

基于预分割的说话人分割方法

郑铁然, 李海峰, 刘先, 韩纪庆

(哈尔滨工业大学 计算机科学与技术学院, 黑龙江 哈尔滨 150001)

摘要: 对传统的基于贝叶斯信息准则(BIC)的说话人分割方法进行了改进, 通过引入预分割环节来降低说话人分割的计算量。理论分析和实验验证表明, 基于预分割的说话人分割方法在基本保证分割性能的前提下, 运算量得到了有效控制。

关键词: 说话人分割; 贝叶斯信息准则; 跳变点检测; 预分割

中图分类号: TP391

文献标识码: A

文章编号: 1000-436X(2009)02-0118-05

Method of speakers segmentation based on pre-segmentation

ZHENG Tie-ran, LI Hai-feng, LIU Xian, HAN Ji-qing

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

Abstract: The traditional method of speakers segmentation based on Bayesian information criterion (BIC) was improved. New speakers segmentation method adopted a pre-segmentation process to decrease calculation spending. Theory analysis and experiments show that the method of speakers segmentation based on pre-segmentation makes the amount of calculation controlled, on the premise of realizing the basic segmentation performance.

Key words: speaker identification; Bayesian information criterion; speaker change detection; pre-segmentation

1 引言

说话人分割技术是从一段含有多个说话人的对话型语音段中找出所有的说话人转变点, 是近些年来语音处理领域中的研究热点之一。利用说话人分割技术, 可以使传统的只能处理单一说话人的说话人识别技术应用于处理对话型语音, 它还是说话人检索技术的基础, 有着重要的应用价值。目前, 说话人分割技术的主要实现方法可以分为以下 4 大类。1) 基于端点检测的分割方法^[1], 利用说话人在讲话间隙出现停顿的特点, 检测所有的静音点作为说话人可能发生变化的潜在点。由于在不同的信噪比环境下, 静音点很难检测, 因而这类方法并不准确。2) 基于模型的分割方法, 对不同类型的语音信

号建立统计模型, 通过这些模型用极大似然准则进行分类, 进而实现跳变点的检测, 被较多采用的模型有 GMM(Gausses mixture model) 和 HMM(hidden markov model)^[2], 文献[3,4]则采用了 UBM (universal background model) 和 GMM 进行增量训练从而实现在线说话人分割的方法。3) 基于距离尺度的分割方法, 这里的距离是指每一点左右相邻的一定窗长的两段信号的距离, 如果这个距离大于设定的门限值, 则认为存在声学属性的跳变点, 常用的尺度包括 KL(Kullback-Levison)距离^[5]、一般似然比距离^[6]等。4) 基于假设检验的方法, 类似于基于尺度的分割方法, 但不是简单地计算音频数据某一点左右两段信号的距离, 而是假设某分割点左右两段的信号服从某一概率分布, 再根据一定的准则去检验

收稿日期: 2008-11-12; 修回日期: 2009-01-05

基金项目: 国家重点基础研究发展计划(“973”计划)基金资助项目(2007CB311100)

Foundation Item: The National Basic Research Program of China (973 Program) (2007CB311100)

这种假设的正确性。其中基于 BIC 假设检验方法^[7]是使用的最多也是最成功的一个。

基于 BIC 的分割方法不需要先验知识, 不需要设定门限阈值, 且具有较高的准确度, 因而为研究者所广泛采用。但这种方法在进行多分割点的说话人分割计算时, 计算量非常大, 无法满足实时处理的要求, 本文提出了一种基于预分割的说话人分割方法, 通过预先分割语音数据, 将多分割点检测问题简化为单分割点检测, 从而有效缩小了计算量。

2 基于 BIC 的说话人分割方法

2.1 BIC

Schwarz 于 1978 年提出了贝叶斯信息准则(BIC)^[8]来解决模型选取问题。设 x_1, x_2, \dots, x_n 是一列独立同分布的变量, 其概率函数为 $f(\cdot|\theta)$, 其中 f 是一个含有 k 个参数的模型函数。定义 $M(k)$ 为

$$M(k): \{f(\cdot|\theta) | \theta = (\gamma_1, \gamma_2, \dots, \gamma_k), \theta \in \Theta_k\} \quad (1)$$

