

高斯混合分布之间 K-L 散度的近似计算

王欢良¹ 韩纪庆¹ 郑铁然¹

摘要 高斯混合分布之间的 K-L 散度没有闭式解, 通常采用其上界来近似. 对于具有相同高斯数的混合分布, 基于相对熵链规则推导其 K-L 散度上界, 提出一种更紧上界的计算方法. 为计算具有不同高斯数的混合分布之间的 K-L 散度上界, 提出基于最佳高斯分量复制的方法. 在中文声韵母声学模型上的实验结果显示, 所提出方法可更好地近似等高斯数的混合分布之间的 K-L 散度, 并能有效处理具有不同高斯数的混合分布.

关键词 K-L 散度 (KLD), 高斯混合分布 (GMD), 相对熵, K-L 散度上界
中图分类号 TP391

Approximation of Kullback-Leibler Divergence between Two Gaussian Mixture Distributions

WANG Huan-Liang¹ HAN Ji-Qing¹ ZHENG Tie-Ran¹

Abstract For no closed-form expression is available for Kullback-Leibler divergence (KLD) between two Gaussian mixture distributions (GMDs), the upper-bound of its solution is used to approximate it. In this paper, the upper-bound of KLD of two GMDs with the same number of components is deduced according to the relative entropy link rule, and then a tighter upper-bound is proposed. In the case that two GMDs have different numbers of components, a method, named optimal Gaussian duplication (OGD), is proposed to approximate their KLD. The evaluation experiments are performed on the acoustic models of the initial and the final, which all are modeled by GMD based HMM in speech recognition. The experimental results show that the tighter upper-bound can more perfectly approximate the KLD than other methods, and the proposed OGD method can effectively compute the upper-bound of KLD between two GMDs with different numbers of components.

Key words K-L divergence (KLD), Gaussian mixture distribution (GMD), relative entropy, upper-bound of K-L divergence

K-L 散度 (Kullback-Leibler divergence, KLD), 也称为相对熵 (Relative entropy), 用来度量两个概率分布之间的差异^[1], 是一种很有意义的统计测度. 设统计模型 \mathcal{M} 和 $\tilde{\mathcal{M}}$ 分别表示两个 N 维概率分布函数, 则这两个模型之间的 KLD 定义为

$$D(\mathcal{M}||\tilde{\mathcal{M}}) = \int_{\mathbf{R}^N} P(x|\mathcal{M}) \log \frac{P(x|\mathcal{M})}{P(x|\tilde{\mathcal{M}})} dx \quad (1)$$

上式的物理意义是: 给定参照统计模型 \mathcal{M} , 竞争统计模型 $\tilde{\mathcal{M}}$ 与它的差异程度. 显然, $D(\mathcal{M}||\tilde{\mathcal{M}}) \neq D(\tilde{\mathcal{M}}||\mathcal{M})$. 如果不考虑参照模型, 则可定义对称 K-L 散度为

$$J(\mathcal{M}, \tilde{\mathcal{M}}) = D(\mathcal{M}||\tilde{\mathcal{M}}) + D(\tilde{\mathcal{M}}||\mathcal{M}) \quad (2)$$

根据信息不等式, KLD 是半正定的^[2]. 在很多文献

中也称它为距离, 但它不满足三角不等式, 因此并不是严格意义上的距离.

利用 KLD, 可以直接比较统计模型之间的差异, 从而可进行模型分析, 并指导模型的训练和应用过程. KLD 已成功应用于很多模式分类和模型分析任务, 如文本检索^[3]、机器翻译^[4]、话者识别^[5]、模型聚类^[6] 和区分性训练^[7-8] 等.

高斯混合分布 (Gaussian mixture distribution, GMD) 作为一种基本建模方法能够近似很多复杂的概率分布情况, 已广泛应用于语音和图像处理等领域. 对于 GMD 来说, 其 KLD 不存在闭式解, 因此只能通过数值方法来近似计算. 常用的近似方法大致可分为两类: 一类是基于采样的 Monte-Carlo 方法^[9]; 另一类通过 KLD 上界来近似^[10-11]. 随着采样数据的增加, Monte-Carlo 方法能够不断逼近统计模型之间真实的 KLD. 但它需要预先生成大量特定分布的样本数据, 因此计算代价很高. KLD 上界可根据模型参数直接估计, 计算代价很低, 但它与真实 KLD 的误差较大, 并且通常要求两个 GMD 具有相同高斯数.

