

# 新型几何谱减语音增强方法

邬鑫锋,曾以成,刘伯权

WU Xin-feng, ZENG Yi-cheng, LIU Bo-quan

湘潭大学 光电工程系,湖南 湘潭 411105

Department of Photoelectric Engineering, Xiangtan University, Xiangtan, Hunan 411105, China

E-mail: xinfengwu@126.com

WU Xin-feng, ZENG Yi-cheng, LIU Bo-quan. Speech enhancement based on new spectral subtraction of geometric approach. *Computer Engineering and Applications*, 2010, 46(23): 144-147.

**Abstract:** In order to get a better performance in processing the speech which changes rapidly, an improved speech enhancement algorithm based on self-adaptive averaging factor method is proposed in this paper. Simulation with Itakura-Saito distance, speech wave picture, spectral picture. Experimental results demonstrate that the proposed algorithm has better performance of speech articulation without any perceptual musicality, compared to another modified spectral subtraction algorithms under the same level of noise reduction, and this superiority is more significant at very low SNR.

**Key words:** speech enhancement; spectral subtraction of geometric approach; Minimum Mean Square Error (MMSE); Itakura-Saito distance

**摘 要:** 针对几何谱减算法在处理快速变化的语音时产生语音畸变的缺点, 提出一种基于最小均方误差算法估计每帧语音信号的每一个频率分量上的平滑系数, 产生自适应帧频率分量平滑系数代替固定值的平滑系数来估计先验信噪比, 从而得到更加接近于真实情况的先验信噪比。通过计算板仓-斋藤距离, 及利用仿真波形图、语谱图对算法进行客观测试, 结果表明新算法相对于其他谱减法在相同的去噪度下, 语音畸变度最小且几乎察觉不到音乐噪声; 特别是在低信噪比非平稳环境下, 相对其他谱减法的优势更加显著。

**关键词:** 语音增强; 几何谱减法; 最小均方误差; 板仓-斋藤距离

DOI: 10.3778/j.issn.1002-8331.2010.23.041 文章编号: 1002-8331(2010)23-0144-04 文献标识码: A 中图分类号: TN912

## 1 引言

语音增强的目的是从带噪语音中提取尽可能纯净的原始语音, 提高语音可懂度。常用的语音增强方法, 大致可分为三类: 第一类是时域方法, 例如基于参数和模型的方法<sup>[1-2]</sup>、子空间的方法<sup>[3-4]</sup>等; 第二类是频域方法, 例如谱减法<sup>[5-7]</sup>、自适应滤波法<sup>[8]</sup>以及基于马尔可夫模型滤波方法<sup>[9]</sup>等; 第三类是其他方法, 例如小波变换法、听觉掩蔽法<sup>[10]</sup>等。谱减法具有易于实现和运算量小等优点, 一直都是语音去噪的最主要的方法。它是通过从含噪语音功率谱中减去噪声功率谱来估计纯净语音的功率谱。因为人耳对语音的感知是通过响应语音信号中各频谱分量的幅度获取的, 而对各分量的相位不敏感, 因此, 就可认为在语音增强过程中, 语音信号的相位不变, 在求得最佳纯净语音幅度值后, 加上原始带噪语音分离出来的相位即可。

在传统谱减算法中, 假设噪声与语音相互独立, 且是零均值的高斯分布, 从而纯净语音与噪声的相关项为零。但现实中的噪声不一定是零均值高斯分布的, 也可能是非零均值、非高斯分布的; 因而传统谱减算法在处理非零均值分布的噪声