其中, Θ_k 为参数空间。假设没有约束参数, 则模型中自由参数的个数为 k 。模型选取问题就是如何从 $M(k), (k=1, 2, \dots)$ 中选取一个最优拟合给定样本的模型函数 f 。

对候选模型 $f \in M(k)$, BIC 定义如下

$$B(f) = \lg L_{\hat{\theta}_k}(X) - \frac{k}{2} \lg n \quad (2)$$

其中, $L_{\hat{\theta}_k}(X)$ 是模型 $f(\cdot|\hat{\theta})$ 在样本集 X 上的极大似然函数, $\frac{k}{2} \lg n$ 是对模型复杂程度 (即模型参数数目) 的惩罚项, 在维数过大且训练样本数据相对较少的情况下可以有效避免出现维度灾难现象。在式(2)中, 使得 BIC 值最大的模型 $\hat{f} = \arg \max_{f \in M(k), k=1, 2, \dots} B(f)$ 就是在 BIC 下的最优模型。

2.2 BIC 在说话人分割中的应用

Chen 等人于 1998 年首次将 BIC 用于说话人分割^[9]。给定的样本序列 $X = \{x_i | x_i \in R^d, i=1, 2, \dots, n\}$ (R^d 为 d 维样本空间), 假定在该样本序列中要么没有说话人转变点, 要么仅有一次说话人转变点, 那么用于描述样本序列的模型就有 2 种形态: 1) 样本序列是由同一个声道模型描述的; 2) 前 k 个样本是由一个声道模型描述的, 之后的 $n-k-1$ 个样本是由另一个声道模型描述的, 并且前后 2 个声道模型不同。Chen 等在其研究中, 将语音的特征序列视为一个高斯 (Gaussian) 过程。

设没有发生说话人转变的待选模型为 f_{θ_0} , 其具体形态为

$$x_i \propto N(\mu, \Sigma), \quad i=1, \dots, n \quad (3)$$

则有

$$B(f_{\theta_0}) = \sum_{i=1}^n -\frac{1}{2} (x_i - \hat{\mu})' \hat{\Sigma}^{-1} (x_i - \hat{\mu}) - \frac{d}{2} n \lg 2\pi - \frac{1}{2} \lg |\hat{\Sigma}| - \lambda \frac{1}{2} \left(d + \frac{d(d+1)}{2} \right) \lg n \quad (4)$$

未知参数个数为 $\frac{d(d+1)}{2}$, 在样本集 X 上, 可以得到参数 μ 的极大似然估计 $\hat{\mu}$, 以及协方差矩阵 Σ 的极大似然估计 $\hat{\Sigma}$ 。

发生了说话人转变的模型为 f_{θ_1} , 其具体形态为

$$x_i \propto \begin{cases} N(\mu_1, \Sigma_1), & i=1, \dots, k \\ N(\mu_2, \Sigma_2), & i=k+1, \dots, n \end{cases} \quad (5)$$

其中, k 为相应的说话人分割点。则有

$$B(f_{\theta_1}) = \sum_{i=1}^k -\frac{1}{2} (x_i - \hat{\mu}_1)' \hat{\Sigma}_1^{-1} (x_i - \hat{\mu}_1) + \sum_{i=k+1}^n -\frac{1}{2} (x_i - \hat{\mu}_2)' \hat{\Sigma}_2^{-1} (x_i - \hat{\mu}_2) - \frac{d}{2} n \lg 2\pi - \frac{k}{2} \lg |\hat{\Sigma}_1| - \frac{n-k}{2} \lg |\hat{\Sigma}_2| - \lambda \left(d + \frac{d(d+1)}{2} \right) \lg n \quad (6)$$

其中, $\hat{\mu}_1, \hat{\mu}_2, \hat{\Sigma}_1, \hat{\Sigma}_2$ 分别为 $\mu_1, \mu_2, \Sigma_1, \Sigma_2$ 的极大似然估计。