本文主要研究了任意 GMD 之间更紧 KLD 上

收稿日期 2006-11-28 收修改稿日期 2007-08-15
Received November 28, 2006; in revised form August 15, 2007
国家重点基础研究发展计划 (973 计划) (2007CB311100) 资助
Supported by National Basic Research Program of China (973 Program) (2007CB311100)
1. 哈尔滨工业大学计算机学院 哈尔滨 150001
1. School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001
DOI: 10.3724/SP.J.1004.2008.00529

界的高效计算方法. 首先根据相对熵的链规则重新推导 KLD 上界, 然后据此提出一个更紧 KLD 上界的计算方法. 针对具有不同高斯数的 GMD, 提出了最优高斯分量复制方法来计算其 KLD 上界. 最后在中文声韵母对应的声学模型上, 通过计算状态观察概率分布之间的 KLD 上界来验证本文所提方法.

1 相同高斯数 GMD 之间的 KLD 上界

1.1 基于相对熵链规则的 KLD 上界

引理 1 (相对熵的链规则). 设 X 和 Y 是两个随机变量, $P(X, Y)$ 和 $\tilde{P}(X, Y)$ 是两个联合分布, 则^[2]

$$D(P(X, Y) || \tilde{P}(X, Y)) = D(P(X) || \tilde{P}(X)) + D(P(Y|X) || \tilde{P}(Y|X)) \quad (3)$$

设 S 是定义在空间 $\mathcal{S} = \{s_1, s_2, \dots, s_I\}$ 上的一个不可观测的离散随机变量, X 是空间 \mathbf{R}^N 上的一个可观测的连续随机变量, $P(X, S)$ 和 $\tilde{P}(X, S)$ 是 $(\mathbf{R}^N \times \mathcal{S})$ 上的两个联合分布, 则

$$D(P(X, S) || \tilde{P}(X, S)) = D(P(X) || \tilde{P}(X)) + D(P(S|X) || \tilde{P}(S|X)) = D(P(S) || \tilde{P}(S)) + D(P(X|S) || \tilde{P}(X|S)) \quad (4)$$

由于 $D(P(S|X) || \tilde{P}(S|X)) \geq 0$, 因此有

$$D(P(X) || \tilde{P}(X)) \leq D(P(S) || \tilde{P}(S)) + D(P(X|S) || \tilde{P}(X|S)) \quad (5)$$

如果令 $c_i = P(S = s_i)$, $\tilde{c}_i = \tilde{P}(S = s_i)$, $f_i(x) = P(X = x|S = s_i)$, $\tilde{f}_i(x) = \tilde{P}(X = x|S = s_i)$, 并且边缘分布 $P(X = x) = \sum_S P(X = x, S) = \sum_{i=1}^I c_i f_i(x)$, $\tilde{P}(X = x) = \sum_S \tilde{P}(X = x, S) = \sum_{i=1}^I \tilde{c}_i \tilde{f}_i(x)$, 令 $f(x) = P(X = x)$, $\tilde{f}(x) = \tilde{P}(X = x)$, 则

$$D(f || \tilde{f}) = D(P(X) || \tilde{P}(X)) \leq \sum_{i=1}^I c_i \log \frac{c_i}{\tilde{c}_i} + \sum_{i=1}^I c_i D(f_i || \tilde{f}_i) \quad (6)$$

如果 $P(X|S = s_i)$ 和 $\tilde{P}(X|S = s_i)$ 服从高斯分布, 则这个上界就是两个相同高斯数 GMD 之间的 KLD 上界. 这个上界与通过对数求和不等式推导的 KLD 上界^[10] 是一致的 (参见附录 A).