时, 增强效果并不是很好, 而且会给结果带来音乐噪声, 不仅使听觉效果受影响, 还影响后续处理, 如语音识别等。为克服传统谱减法的缺陷, 出现了很多改进的算法如多频带谱减法<sup>[11]</sup>、自适应双谱多窗口谱减法、基于人耳掩蔽效应谱减法<sup>[12]</sup>等, 这些改进算法在消除音乐噪声、背景噪声和减少语音失真方面都有一定效果, 但并不理想。EVIN教授根据带噪语音在几何上的关系提出了几何谱减法<sup>[13]</sup>, 此算法能有效地去除音乐噪声, 提高信噪比, 但是增强后的语音畸变度较大。针对这个缺点, 提出用自适应帧频率分量平滑系数代替几何谱减法中固定平滑系数来估计先验信噪比, 即通过估计每帧频率分量平滑系数, 得到最佳先验信噪比, 进而进行几何谱减法语音增强处理, 新算法能够有效去除背景噪声, 且能在去除残留“音乐噪声”度和语音畸变度之间做出很好的折中。

## 2 几何谱减法

传统谱减法存在两个主要缺点: 条件假设不合理与产生音乐噪声, 几何谱减法是克服这些缺点而提出的, 通过利用

**基金项目:** 湖南省自然科学基金(the Natural Science Foundation of Hunan Province of China under Grant No.08JJ5031)。

**作者简介:** 邬鑫锋(1982-), 男, 硕士研究生, 研究方向: 语音信号处理; 曾以成(1962-), 男, 博士, 教授, 博士生导师, 研究方向: 信号处理; 刘伯权(1985-), 男, 硕士研究生, 研究方向: 语音信号处理。

**收稿日期:** 2009-01-15 **修回日期:** 2009-03-30

带噪语音的几何关系得出先验信噪比估计的语音增强算法。设 $\{a_Y, a_X, a_D\}, \{\theta_Y, \theta_X, \theta_D\}$ 分别是含噪语音、纯净语音、噪声的幅值和相位, 有:

$$a_Y e^{j\theta_Y} = a_X e^{j\theta_X} + a_D e^{j\theta_D} \quad (1)$$

由方程(1)可作出带噪语音的几何关系图, 如图1所示。

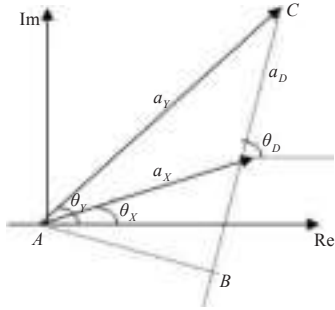


图1 语音的几何关系图

直角三角形 $ABC$ 中体现了含噪语音、纯净语音和噪声三者之间的关系, 其中 $\overline{AB} \perp \overline{BC}$ 。可得:

$$\begin{aligned} \overline{AB} &= a_Y \sin(\theta_D - \theta_Y) = a_X \sin(\theta_D - \theta_X) \Rightarrow \\ a_Y^2 (1 - c_{YD}^2) &= a_X^2 (1 - c_{XD}^2) \end{aligned} \quad (2)$$

其中 $c_{YD} = \cos(\theta_Y - \theta_D)$ ,  $c_{XD} = \cos(\theta_X - \theta_D)$ , 从以上方程可以得到增益函数:

$$H_{GA} = \frac{a_X}{a_Y} = \sqrt{\frac{1 - c_{YD}^2}{1 - c_{XD}^2}} \quad (3)$$

从而可得到纯净语音功率谱的估计:

$$|\hat{X}(\omega_k)|^2 = H_{GA}^2(\omega_k) |Y(\omega_k)|^2 \quad (4)$$

当纯净语音和噪声的向量是正交时(即 $C_{XD}=0$ )两者不相关, 几何谱减法的增益函数和传统谱减法的增益函数是相同的, 用传统谱减法处理即可, 而当遇到其他情况时, 用传统谱减法处理就有误差, 此时纯净语音和噪声的相位关系对增强效果是有影响的, 不能把它们的相关项当零处理。