定义 $\Delta B(k)$ 为

$$\Delta B(k) = B(f_{\theta_1}) - B(f_{\theta_0}) \quad (7)$$

可以化简为

$$\Delta B(k) = -k \lg |\hat{\Sigma}_1| - (n-k) \lg |\hat{\Sigma}_2| + n \lg |\hat{\Sigma}| - \frac{\lambda}{2} \left(d + \frac{d(d+1)}{2} \right) \lg n \quad (8)$$

当 $\Delta B(k) > 0$ 时, 认为样本集 X 中存在说话人分割点, 并且分割点为 $\arg \max_k \Delta B(k)$ 。

图 1 是利用 BIC 进行单一说话人转变点检测时得到的 $\Delta B(k)$ 曲线。图中的圆点标示出了真实说话人转变点的发生时刻, 可以看出 $\Delta B(k)$ 曲线在该位置很好地形成了一个大于 0 的峰值。

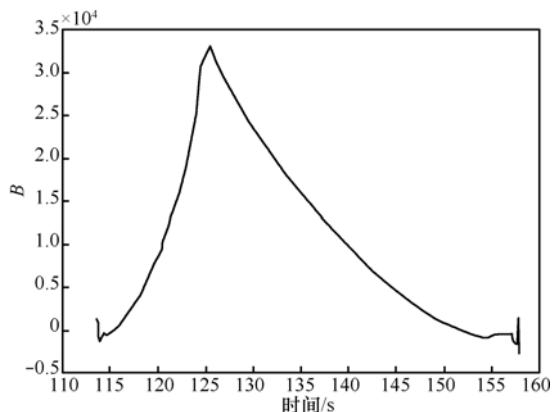


图1 单分割点检测曲线

在实际使用中,待分析语音往往包含多个说话人跳变点,Chen 等人在文献[9]中给出了一种多分割点的检测算法,首先在一个只包含2个样本的集合上使用BIC测试分割点是否存在,如果不存在分割点,则将下一个样本也包含进来,接着在新的样本集上再次测试,如果一直检测不到分割点则不断扩大样本集,直到找到分割点。当找到分割点后,新的样本集将从分割点之后的下2个样本开始,然后再按照前述的方法寻找分割点。可以看出,算法在无法确知分割点个数的情况下采取了一种“摸着石头过河”的方法,从一个最小的样本集开始逐一接纳新的样本。

2.3 BIC在说话人分割中的应用

现分析多分割点时分割算法的时间复杂度。设样本序列为 $X = \{x_i | i = 1, \dots, n\}$, 算法产生的说话人分割点序列为

$$S = \{s_j | j = 1, \dots, w, w+1\} \quad (9)$$

其中, $0 \leq w < n$, $2 \leq s_1 < s_2 < \dots < s_w < n$, $s_{w+1} = n$ 。 w 是分割点的个数,可以为0,当 $w=0$ 时 S 中只有一个元素 n ,表示算法在 X 集中没有检测到分割点。

设算法检测到第 i 个分割点 s_i 时,待检测的样本集 D 被重置为 $\{x_{s_i}, x_{s_i+1}\}$,接着应用BIC进行测试,没有发现转变点,继续将 x_{s_i+2} 添加到待检测样本集 D 。当 $D = \{x_{s_i}, x_{s_i+1}, \dots, x_{s_i+u}\}$ 时,有 $s_i < k \leq s_i + u$,因而需要计算 u 次 $\Delta B(k)$ 。算法不断进行,直到 $D = \{x_{s_i}, x_{s_i+1}, \dots, x_{s_{i+1}}\}$ 发现新的分割点 s_{i+1} 为止。由前面的分析可知,从 $D = \{x_{s_i}, x_{s_i+1}\}$ 到检测出新的分割点 s_{i+1} 计算 $\Delta B(k)$ 的次数为 $\frac{(s_{i+1} - s_i + 1)(s_{i+1} - s_i)}{2}$ 。