1.2 更紧的 KLD 上界

从式 (4) 可以看出, 联合分布的 KLD 与其边缘分布的 KLD 之间的差异造成了 GMD 的 KLD 与其上界之间的差异. 这表明在这个 KLD 上界中比较

了模型之间潜在结构的差异, 即在 S 和 X 之间的映射关系. 因此, 为得到更紧的 KLD 上界, 需要最小化这种模型内部潜在映射关系之间的 KLD. 值得注意的是, 在推导过程中, 并没有强制限定两个 GMD 中的高斯分量之间的对应关系. 因此, 可以通过在两个 GMD 中的高斯分量之间建立最优对应关系来最小化 KLD 上界, 从而更好地近似 KLD.

设 $f(x) = \sum_{i=1}^I c_i f_i(x)$ 和 $\tilde{f}(x) = \sum_{i=1}^I \tilde{c}_i \tilde{f}_i(x)$ 是两个 GMD, 其中 $f_i(x) \sim \mathcal{N}(x; \mu_i, \Sigma_i)$, $\tilde{f}_i(x) \sim \mathcal{N}(x; \tilde{\mu}_i, \tilde{\Sigma}_i)$. 定义一一映射函数 $\beta: I \rightarrow I$, 表示两个 GMD 中高斯分量之间的对应关系, 则更紧的 KLD 上界为

$$D(f || \tilde{f}) \leq \min_{\beta} \sum_{i=1}^I c_i \left\{ \log \frac{c_i}{\tilde{c}_{\beta(i)}} + D(f_i || \tilde{f}_{\beta(i)}) \right\} \quad (7)$$

寻找最优映射函数 $\beta(\cdot)$ 使式 (7) 右端最小化需要 $I!$ 次高斯分布的排列运算. 显然, 这是难以实现的. 我们采用贪心算法给出一个次优的映射函数, 具体算法如下:

1) 根据高斯权值 c_i 从大到小排序参照 GMD 中的所有高斯分量, 使得 $c_1 \geq c_2 \geq \dots \geq c_I$, 令 $i = 1$, $A = \emptyset$;

2) 根据下式依次构造映射函数

$$\beta(i) = \arg \min_{j \in \{1, \dots, I\} \setminus A} \{D(f_i || \tilde{f}_j) - \log \tilde{c}_j\} \quad (8)$$

其中 $A = \{\beta(t) | t = 1, \dots, i-1\}$;

3) $A = A \cup \{\beta(i)\}$, 如果 $i < I$, $i = i + 1$, 转 2); 否则, 结束.

上面贪心算法的计算复杂度为 $O(I^2)$, 其中 I 为 GMD 中的高斯数.

2 不同高斯数 GMD 之间的 KLD 上界

设 $f(x) = \sum_{i=1}^I c_i f_i(x)$ 和 $\tilde{f}(x) = \sum_{i=1}^{\tilde{I}} \tilde{c}_i \tilde{f}_i(x)$ 是两个 GMD, 其中 $f_i(x) \sim \mathcal{N}(x; \mu_i, \Sigma_i)$, $\tilde{f}_i(x) \sim \mathcal{N}(x; \tilde{\mu}_i, \tilde{\Sigma}_i)$. 如果 $I \neq \tilde{I}$, 则无法直接根据第 1.1 节的推导获得 $D(f || \tilde{f})$ 的上界, 因为这两个 GMD 的隐变量空间 \mathcal{S} 是不同的. 为此, 需要在保持原始分布函数不变的前提下使两个 GMD 具有同样的隐变量空间. 我们提出采用最优高斯分量复制的办法来增加混合高斯数, 最终使两个 GMD 具有相同高斯数. 下面分两种情况进行讨论.