图1中进一步由三角函数定理可得以下关系:

$$c_{YD} = \frac{a_Y^2 + a_D^2 - a_X^2}{2a_Y a_D} \quad (5)$$

$$c_{XD} = \frac{a_Y^2 - a_X^2 - a_D^2}{2a_X a_D} \quad (6)$$

分子分母同除以 $a_D^2$ 得:

$$c_{YD} = \frac{\gamma + 1 - \xi}{2\sqrt{\gamma}} \quad (7)$$

$$c_{XD} = \frac{\gamma - 1 - \xi}{2\sqrt{\xi}} \quad (8)$$

其中变量 $\xi = \frac{a_X^2}{a_D^2}$ ,  $\gamma = \frac{a_Y^2}{a_D^2}$ , 分别表示先验信噪比和后验信噪比, 从而可以得到用先验和后验信噪比表示的增益函数:

$$H_{GA}(\xi, \gamma) = \sqrt{\frac{1 - \frac{(\gamma + 1 - \xi)^2}{4\gamma}}{1 - \frac{(\gamma - 1 - \xi)^2}{4\xi}}} \quad (9)$$

基于当前帧的后验信噪比 $\hat{\gamma}(\lambda, k)$ 可以通过噪声谱的估计计算 $a_Y^2/a_D^2$ 获得, 而基于当前帧的先验信噪比 $\hat{\xi}(\lambda, k)$ 则需要使用估计的方法得到, 即可利用当前和过去的谱信息, 更确切的说, 可以利用增强了的前一帧幅度谱估计先验信噪比 $\xi$ :

$$\hat{\xi}_l(\lambda, k) = \frac{\hat{a}_X^2(\lambda - 1, k)}{\hat{a}_D^2(\lambda - 1, k)} \quad (10)$$

其中 $\hat{\xi}_l(\lambda, k)$ 表示语音信号第 $\lambda$ 帧第 $k$ 个频率分量上的帧频率分量先验信噪比即时值, 上式的即时先验信噪比仅仅利用了过去的谱信息。

由式(7)得:

$$\xi = \gamma + 1 - 2\sqrt{\gamma} c_{YD} \quad (11)$$

可以利用上式中 $\xi$ 和 $\gamma$ 的关系以及当前帧获得的谱信息来求得 $\xi$ 的估计值:

$$\begin{aligned} \text{因为 } -1 < c_{YD} \leq 1, \text{ 可由式(11)得到 } \xi \text{ 的范围 } (\sqrt{\gamma} - 1)^2 \leq \xi \leq \\ (\sqrt{\gamma} + 1)^2, \text{ 用它的下限作为估计值得:} \end{aligned}$$

$$\hat{\xi}(\lambda, k) = (\sqrt{\hat{\gamma}(\lambda, k)} - 1)^2 \quad (12)$$

其中 $\hat{\gamma}(\lambda, k)$ 表示后验概率 $\gamma$ 在 $\lambda$ 帧 $k$ 频率分量的估计值。综合基于前一帧和当前帧的两种估计值, 就可以得到以下估计方程:

$$\hat{\xi}(\lambda, k) = \alpha \cdot \left[ \frac{\hat{a}_X(\lambda - 1, k)}{\hat{a}_D(\lambda - 1, k)} \right]^2 + (1 - \alpha) \cdot (\sqrt{\hat{\gamma}(\lambda, k)} - 1)^2 \quad (13)$$

$\alpha$ 是平滑系数, 取值介于0.95~1.0之间, 一般情况下取固定值为0.98。 $\hat{a}_D(\lambda, k)$ 是噪声谱的估计值, 若增强后的语音幅值很小就会导致整个结果的衰减, 从而使得语音信号出现谱失真情况, 产生畸变。因此就需要在处理过程的最后加上一个平滑系数, 用以防止处理结果的语音信号出现谱失真的情况。