利用该结论可以推知,整个说话人分割过程中,计算 ΔB 的总次数 $F(n)$ 为

$$F(n) = \frac{n-1}{2} + \frac{(s_1-1)^2}{2} + \sum_{i=2}^{w+1} \frac{(s_i - s_{i-1})^2}{2} \quad (10)$$

从式(10)可以看出,算法的时间复杂度和可能的转变点个数有关,式中 $(s_i - s_{i-1})$ 表示的是分割段长度,可以证明分割段长度是服从均匀分布的。因而为了分析的方便,采用分割段的平均段长 $\frac{n}{w+1}$ 来代替式(10)中的 $(s_i - s_{i-1})$ 项,得到

$$F(n) = \frac{n-1}{2} + \frac{n^2}{w+1} \quad (11)$$

当 w 与 n 接近时, $F(n)$ 接近于 $O(n)$ 。在实际应用中, w 通常要比 n 小很多,以每20s发生一次说话人转变为例,如果帧率为80frame/s的话,则 $n/w=1600$ 。所以在通常情况下,该算法需要做将近 $O(n^2)$ 次的 ΔB 计算。而计算单个 ΔB 的时间复杂度为 $O(n^2)$ 。由此可以看出整个算法的时间复杂性是非常高的。

3 基于预分割的说话人分割方法

从第2节的分析中看出,Chen 等人提出的说话人分割算法时间复杂性过高,主要原因在于其针对多说话人的分割方法本质上仍旧是单一说话人的检测方法。由于事先无法预估跳变点的位置,只好从一个最短的序列(只包含2个样本点)开始,逐点地扩充待检测集,不断地进行说话人转变点检测。

下面提出一种基于预分割的方法,减少计算 ΔB 的次数。先通过一种相对简单的方法产生一组潜在分割点,从而将原始的音频特征序列划分成了若干小段,理想情况下,每2个相邻小段之间将存在一个说话人转变点。然后再使用单一转变点检测的BIC方法对潜在分割点进行逐一验证。

3.1 预分割方法

预分割是通过在一个滑动窗中计算出窗口中心点的 $\Delta B(t_0)$ 值,然后通过滑动该窗口得到一条曲线, $\Delta B(k)$ 越大表示位于第 k 个样本点两端的信号差异程度愈大,因而 $\Delta B(k)$ 曲线上的峰值点对应了说话人的转变点。

使用固定窗长为 $2N_w$ 的滑动窗,在落入窗口的样本子集 $\{x_{t_0-N_w+1}, \dots, x_{t_0}, x_{t_0+1}, \dots, x_{t_0+N_w}\}$ 上计算出位于窗口中心处 t_0 点的 ΔB 值为

$$\Delta B(t_0) = -N_w \lg |\hat{\mathcal{S}}_1| - N_w \lg |\hat{\mathcal{S}}_2| + 2N_w \lg |\hat{\mathcal{S}}| \left(d + \frac{d(d+1)}{2} \right) \lg 2N_w \quad (12)$$

接着窗口滑动 ΔT 个样本, 中心点落在 $\Delta T + t_0$ 位置, 再次计算得到 $\Delta B(t_0 + \Delta T)$ 。不断地如此滑动窗口, 得到序列 $\{\Delta B(t_0 + m\Delta T) | m = 0, 1, \dots\}$, 在此序列上大于 0 的局部峰值点表示可能发生说话人转变的位置。首先, 根据 BIC 理论只有当 ΔB 大于 0 时, 说话人发生转变模型才比说话人未发生转变的模型更优; 其次, BIC 公式中的第一项是极大似然, 因而当窗口中心点接近真实说话人分割点时其计算出的 ΔB 越大。

图 2 所示的就是在一段长约 120s 的语音样本上通过滑动窗计算出来的 $\{\Delta B(t_0 + m\Delta T)\}$ 序列, 图中的菱形标示出了真实的说话人转变点的位置。