2.1 $I > \tilde{I}$ 的情况

此时, 竞争 GMD 中的高斯数少于参照 GMD, 因此需要增加竞争 GMD 中的高斯数. 新添加的高斯分量应该是竞争 GMD 中与当前高斯分量具有最小 KLD 值的高斯分量的副本. 具体算法如下:

1) 根据高斯权值 c_i 从大到小排序参照 GMD 中的所有高斯分量, 使得 $c_1 \geq c_2 \geq \dots \geq c_I$, 令 $i = 0$, $A = \emptyset$;

2) 根据式 (8) 依次构造映射函数 $\beta(i)$, $A = \{\beta(t) | t = 1, \dots, i-1\}$;

3) $A = A \cup \{\beta(i)\}$, $i = i + 1$, 如果 $i \leq \tilde{I}$, 转 2);

4) 从竞争 GMD 中寻找可用于复制的最优高斯分量 (推导过程见附录 B):

$$k = \arg \min_{j \in \{1, \dots, i-1\}} \left\{ c_{\beta^{-1}(j)} \log \frac{c_{\beta^{-1}(j)}}{\tilde{c}_j - \tilde{c}_n} + c_i \log \frac{c_i}{\tilde{c}_n} + c_{\beta^{-1}(j)} D(f_{\beta^{-1}(j)} || \tilde{f}_j) + c_i D(f_i || \tilde{f}_i) \right\} \quad (9)$$

其中, $\tilde{c}_n = \tilde{c}_j c_i / (c_{\beta^{-1}(j)} + c_i)$, $\beta^{-1}(\cdot)$ 表示 $\beta(\cdot)$ 的反函数;

5) 复制最优高斯分量: $\tilde{f}_i(x) = \tilde{f}_k(x)$, 最优权值为 $\tilde{c}_i = \tilde{c}_k c_i / (c_{\beta^{-1}(k)} + c_i)$, $\tilde{c}_k = \tilde{c}_k - \tilde{c}_i$, 并令 $\beta(i) = i$;

6) $i = i + 1$, 如果 $i \leq I$, 转 4); 否则, 令 $\tilde{I} = I$, 结束.

显然, $\sum_{i=1}^I \tilde{c}_{\beta(i)} \tilde{f}_{\beta(i)}(x) = \tilde{f}(x)$.

2.2 $I < \tilde{I}$ 的情况

此时, 参照 GMD 中的高斯数少于竞争 GMD, 因此需要在参照 GMD 中添加高斯分布函数. 添加的高斯分量应该是与竞争 GMD 中的当前高斯分量具有最小 KLD 值的高斯分布函数的副本. 具体算法如下:

1) 根据高斯权值 c_i 从大到小排序参照 GMD 中的所有高斯分量, 使得 $c_1 \geq c_2 \geq \dots \geq c_I$, 令 $i = 0$, $A = \emptyset$;

2) 根据式 (8) 依次构造映射函数 $\beta(i)$, $A = \{\beta(t) | t = 1, \dots, i-1\}$;

3) $A = A \cup \{\beta(i)\}$, $i = i + 1$, 如果 $i \leq I$, 转 2);

4) 从参照 GMD 中寻找可用于复制的最优高斯分量 (推导过程同附录 B):

$$k = \arg \min_{j \in \{1, \dots, i-1\}} \left\{ c_m \log \frac{c_m}{\tilde{c}_{\beta(j)}} + c_n \log \frac{c_n}{\tilde{c}_i} + c_m D(f_j || \tilde{f}_{\beta(j)}) + c_n D(f_j || \tilde{f}_i) \right\} \quad (10)$$

其中

$$c_m = c_j - c_n$$

$$c_n = c_j \tilde{c}_i B / (\tilde{c}_{\beta(j)} + \tilde{c}_i B)$$

$$B = r^{D(f_j || \tilde{f}_{\beta(j)}) - D(f_j || \tilde{f}_i)}, r \text{ 为 } \log \text{ 函数的底数}$$

5) 复制最优高斯分量 $f_i(x) = f_k(x)$, 最优权值 $c_i = c_k \tilde{c}_i B / (\tilde{c}_{\beta(k)} + \tilde{c}_i B)$, $c_k = c_k - c_i$, $B = r^{D(f_k || \tilde{f}_{\beta(k)}) - D(f_k || \tilde{f}_i)}$, 并令 $\beta(i) = i$;

6) $i = i + 1$, 如果 $i \leq \tilde{I}$, 转 4); 否则, 令 $I = \tilde{I}$, 结束.

显然, $\sum_{i=1}^{\tilde{I}} c_i f_i(x) = f(x)$.

最后, 两个 GMD 具有相同的高斯数, 只要基于映射函数 $\beta(\cdot)$, 即可应用式 (7) 来计算两个 GMD 的 KLD 上界.