### 3 几何谱减法的改进

为了准确估计出带噪语音信号的先验信噪比, 公式(13)的估计方法计算当前帧的先验信噪比 $\hat{\xi}(\lambda, k)$ 对于如何合适地选取参数 $\alpha$ 又是十分关键的问题。平滑系数 $\alpha$ 的取值对去噪度和语音畸变度影响很大。例如, 在估计 $\hat{\xi}(\lambda, k)$ 时平滑系数 $\alpha$ 是在跟踪语音变化的准确度和 $\hat{\xi}(\lambda, k)$ 的平滑度之间权衡, 即在语音畸变度和残留音乐噪声度之间折中; $\hat{\xi}(\lambda, k)$ 的标准方差( $\hat{\xi}(\lambda, k)$ 的平滑度影响音乐噪声的大小)随 $\alpha-1$ 增大而增大, 即随着 $\alpha$ 的增大,  $\hat{\xi}(\lambda, k)$ 越不平滑, 带噪语音的变化程度称为“平稳度”, 平滑系数依据“平稳度”自适应变化。因此, 选取一个可靠的平稳度参数是至关重要的。其平滑系数是根据实验取自固定的经验值。然而, 实验表明先验信噪比估计中平滑系数 $\alpha$ 较小时, 语音畸变和噪声残留都较少, 但残留的音乐噪声显著; 反之,  $\alpha$ 大时, 语音畸变和噪声残留较多, 但残留的音乐噪声较少。所以, 对于纯噪声帧和带噪语音帧的平滑系数取值应该是不一样的, 较好的解决办法是随语音出现的状态而时变, 带噪语音帧中平滑系数取较小值, 纯噪声帧中平滑系数取较大值。至于平滑系数时变的依据, 可根据均方最小误差算法估计最佳平滑系数。

在这里采用基于均方最小误差的估计方法估算各帧各个频率分量上的平滑系数 $\alpha_{\lambda, k}$ , 以期得到每帧内各频率分量上的先验信噪比估计值 $\hat{\xi}(\lambda, k)$ , 从而达到在去噪度、残留“音乐噪声”和语音畸变度之间取得很好的均衡。由公式(13)得出先验信噪比的修正方程:

$$\hat{\xi}(\lambda, k) = \alpha_{\lambda, k} \cdot \left[ \frac{\hat{a}_X(\lambda - 1, k)}{\hat{a}_D(\lambda - 1, k)} \right]^2 + (1 - \alpha_{\lambda, k}) \cdot (\sqrt{\hat{\gamma}(\lambda, k)} - 1)^2 \quad (14)$$

其中前一帧各频率分量的先验信噪比为:

$$\hat{\xi}(\lambda - 1, k) = \left[ \frac{\hat{a}_X(\lambda - 1, k)}{\hat{a}_D(\lambda - 1, k)} \right]^2 \quad (15)$$

采用均方最小误差方法来估计第 $\lambda$ 帧第 $k$ 个频率分量的平

滑系数 $\alpha_{\lambda,\kappa}$ 。

由均方最小误差,

$$J_a = E \left\{ \left[ \hat{\xi}(\lambda, \kappa) - \xi(\lambda, \kappa) \right]^2 / \hat{\xi}(\lambda - 1, \kappa) \right\} \quad (16)$$

将公式(14)、(15)代入公式(16)可以推出:

$$J_a = \alpha_{\lambda,\kappa}^2 \left[ \hat{\xi}(\lambda - 1, \kappa) - \xi(\lambda, \kappa) \right]^2 + (1 - \alpha_{\lambda,\kappa})^2 \left[ \xi(\lambda, \kappa) + 1 \right]^2 \quad (17)$$

然后令  $\partial J_a / \partial \alpha_{\lambda,\kappa}$  为零,就可算出基于均方最小误差估计的每帧各频率分量上 $\alpha_{\lambda,\kappa}$ 的最优解:

$$\alpha_{\lambda,\kappa} = \frac{1}{1 + \left[ \frac{\xi(\lambda, \kappa) - \hat{\xi}(\lambda - 1, \kappa)}{\xi(\lambda, \kappa) + 1} \right]^2} \quad (18)$$