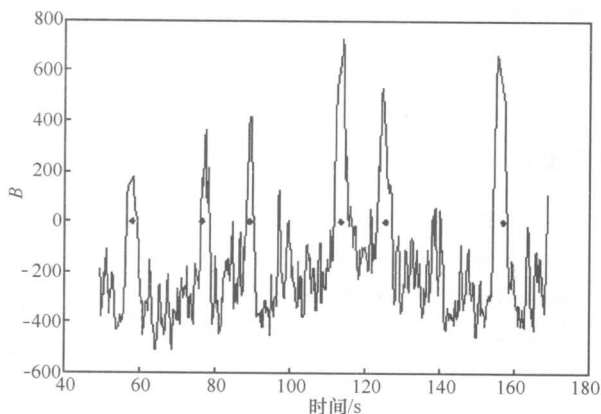


图 2 $\Delta B(t_0 + t\Delta T)$ 序列

3.2 分割点确认

设在预分割阶段完成后, 产生的预分割点序列为 $P = \{p_i | i = 0, 1, \dots, M\}$, 其中, $p_0 = x_1, p_M = x_n$ 。那么在理想情况下, $P_i = \{x_{p_{i-1}+1}, \dots, x_{p_i}, \dots, x_{p_{i+1}-1}\}$ 是一段包含有一个说话人转变点 x_{p_i} 的样本集。在该样本集上使用 BIC 求取出 $\max_k \Delta B(k)$, 如果 $\max_k \Delta B(k) > 0$, 根据 BIC 理论, 可以证明在 P_i 中确实存在说话人转变点。由于在预分割过程是基于滑动窗的, 且每隔 ΔT 个样本点才计算一次 B 值, 因而在预分割过程中分割点的时间分辨率有限, 使得求取的 p_i 可能会因此而与真实的分割点之间有所偏差。而在 P_i 集合中计算 $\Delta B(k)$ 时是逐样本点计算的, 时间分辨率要高出很多。因此, 如果 $\arg \max_k \Delta B(k) \neq p_i$ 的话, 最终还是采用 $\arg \max_k \Delta B(k)$

作为说话人分割点。

分割点确认的具体算法如下。

- 1) 输入潜在分割点个数 M 。
- 2) 初始化 $i=1$, 待检测样本集 $P_i = \{x_{p_{i-1}}, \dots, x_{p_i}, \dots, x_{p_{i+1}}\}$ 。
- 3) 令 $k = p_{i-1}$ 。
- 4) 计算 $\Delta B(k)$, 并保存到队列中。
- 5) $k = k + 1$, if $k \leq p_{i+1}$, 跳转到 4)。
- 6) if $\max_k \Delta B(k) > 0$, 则接受说话人转变点, $Q = Q \cup \{\arg \max_k \Delta B(k)\}$ 。
- 7) $i = i + 1$, 检验下一个潜在分割点, 重置 $P_i = \{x_{p_{i-1}}, \dots, x_{p_{i+1}}\}$ 。
- 8) if $i < M$, 跳转到 3)。

3.3 算法时间复杂性分析

基于预分割的说话人分割算法的时间复杂性分为 2 部分: ①预分割部分的时间复杂性, 由于分割算法可以看出窗口每次滑动 ΔT 长度并计算一次 ΔB , 故计算 ΔB 的总次数为 $\lceil n / \Delta T \rceil$, 属于 $O(n)$ 级。②在分割点验证阶段, 计算 ΔB 的次数为 $n + p_{M-1} - p_1$ 。在每次验证中, 分割点确认算法第 4) 步需要计算 $(p_{i+1} - p_{i-1})$ 次 ΔB , 总的计算次数可证是小于 $2n$ 的, 从而在分割点确认阶段也只需计算 $O(n)$ 次 ΔB 。

从上面的分析可知, 基于预分割的说话人分割算法在预分割阶段和分割点验证阶段的都只需要计算 $O(n)$ 次 ΔB , 故总共也只需计算 $O(n)$ 次 ΔB 。

4 实验结果及分析

实验语料录制于网络新闻广播节目, 共录制了 6 段, 每段录制时间约 15min, 表 1 为各段实验语料的时长和所包含的实际分割点的情况。

表 1 说话人分割的实验语料

语料段	时长/s	实际分割点数目
#1	1 205	61
#2	1 306	55
#3	1 027	70
#4	945	37
#5	1 276	49
#6	1 257	62