3 实验与分析

GMD 的一个典型应用就是在语音识别中作为隐马尔可夫模型 (Hidden Markov model, HMM) 的状态观察概率分布函数. 为验证所提算法的有效性, 采用如下中文声韵母对应的声学模型作为实验数据:

声母: b, p, m, f, d, t, n, l, g, k, h, j, q, x, zh, ch, sh, r, z, c, s

韵母: a, o, e, i, u, ui, iu, ai, ei, ao, ou, an, en, ang, eng, ong, un

声学模型为 3 状态自左向右结构的 HMM, 特征为 39 维标准 Mel 频率倒谱系数 (Mel-frequency cepstral coefficient, MFCC) 特征. 在 863 数据库上训练两组声学模型, 其状态上 GMD 的高斯数分别为 8 和 16. 采用 HTK 工具包完成声学模型训练^[12]. 由于模型前后状态易受上下文影响, 因此实验只采用中间状态的 GMD.

采用两个评价指标来评测本文所提方法. 一个是放大比 (Amplified ratio), 它是近似 KLD 与真实 KLD 的比值, 反映了近似值的准确程度. 实验中, 以采用 Monte-Carlo 方法估计的 KLD 作为真实 KLD, 每次随机生成 5 000 个观察样本来计算它. 另一个指标是等级 (Rank) 相关系数, 它是所有 KLD 近似值与真实值的排序等级之间的相关程度, 反映了它们在相对大小上的一致性. 实际上, KLD 主要用来进行模型比较和分析, 因此其相对大小要比具体值更有意义. 相关系数越大, 表明近似 KLD 越能更好地保持真实的等级 (排序) 关系.

3.1 相同高斯数 GMD 的 KLD 近似方法比较

采用 8 高斯的 GMD, 验证在等高斯情况下本文所提 KLD 上界的紧度. 总共可计算 $38 \times (38 - 1) = 1406$ 个 KLD. 为了比较, 实现了随机匹配近似方法 (Random match KLD, RM-KLD)^[10] 和 APA (Asymptotic likelihood approximation) 近似方法 (APA-KLD)^[11]. 图 1 (见下页) 给出了分别采用 RM-KLD 方法、APA-KLD 方法和本文所提方法计算的近似 KLD 的放大比直方图.

从图 1 可以看出, 大部分的 KLD 近似值是其真实值的 1.15 倍. 与 RM-KLD 方法和 APA-KLD 方法相比, 本文所提方法获得的 KLD 近似值更加集中, 并且更接近其真实值. 这表明本文所提方法能得到一个更紧的 KLD 上界. 表 1 给出了分别采用 RM-KLD 方法、APA-KLD 方法和本文所提方法计算得到的 KLD 近似值与其真实值之间的等级相关系数.

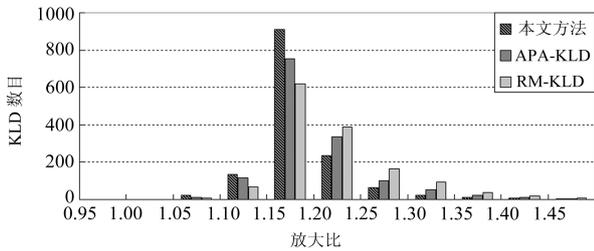


图 1 等高斯情况下, 采用不同方法计算的近似 KLD 的放大比直方图

Fig. 1 Histogram of amplified ratios of the approximated KLD estimated by three methods in the case of same Gaussian numbers

表 1 等高斯情况下, 不同方法计算的 KLD 近似值与其真实值之间的等级相关系数

Table 1 Rank correlation between the true KLD and its approximation obtained by different approaches in the case of same Gaussian numbers in two GMDs

KLD 计算方法	等级相关系数
RM-KLD 方法	0.874
APA-KLD 方法	0.918
本文所提方法	0.941

从表 1 可以看出, 本文所提方法对应的等级相关系数最大. 这说明在对等高斯 GMD 模型进行相似性比较时, 采用本文所提方法计算的近似 KLD 与真实 KLD 在相对大小上能够保持更高的一致性. 因此, 在实际应用中采用本文所提方法对 GMD 模型进行比较和分析更为可靠.