由于 $\xi(\lambda, \kappa)$ 是未知量,故直接通过以上公式是无法求出 $\alpha_{\lambda,\kappa}$ 的,但又有以下近似的关系: $E \{ \hat{\xi}(\lambda, \kappa) \} \cong \xi(\lambda, \kappa)$ 。于是就可用 $\hat{\xi}(\lambda, \kappa) = (\sqrt{\hat{\gamma}(\lambda, \kappa)} - 1)^2$ 替代 $\xi(\lambda, \kappa)$ ,并代入公式(18)求出 $\alpha_{\lambda,\kappa}$ 的最优估计值。

可见,利用均方最小误差估计平滑系数 $\alpha_{\lambda,\kappa}$ 能较真实地反

映语音变化情况,在语音变化较为平缓的帧中(纯噪声帧),平滑系数 $\alpha_{\lambda,\kappa}$ 的取值就接近于1;而在语音变化非常剧烈的帧中(带噪声语音帧),平滑系数 $\alpha_{\lambda,\kappa}$ 估计值会变得比较小,以使得估计的先验信噪比能够快速适应语音的变化。

## 4 实验结果与分析

纯净语音取自标准语音库的一段男声:“1至10”10个阿拉伯数字的标准汉语发音。噪声取自 NoiseX92 的白噪声、工厂噪声和 Babble 噪声。将它们合成信噪比为-10~5 dB 的带噪声语音,经过 16 kHz 采样,512 点分帧,帧移 256 点。分别采用传统谱减、几何谱减及本文改进的几何谱减法进行增强实验。

图2所示分别是受 0 dB Babble 噪声污染的语音及经过三种不同的方法处理后的波形图,从图中可以清晰地看出,经过本文方法处理后,语音含有的噪声量相比传统谱减法增强后大为减少。虽然去噪效果没几何谱减法强,但是原始语音得到了很好的恢复。图3所示是受 0 dB 白噪声污染的语音及经过三种不同的方法处理后的语谱图,从语谱图中看出,处理后