采用召回率和正确率来评价实验结果。在相同的语料集上分别使用 Chen 等所提出的分割算法和

基于预分割的说话人分割算法进行实验,表 2 是使用 Chen 等所提出的分割算法实验得出的分割结果,表 3 是使用基于预分割的方法得到的实验结果。为了更好地分析实验结果,在表 3 中除了给出最终的分割结果,还给出了分割点确认之前的召回率及准确率。

表 2 传统的说话人分割方法的实验结果

语料段	召回率/%	准确率/%
#1	86.9	72.6
#2	83.6	65.7
#3	85.7	77.9
#4	81.1	66.7
#5	79.6	75.1
#6	83.8	71.6

表 3 基于预分割方法的实验结果

语料段	召回率		准确率	
	预分割/%	确认后/%	预分割/%	确认后/%
#1	77.1	75.4	70.1	88.5
#2	76.4	72.7	60.0	78.4
#3	78.6	72.9	64.0	82.2
#4	70.3	67.6	59.1	73.5
#5	75.5	75.5	67.3	80.4
#6	77.4	77.4	73.8	76.2

观察表 2 可见,传统的分割方法其召回率基本在 80%以上,准确率大约在 70%左右。在实验中发现,如果在这个基础上再提高召回率,相应的准确率将发生比较陡峭的下降。表 2 中的结果是综合考虑召回率和准确率的情况下选取的实验结果。

再观察表 3,基于预分割的方法通过分割点确认这一步,召回率在预分割的基础上发生了小幅度的下降,其下降的程度基本在 5 个百分点以下,而准确率得到了较大的提高,一般都上升了 10 个百分点。这是因为分割点确认的目的就是要排除预分割时产生的错误分割点,当然在排除错误的分割点的时候也可能会错误地将个别几个本来正确的分割点去除掉,但是实验表明这样的情况发生的很少,分割点确认算法达到了预期的目的。

众观表 2 和表 3,可以发现由于在预分割阶段说话人分割的召回率低于传统的方法,因而最终的召回率低了近 10 个百分点。究其原因,是因为

预分割算法为了降低算法复杂度而不得不变得相对粗糙,在分割点的检出能力上受到制约,即便放大大对准确率降低的容忍程度都很难使得其召回率进一步得到较大的提高。

在同一台 PC 机上进行说话人分割实验,以考核其运行时间指标。对约 15min 的语料段,传统的 BIC 分割方法平均处理时间为 180min 左右,约为 9 倍实时,显然无法满足实时处理的要求。而基于预分割的说话人分割方法平均处理时间只需要 20min,基本上达到了实时处理的要求。

5 结束语

针对传统的基于 BIC 的说话人分割方法,在进行多分割点说话人分割时计算时间复杂度太大的问题,提出了一种基于预分割的说话人分割算法,通过将语音预分割成仅包含一个分割点的若干小段,从而有效减少了计算量,在处理速度上有近 9 倍的提升,基本上达到了实时处理的要求。

参考文献:

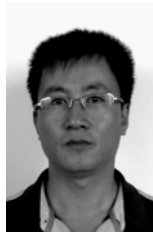
- [1] LU L, ZHANG H J. Speaker change detection and tracking in real-time news broadcasting analysis[A]. Proc ACM Multimedia, Juan-les-Pins[C]. France, 2002.602-610.
- [2] BAKIS R, CHEN S, GOPALAKRISHNAN P S. Transcription of broadcast news shows with the IBM large vocabulary speech recognition system[A]. Proc of DARPA Speech Recognition Workshop[C]. Chantilly, Virginia, USA, 1997.67-72.
- [3] WU T Y, LU L, CHEN K, *et al.* UBM-based real-time speaker segmentation for broadcasting news[A]. International Conference on Acoustic, Speech and Signal Processing[C]. Hong Kong, 2003. 193-196.
- [4] WU T Y, LU L, CHEN K, *et al.* UBM-based incremental speaker adaptation[A]. Proc of IEEE International Conference on ICME[C]. Baltimore, USA, 2003. 721-724.
- [5] SEIGLER M A, RAJ U, STERN R M. Automatic segmentation classification and clustering of broadcast news audio[A]. Proc of DARPA Speech Recognition Workshop[C]. Chantilly, Virginia, USA, 1997. 97-99.
- [6] GISH H, SCHMIDT N. Text-independent speaker identification[J]. IEEE Signal Processing Magazine, 1994,11(4):18-32.
- [7] YAO Y C, DAVIS R A. The asymptotic behavior of the likelihood ratio statistics for testing shift in mean in a sequence of independent normal varieties[J]. Sankhya, 1986, 48:339-353.