3.2 不同高斯数 GMD 的 KLD 近似

计算所有 8 高斯和 16 高斯 GMD 之间的 KLD 上界, 这样总共可得到 $38 \times 38 = 1444$ 个近似 KLD. 图 2 给出了采用本文所提方法计算的近似 KLD 的放大比直方图.

从图 2 可以看出, 在不同高斯数情况下, 绝大部分 KLD 近似值都在其真实值的 1.3~1.6 倍范围内, 集中程度较高. 这表明本文所提方法能够比较准确地近似不同高斯数 GMD 的 KLD. 图 3 是所有模型

根据 KLD 近似值和其真实值分别排序后对应关系的散点图.

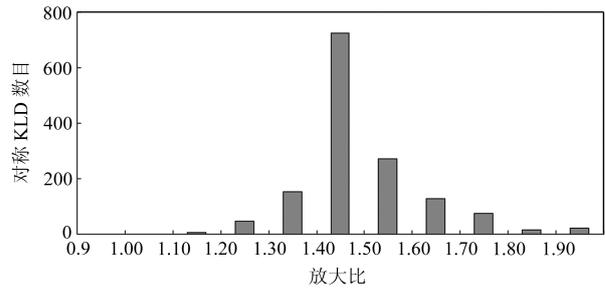


图 2 不同高斯数情况下, 本文所提方法计算的近似 KLD 的放大比直方图

Fig. 2 Histogram of amplified ratios of the approximated KLD estimated by the proposed method in the case of different Gaussian numbers

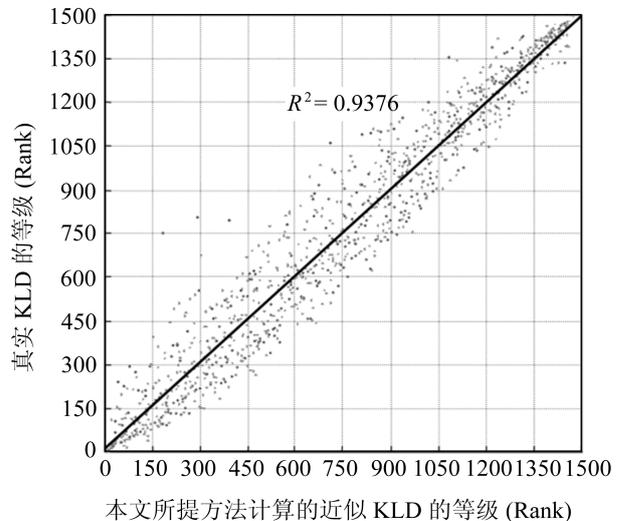


图 3 不同高斯数情况下, 所有近似 KLD 与真实 KLD 的等级对应关系的散点图和等级相关系数

Fig. 3 Scatter graph and the correlation of rank order of the approximated KLDs against their true values in the case of different Gaussian numbers

从图 3 可以看到, 大部分等级对应点分布在对角线附近, 相对比较集中, 并且等级相关系数 $R^2 = 0.9376$. 这说明采用本文所提方法计算的 KLD 近似值和其真实值在等级顺序上具有较高的一致性. 也就是说, 对于不同高斯数的 GMD, 本文所提方法能够使其 KLD 近似值之间的相对大小关系和其真实值较好地保持一致. 因此, 本文所提方法可用于对任意 GMD 模型进行有效地比较和分析.

4 结论

本文利用 KLD 上界来近似两个 GMD 的 KLD. 对于具有相同高斯数的 GMD, 提出了计算更紧 KLD 上界的方法. 对于具有不同高斯数的 GMD, 提出了最优高斯分量复制方法来计算其 KLD 上界. 在中文声韵母声学模型上的对比实验表明, 本文所提方法计算的近似 KLD 比其他方法更接近真实值, 并且能有效计算任意 GMD 之间的 KLD 上界.