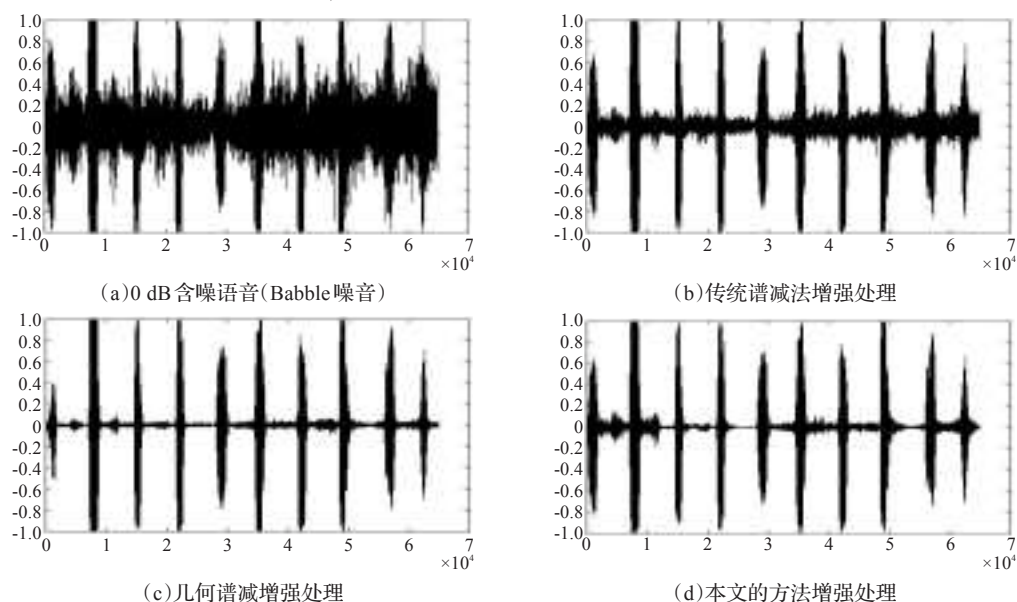


图2 波形图实验

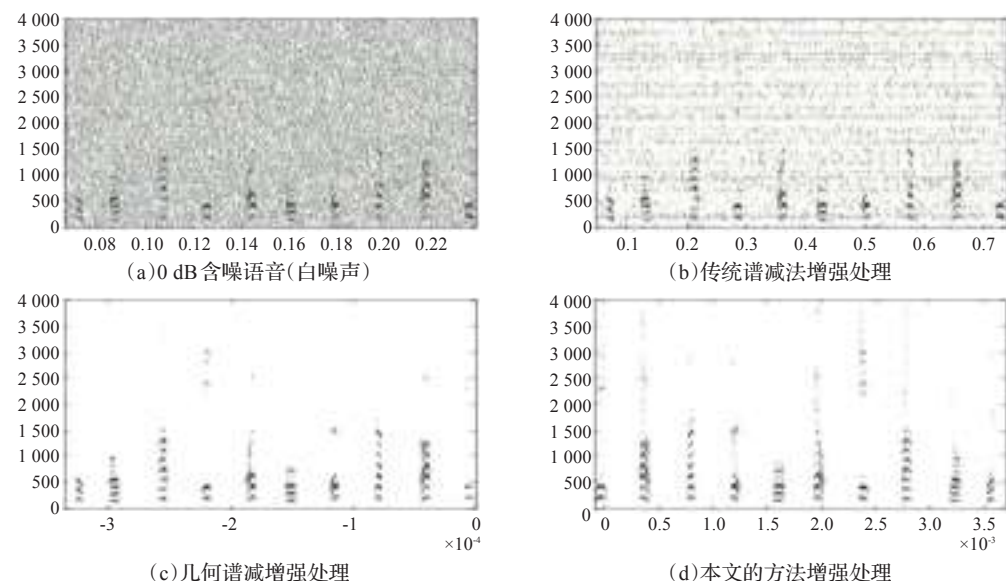


图3 语谱图实验



的语音中残存的白噪声量也减少了,图3(d)中的语谱比前两种方法处理得到的语谱更清晰,特别是发“10”的语音得到很好的恢复。

板仓-斋藤(Itakura-Saito(IS))距离可以作为衡量增强处理后语音的畸变程度的指标。受不同信噪比工厂环境噪声污染的语音在三种不同增强方法处理后与原始语音的板仓距离如图4所示,很明显,在非平稳环境下,基于本文的语音增强算法,有效地降低了板仓距离,语音最为逼近原始语音,即语音畸变度最小。

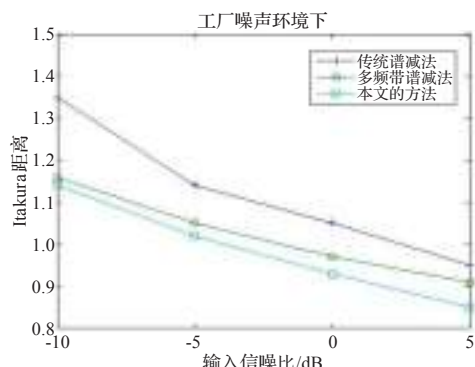


图4 不同输入信噪比下语音增强处理后的I-S距离

主观听觉实验中,本文方法和几何谱减法都感觉不到残留“音乐噪声”和背景噪声,而本文的方法语音畸变度最小,耳听舒适,在有效去除噪声的同时,在残留“音乐噪声”度和语音畸变度之间能找到很好的折中。运算量上,本文方法相对几何谱减法就多了一步均方最小误差计算,运算量增加不多。