(下转第 128 页)

2006, 29(8): 1375-1382.

- [6] JHA S, SHEYNER O, WING J. Two formal analyses of attack graphs[A]. Proc the 15th Computer Security Foundations Workshop[C]. Nova, Scotia, 2002.49-63.
- [7] SHEYNER O, HAINES J, JHA S. Automated generation and analysis of attack graphs[A]. Proc 2002 IEEE Symposium on Security and Privacy[C]. Oakland, California, USA, 2002.254-265.
- [8] PAMULA J, AMMANN P. A weakest-adversary security metric for network configuration security analysis[A]. Proc of the 2nd ACM Workshop on Quality of Protection[C]. Alexandria, Virginia, USA, 2006. 31-38.
- [9] WANG L Y, NOEL S, JAJODIA S. Minimum-cost network hardening using attack graphs[J]. Computer Communications, 2006, 29: 3812-3824.
- [10] 张永铮, 云晓春, 胡铭曾. 基于特权提升的多维量化属性弱点分类法的研究. 通信学报, 2004, 25(7): 107-114.
- ZHANG Y Z, YUN X C, HU M C. Research on privilege-escalating based vulnerability taxonomy with multidimensional quantitative attribute[J]. Journal on Communications, 2004, 25(7): 107-114.
- [11] MAN D P, ZHANG B, YANG W. A method for global attack graph generation[A]. Proc of 2008 IEEE International Conference on Networking, Sensing and Control[C]. Sanya, China, 2008.236-241.

作者简介:



司加全 (1972-), 男, 山东东阿人, 哈尔滨工程大学信息安全研究中心博士生, 主要研究方向为网络风险评估。



张冰 (1973-), 男, 黑龙江哈尔滨人, 博士, 国家计算机网络应急技术处理协调中心高级工程师, 主要研究方向为网络安全、复杂系统仿真等。

苟大鹏 (1980-), 男, 辽宁抚顺人, 哈尔滨工程大学信息安全研究中心博士生, 主要研究方向为网络风险评估、安全投资等。

杨武 (1974-), 男, 辽宁宽甸人, 博士, 哈尔滨工程大学教授, 主要研究方向为网络安全。

(上接第 122 页)

- [8] SCHWARZ G. Estimation of the dimension of a model[J]. The Annals of Statistics, 1978, 6:461-464.
- [9] CHEN S S, GOPALAKRISHNAN P S. Speaker, environment and channel change detection and clustering via the bayesian information criterion[A]. DARPA Broadcast News Transcription and Understanding Workshop[C]. Lansdowne, Virginia, USA, 1998.127-132.

作者简介:



郑铁然 (1972-), 男, 黑龙江哈尔滨人, 哈尔滨工业大学计算机学院讲师, 主要研究方向为语音识别、语音文档检索等。



李海峰 (1969-), 男, 黑龙江哈尔滨人, 哈尔滨工业大学计算机学院教授, 主要研究方向为语音信号处理、音频信号处理、人工神经网络。

刘先 (1982-), 男, 浙江杭州人, 硕士, 主要研究方向为语音分割。

韩纪庆 (1964-), 男, 黑龙江哈尔滨人, 哈尔滨工业大学计算机学院教授, 主要研究方向为语音信号处理、音频信号处理、网络信息安全。