基于本文所提方法, 可以对任意的 GMD 模型进行有效地比较和分析. 在语音识别应用中, 利用 GMD 的 KLD 可以指导状态绑定过程, 使得数据共享更为合理, 更好地解决数据稀疏问题. 在模型训练过程中, KLD 可用来指示两个模型的可区分性或者作为识别错误的软指标^[8], 从而使训练更有针对性. 在文本或图像检索任务中, GMD 的 KLD 可用来作为决策准则, 因此 KLD 近似值和其真实值保持相同的等级顺序显得尤为重要. 此外, 本文所提方法还可用于更好地估计两个 HMM 之间的 KLD.

附录 A. 基于对数求和不等式的 KLD 上界推导

引理 1 (对数求和不等式). 对于非负数 a_1, a_2, \dots, a_N 和 b_1, b_2, \dots, b_N , 下式成立^[2]

$$\left(\sum_{n=1}^N a_n\right) \log \frac{\sum_{n=1}^N a_n}{\sum_{n=1}^N b_n} \leq \sum_{n=1}^N a_n \log \frac{a_n}{b_n} \quad (\text{A1})$$

如果 $a_n/b_n = C$, 则等式成立.

设 $f(x) = \sum_{i=1}^I c_i f_i(x)$ 和 $\tilde{f}(x) = \sum_{i=1}^I \tilde{c}_i \tilde{f}_i(x)$ 是两个 GMD, 其中 $f_i(x) \sim \mathcal{N}(x; \mu_i, \Sigma_i)$, $\tilde{f}_i(x) \sim \mathcal{N}(x; \tilde{\mu}_i, \tilde{\Sigma}_i)$, 则

$$D(f||\tilde{f}) = \int_{\mathbf{R}^N} \sum_{i=1}^I c_i f_i(x) \log \frac{\sum_{i=1}^I c_i f_i(x)}{\sum_{i=1}^I \tilde{c}_i \tilde{f}_i(x)} dx \quad (\text{A2})$$

根据引理 1, 可得

$$D(f||\tilde{f}) \leq \int_{\mathbf{R}^N} \sum_{i=1}^I c_i f_i(x) \log \frac{c_i f_i(x)}{\tilde{c}_i \tilde{f}_i(x)} dx = \sum_{i=1}^I c_i \log \frac{c_i}{\tilde{c}_i} + \sum_{i=1}^I c_i D(f_i||\tilde{f}_i) \quad (\text{A3})$$

上式就是两个具有相同高斯数 GMD 的 KLD 上界.

附录 B. 确定最优高斯分量权值的推导

以当前高斯分量 $c_i f_i(x)$ 为参照, 不妨设竞争 GMD 中将要被复制的高斯分量为 $\tilde{c}_j \tilde{f}_j(x)$, 新复制高斯分量的权为 \tilde{c}_n , 原始高斯分量的权变为 $\tilde{c}_j - \tilde{c}_n$, 则新的 KLD 上界 $g(\tilde{c}_n)$

为

$$g(\tilde{c}_n) = c_{\beta-1(j)} \log \frac{c_{\beta-1(j)}}{\tilde{c}_j - \tilde{c}_n} + c_{\beta-1(j)} D(f_{\beta-1(j)}||\tilde{f}_j) + c_i \log \frac{c_i}{\tilde{c}_n} + c_i D(f_i||\tilde{f}_i) + C \quad (\text{B1})$$

其中 C 为常数, 表示其余已确定对应关系的高斯分量之间的 KLD. 在给定被复制的高斯分量后, 需要确定复制后高斯分量的权 \tilde{c}_n 来最小化式 (B1).

由于

$$g''(\tilde{c}_n) = (c_{\beta-1(j)}(\tilde{c}_j - \tilde{c}_n)^{-2} + c_i(\tilde{c}_n)^{-2}) \log e \geq 0 \quad (\text{B2})$$

因此, 式 (B1) 存在极小值, 令

$$g'(\tilde{c}_n) = 0 \quad (\text{B3})$$

可得

$$\tilde{c}_n = \frac{\tilde{c}_j c_i}{c_{\beta-1(j)} + c_i} \quad (\text{B4})$$

显然, 要被复制的最佳高斯分量应该为使式 (B1) 最小化的高斯分量, 省略常数 C , 即为

$$k = \arg \min_{j \in \{1, \dots, i-1\}} \left\{ c_{\beta-1(j)} \log \frac{c_{\beta-1(j)}}{\tilde{c}_j - \tilde{c}_n} + c_i \log \frac{c_i}{\tilde{c}_n} + c_{\beta-1(j)} D(f_{\beta-1(j)}||\tilde{f}_j) + c_i D(f_i||\tilde{f}_i) \right\} \quad (\text{B5})$$

其中 \tilde{c}_n 由式 (B4) 定义.