## 5 总结

在考虑基于几何谱减增强算法的基础上,自适应地调整先验信噪比估计中的平滑系数,提出了在几何模型下的谱减算法的一种修正形式。实验表明,本文的自适应几何谱减算法优于传统的几何谱减算法,能有效去除噪声的同时,在残留“音乐噪声”度和语音畸变度之间做出很好的平衡,这是因为该算法能实时地自适应地跟踪语音的变化,从而使得几何谱减法中的先验信噪比估计更符合假设的语音模型和人耳听觉特性。

## 参考文献:

- [1] Gannot S, Burshtein D, Weinstein E. Iterative and sequential Kalman filter-based speech enhancement algorithms[J]. IEEE Trans Speech and Audio Processing, 1998, 6(4): 373-385.
- [2] Kin J B, Lee K Y, Lee C W. On the applications of the interacting multiple model algorithm for enhancing noisy speech[J]. IEEE Trans Speech and Audio Process, 2000, 8(3): 349-352.
- [3] Ephraim Y, Trees H L. V.A signal subspace approach for speech enhancement[J]. IEEE Trans Speech and Audio Processing, 1995, 3(7): 251-266.
- [4] Jabloun F, Champagne B. A multi-microphone signal subspace approach for speech enhancement[C]//Proc IEEE ICASSP, 2001, 9(1): 205-208.
- [5] Boll S. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE Trans on Acoustic Speech and Signal Processing, 1979, 27(2): 113-120.
- [6] Fan Ningping. Low distortion speech denoising using an adaptive parametric Wiener filter[C]//IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2004, 5(1): 17-21.
- [7] Ephraim Y, Malah D. Speech enhancement using a minimum mean square error short time spectral amplitude estimator[J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1984, 32(60): 1109-1121.
- [8] 高鹰, 谢胜利. 一种变步长LMS自适应滤波算法及分析[J]. 电子学报, 2001, 29(8): 1094-1097.
- [9] Jax P, Vary P. Artificial bandwidth extension of speech signals using MMSE estimation based on a hidden Markov model[C]//IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2003, 8(1): 680-683.
- [10] Virag N. Signal channel speech enhancement based on masking properties of the human auditory system[J]. IEEE Transactions on Speech and Audio Processing, 1999, 7(2): 126-137.
- [11] Kamath S, Loizou P. A multi-band spectral subtraction method for enhancing speech corrupted by colored noise[C]//IEEE Trans Acoust, Speech Signal Process, 2002, 8(4): 164-168.
- [12] Pu F L, Wang W M, Dai Q J, et al. Optimizing speech enhancement based on noise marked probability[J]. Journal of Electronics & Information Technology, 2005, 27(5): 753-756.
- [13] Lu Yang, Loizou P C. A geometric approach to spectral subtraction[J]. Speech Communication, 2008, 7(50): 453-466.

(上接143页)

的运算转化为类别数的运算。改进后类中心分类法对高特征维, 大量文档分类是有效的。

## 参考文献:

- [1] 焦玉英, 宋晓晴. 基于VSM的文档信息检索改进[J]. 情报理论与实践, 2007, 30(1): 97-104.
- [2] 陈治纲. 于向量空间模型的文本分类系统研究与实现[D]. 天津: 天津大学, 2005.

- [3] 熊忠阳, 李智星, 张玉芳, 等. 基于正交分解的文本分类模型[J]. 计算机工程, 2009, 35(14): 227-229.
- [4] 宋卫东. 析几何[M]. 北京: 高等教育出版社, 2003: 69-107.
- [5] 刘丽珍. 本分类中的特征选取[J]. 计算机工程, 2004(4).
- [6] 刘健, 张维明. 基于互信息的文本特征选择方法研究于改进[J]. 计算机工程与应用, 2008, 44(10): 135-137.
- [7] 吴春颖, 王士同, 蔡崇超. 一种基于新闻发现. Web文本表示方法[J]. 计算机应用, 2008(3).
- [8] 罗杰, 陈力, 夏德麟, 等. 基于新的关键词提取方法的快速文本分类系统[J]. 计算机应用研究, 2006(4): 32-34.