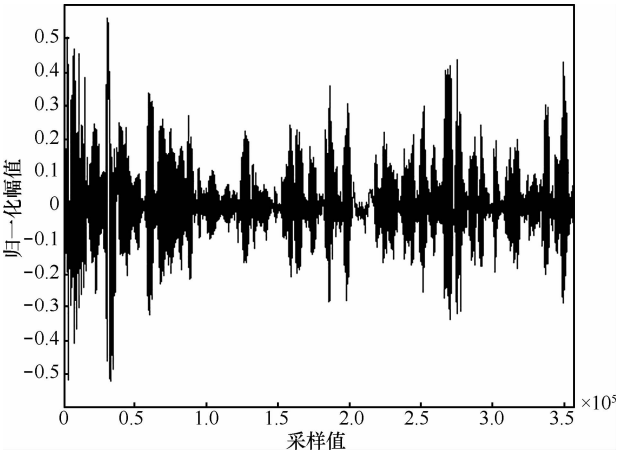
4.3　抗攻击性能

为了验证算法在抗攻击条件下检测水印时的效果，选择长度为 1 000 的 0，1 信号组成的随机序列作为水印，在训练过的音频样本中选取一段长度为 10 s

的音频，将水印嵌入其中。如图 2 所示，图 2(a)是原始语音信号，图 2(b)是嵌入了水印的语音信号。



(a) 原始音频信号



(b) 嵌入了水印的音频信号

图 2 原始音频信号与嵌入了水印的音频信号

由于文献[5~7]已经用实验证明了 EM 算法相比 SVM 和线性相关检测法在检测音频水印时的性能优势，所以主要与 EM 算法[6,7]进行比较。水印检测的性能用误检率来表示，误检率越低，则检测性能越好。水印的检测考虑在音频信号受到攻击干扰的情况下进行，在实验中采用 StirMark[22]对音频信号进行各种攻击，实验结果表明，本方法相比 EM 算法在各种攻击下，性能均有不同程度的提升。

从表 1 中可以看出，选取了 8 种典型的攻击方法进行测试，在这 8 种攻击下，VBEM 算法相比 EM 算法明显具有较低的误检率。这归因于 VBEM 算法可以自行确定高斯混合模型的混合度，从而找到一个最合适的值，而 EM 算法在实验中不能自行确定混合度，根据文献[7]，在混合度大于等于 6 时，拟合效果较好，混合度在大于 6 的时候，对数似然函数值增加缓慢，拟合效果提升不明显，因此在 EM 算法中将混合度固定为 6。此外，由于 VBEM 算法与 EM 算法在模型求解以及似然计算的过程中具有相同的复杂度，因此 2 种方法的时间开销是相同的，也就是说，VBEM 算法没有造成额外的开销。

表 1 VBEM 算法与 EM 算法在恶意攻击情况下的误检率比较

攻击类型	误检率/%	
	VBEM	EM
Addbrumm 10 100	9.5	9.9
Compressor	11.4	12.8
FFT invert	4.8	6.1
Lsbzero	4.8	6.3
Normalize	4.7	6.5
Stat1	33.3	40.1
rc_highpass	5.0	6.4
Zerocross	38.2	40.1

4.4 小样本学习情况下的检测性能

为了比较在训练样本减少的情况下，本方法与 EM 算法在检测性能上的差异，在下面的实验中，把用于训练的音频样本长度减小，观察在不同情况下，2 种方法的检测性能差异。从图 3 中可以看出，训练样本长度从 50 s 减小到 20 s 的过程中，2 种算法的误检率都呈逐渐增大趋势，并且 VBEM 算法下的误检率始终低于 EM 算法下的误检率，随着训练样本长度的不断减小，VBEM 算法的误检率缓慢增大，而 EM 算法的误检率增大较快。这就表明，在小样本训练的情况下，VBEM 算法对于音频水印的识别率明显优于 EM 算法。

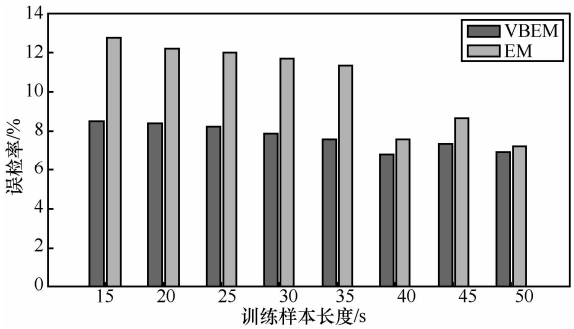


图 3 VBEM 算法与 EM 算法在不同长度训练样本下误检率比较

4.5 训练样本长度不一致情况下检测性能

在实验的过程中对嵌入 0、嵌入 1、未嵌入水

印 3 种情况的音频帧建立了高斯混合模型，在之前的实验中，建立每个高斯混合模型时用到的训练样本长度是相同的，即同时使用 50 s 的音频样本去训练 3 个模型，或者同时减小训练样本的长度。在下面的实验中，将考虑采用不同长度的训练样本，去分别训练 3 个高斯混合模型。为了便于比较，对于嵌入 1 和未嵌入水印的情况，固定训练样本长度为 50 s，而仅改变嵌入 0 时的训练样本长度，本算法与 EM 算法在检测性能上的差异变化。