References

- 1 Kullback S, Leibler R A. On information and sufficiency. *The Annals of Mathematical Statistics*, 1951, **22**(1): 79–86
- 2 Cover T M, Thomas J A. *Elements of Information Theory*. New York: Wiley-Interscience, 1991
- 3 Do M N, Vetterli M. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Transactions on Image Processing*, 2002, **11**(2): 146–158
- 4 Hu Ri-Le, Zong Cheng-Qing, Xu Bo. Approach to automatic translation template acquisition based on statistical learning. *Journal of Chinese Information Processing*, 2005, **19**(6): 1–6
(胡日勒, 宗成庆, 徐波. 基于统计学习的机器翻译模板自动获取方法. 中文信息学报, 2005, **19**(6): 1–6)
- 5 Qu Wei, Liu He-Ping, Zhang Hai-Jun. The SVM method based on KL divergence and its application. *Information and Control*, 2005, **34**(5): 627–630
(屈微, 刘贺平, 张海军. 基于 KL 散度的支持向量机方法及应用研究. 信息与控制, 2005, **34**(5): 627–630)
- 6 Smyth P. Clustering sequences using hidden Markov models. In: *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 1997. 648–654
- 7 Chretien S, Hero A O III. Kullback proximal algorithms for maximum-likelihood estimation. *IEEE Transactions on Information Theory*, 2000, **46**(5): 1800–1810

- 8 Du J, Liu P, Soong F K, Zhou J L, Wang R H. Minimum divergence based discriminative training. In: Proceedings of the 9th International Conference on Spoken Language Processing. Pittsburgh, Pennsylvania: 2006. 1030–1033
- 9 Geman S, Geman D. Stochastic relaxation, Gibbs distribution and Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1984, **6**(6): 721–741
- 10 Singer Y, Warmuth M K. Batch and on-line parameter estimation of Gaussian mixtures based on the joint entropy. In: Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems II. Cambridge, USA: MIT Press, 1998. 578–584
- 11 Vasconcelos N. On the efficient evaluation of probabilistic similarity functions for image retrieval. *IEEE Transactions on Information Theory*, 2004, **50**(7): 1482–1496
- 12 Young S J, Evermann G, Gales M, Kershaw D, Odell J, Ollason D. *The HTK Book (for HTK version 3.1)*. Cambridge University, Engineering Department, 2001



王欢良 青岛科技大学讲师. 2007 年于哈尔滨工业大学获博士学位. 主要研究方向为语音信号处理、大词表连续语音识别.

E-mail: huanliangwang@126.com
(**WANG Huan-Liang** Lecturer at Qingdao University of Science and Technology. He received his Ph. D. de-

gree from Harbin Institute of Technology in 2007. His research interest covers speech signal processing and large vocabulary continuous speech recognition.)



韩纪庆 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、语音识别、音频信息检索. 本文通信作者.

E-mail: jqhan@hit.edu.cn
(**HAN Ji-Qing** Professor at School of Computer Science and Technology, Harbin Institute of Technology. His research interest covers speech signal processing, speech recognition, and audio information retrieval. Corresponding author of this paper.)



郑铁然 哈尔滨工业大学计算机学院博士研究生. 现为哈尔滨工业大学讲师. 主要研究方向为口语文档检索和音频信息处理. E-mail: zhengtieran@hit.edu.cn

(**ZHENG Tie-Ran** Ph. D. candidate at School of Computer Science and Technology, Harbin Institute of Technology. Currently, he is a lecturer at Harbin Institute of Technology. His research interest covers spoken document retrieval and audio information processing.)