从图 4 可以看到，仅改变嵌入 0 时的训练样本长度，当训练样本长度减小时，VBEM 算法在检测水印时相比 EM 算法具有更好的性能。这可以得出 2 个结论，首先，正如上一个实验得出的结论，提出的水印检测方法在小样本训练的情况下具有更好的检测性能。其次，检测方法相比 EM 算法，对于过拟合问题具有相对更好的健壮性，过拟合问题通常发生在这种训练样本长度不一致的情况下。在下面的实验中，将重点考察过拟合问题对 2 个方法的影响。

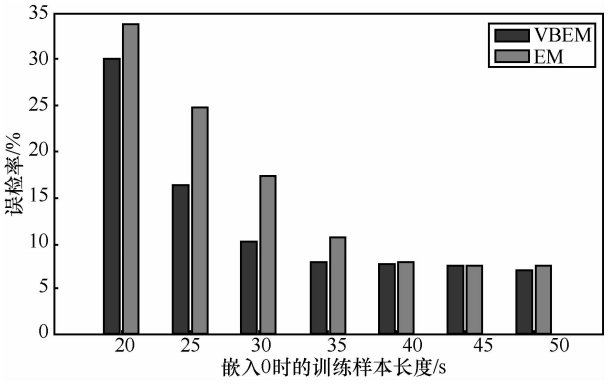


图 4　VBEM 算法与 EM 算法在 0 的训练样本长度改变情况下误检率比较

4.6　过拟合问题下的健壮性

过拟合问题是影响 EM 算法性能的一个主要因素，它是指当求解出的高斯混合模型与训练样本高度符合的时候，检测训练样本中嵌入的音频水印识别率很高，而一旦训练样本被稍加改变，检测性能立刻就会下降。在音频水印盲检测场景中，这种性能的下降体现在测试样本对于训练出的高斯混合模型的似然度降低。

在这部分实验中，通过改变 VBEM 算法与 EM 算法的误差限，在不同的误差限情况下计算出测试样本的对数似然值，来评判 2 种算法受过拟合的影响。从图 5 可以看出，使用 EM 算法的情况下，测试样本的对数似然值随着误差限的减小，首先增大，然后一直减小，这是由于随着误差限的减小，模型的训练越来越精细，出现了过拟合。而在使用 VBEM 算法的情况下，测试样本的对数似然值保持缓慢的增加，并且曲线相对更加平稳。这就表明了本方法能够有效地避免过拟合问题带来的危害。

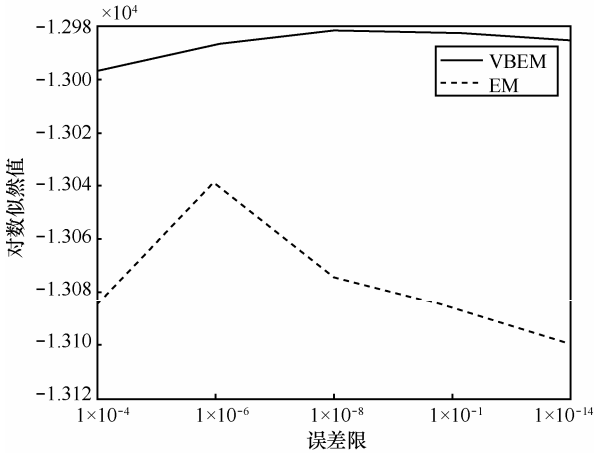


图 5　VBEM 算法与 EM 算法在不同误差限下测试样本对数似然值比较

5　结束语

本文研究了基于变分贝叶斯学习的音频水印盲检测方法。用 VBEM 算法求解高斯混合模型，从而自适应地求得模型的混合度，并解决欠学习和过拟合问题。实验结果表明本方法可以有效地提高音频水印检测的准确率，在音频信号受到噪声干扰和恶意攻击的情况下，相对基于 EM 算法的检测方法，本方法在误检率上有明显降低，并且本方法在小样本训练和过拟合情况下具有相对更好的性能。

参考文献：

[1] 马兆丰, 范科峰, 陈铭等. 支持时空约束的可信数字版权管理安全许可协议[J]. 通信学报, 2008, 29(10): 153-163.
MA Z F, FAN K F, CHEN M, *et al*. Trusted digital rights management protocol supporting for time and space constraint[J]. Journal on Communications, 2008,29(10):153-163.

[2] 蒋铭, 马兆丰, 辛宇等. 基于 DWT 和视觉加权的图像质量评价方法研究[J]. 通信学报, 2011, 32(9): 129-136.
JIANG M, MA Z F, XIN Y, *et al.* Image quality evaluation method base on digital wavelet transform and vision weighted[J]. Journal on Communications, 2011,32(9):129-136.

[3] 冯涛, 韩纪庆. 双声道音频水印的同步及盲检测算法[J]. 通信学报, 2006, 27(10): 62-68.
FENG T, HAN J Q. Synchronization and blind detect algorithm for dual channel audio watermark[J]. Journal on Communications, 2006,

27(10):62-68.

[4] 孙中伟, 朱岩, 冯登国. DCT 域图像水印的局部优化检测性能研究[J]. 电子学报, 2005, 33(5):864-867.
SUN Z W, ZHU Y, FENG D G. Performance analysis of DCT-domain watermark detection based on local optimum detection[J]. Acta Electronica Sinica, 2005, 33(5):864-867.

[5] AKHAEE M A, SAHRAEIAN S M E, MARVASTI F. Contourlet-based image watermarking using optimum detector in a noisy environment[J]. IEEE Transactions on Image Processing, 2010, 19(4): 967-980.

[6] GUNSEL B, ULKER Y, KIRBIZ S. A statistical framework for audio watermark detection and decoding[A]. Multimedia Content Representation, Classification and Security[C]. Springer Berlin Heidelberg, 2006. 241-248.

[7] 林晓丹. 基于高斯混合模型的 DCT 域水印检测方法[J]. 自动化学报, 2012,38(9):1445-1448.
LIN X D. DCT-domain watermark detection using Gaussian mixture model[J]. Acta Automatica Sinica, 2012, 38(9):1445-1448.

[8] DEMPSTER A P, LAIRD N M, RUBIN D B. Maximum likelihood from incomplete data via the EM algorithm[J]. Journal of the Royal Statistical Society. Series B (Methodological), 1977,39(1):1-38.

[9] SHINOZAKI T, OSTENDORF M. Cross-validation EM training for robust parameter estimation[A]. Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal, ICASSP 2007[C]. 2007. IV-437-IV-440.

[10] MOATTAR M H, HOMAYOUNPOUR M M. Text-independent speaker verification using variational Gaussian mixture model[J]. ETRI Journal, 2011, 33(6):914-923.

[11] CORDUNEANU A, BISHOP C M. Variational bayesian model selection for mixture distributions[A]. Artificial Intelligence and Statistics[C]. Waltham, MA: Morgan Kaufmann, 2001.27-34.

[12] COX I J, KILIAN J, LEIGHTON F T, *et al*. Secure spread spectrum watermarking for multimedia[J]. IEEE Transactions on Image Processing, 1997, 6(12): 1673-1687.

[13] KIROVSKI D, MALVAR H S. Spread-spectrum watermarking of audio signals[J]. IEEE Transactions on Signal Processing, 2003, 51(4): 1020-1033.

[14] MALIK H, KHOKHAR A, ANSARI R. Improved watermark detection for spread-spectrum based watermarking using independent component analysis[A]. Proceedings of the 5th ACM Workshop on Digital Rights Management[C]. ACM, 2005.102-111.

[15] BISHOP C M, NASRABADI N M. Pattern Recognition and Machine Learning[M]. New York: Springer, 2006.

[16] WEYCHAN R, MARCINIAK T. Analysis of differences between MFCC after multiple GSM transcodings[J]. Przegląd Elektrotechniczny Selected Full Texts, 2012, 88(6): 24-29.

[17] BESSON O, BIDON S, TOURNERET J Y. Covariance matrix estimation with heterogeneous samples[J]. IEEE Transactions on Signal Processing, 2008, 56(3): 909-920.

[18] GÖRÜR D, RASMUSSEN C E. Dirichlet process Gaussian mixture models: choice of the base distribution[J]. Journal of Computer Science and Technology, 2010, 25(4): 653-664.

[19] MURPHY K P. Conjugate Bayesian Analysis of the Gaussian Distribution[R]. Technical Report, UBC, 2007.

[20] DO M N, VETTERLI M. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance[J]. IEEE Transactions on Image Processing, 2002, 11(2): 146-158.

[21] SHRIBERG E, FERRER L, KAJAREKAR S, *et al.* Modeling prosodic feature sequences for speaker recognition[J]. Speech Communication, 2005, 46(3): 455-472.

[22] STEINEBACH M, PETITCOLAS F A P, RAYNAL F, *et al.* StirMark benchmark: audio watermarking attacks[A]. Proceedings of the 2011 International Conference on Information Technology: Coding and Computing[C]. 2001. 49-54.

作者简介：

唐鑫（1987-），男，江苏南京人，北京邮电大学博士生，主要研究方向为数字版权管理、数字内容安全、数字水印等。

马兆丰（1974-），男，甘肃镇原人，博士，北京邮电大学讲师，主要研究方向为数字版权管理、数字内容安全、计算机网络安全。

钮心忻（1963-），女，浙江湖州人，北京邮电大学教授、博士生导师，主要研究方向为数字水印、信息隐藏、隐写分析。

杨义先（1961-），男，四川盐亭人，北京邮电大学教授、博士生导师，主要研究方向为密码学、计算机网络与信息安全。