

一种基于新型 BDS 模型的语音信号字典构造方法

张宏乐,李凤莲,张雪英

(太原理工大学 信息工程学院,山西 太原 030000)

摘 要: 用字典训练的方法稀疏表示语音信号成为语音信号处理领域的热点课题,针对字典初始规模的选择问题,提出了一种新的基于新型 BDS 模型的字典训练方法,该方法根据最佳字典规模与稀疏比的关系为字典规模建立模型,可以自适应的为语音信号选择恰当的初始字典规模,克服了 K-SVD 方法依靠经验设置字典规模的缺陷.将加入 BDS 模型的训练字典的方法应用于来自太原理工大学数字音频与视频实验室语音库的语音,进行仿真实验并对实验结果进行了分析.实验结果表明:基于 BDS 模型的语音信号字典构造方法实现了自适应选择最佳字典规模目的,可在保证重构语音质量的同时,进一步提高字典训练的效率.

关键词: K-SVD;稀疏表示;字典训练;最佳字典规模

中图分类号: TN912.3

文献标识码: A

文章编号: 1000—7180(2017)01—0030—05

A Speech Signal Dictionary Construction Method Based on New BDS Model

ZHANG Hong-le, LI Feng-lian, ZHANG Xue-ying

(College of Information Engineering, TaiYuan University of Technology, Taiyuan 030000, China)

Abstract: Recently, using a dictionary learning method to sparse represent speech signal becomes a popular research subject. For selecting the initial dictionary size problem, it is proposed a new dictionary training method that based on new BDS model. This method established model for dictionary size according to the relationship between dictionary size and sparse ratio. This model can self-adaptively choose properly initial dictionary for speech signal. More over, it overcomes the defects of selecting initial dictionary size in K-SVD method on experience. Appling this training dictionary method added the BDS model to speech library of lab of audio and video of Taiyuan University of Technology by simulating experiments. The results are analyzed. Simulation results show that the based on BDS model speech signal dictionary construction method not only realized that choose best dictionary size adaptively, but also improved the efficiency and stability of the dictionary training.

Key words: K-SVD; sparse representation; dictionary learning; best dictionary size

1 引言

语音信号在时频上具有近似稀疏性,用语音的稀疏逼近取代原始语音信号可以从本质上降低语音信号处理的成本.稀疏表示中第一个重要的问题是构造字典.字典可以选为一组预先设定有特定功能的原子,如可操纵的小波^[1], curvelets^[2], 和 band-lets^[3],也可以根据给定的一组信号来设计,使字典

内容更适合特定的信号^[4-6].在文献[7]中,Zhou Yan 等人提出了免疫 K-SVD 算法(IK-SVD),将免疫算法嵌入到追踪算法中,极大的提高了 K-SVD 收敛速度和准确性,利用这种算法对含噪语音进行增强,进一步提高了语音信噪比. Heming Zhao 等人^[8]结合了稀疏算法的分解能力和深度信念网络的非线性网络结构,更好的提取语音信号的非线性特征,提高测谎设备识别率.

尽管有很多研究人员对 K-SVD 算法进行深入研究,却鲜有文献关注所构造字典的规模. 实验证明,字典规模的选取严重影响重构语音质量. 本文分析了字典训练过程,针对字典规模设定问题提出了一种基于最佳字典规模 BDS(Best Dictionary Size)模型的字典训练算法.

2 基于 K-SVD 的字典构造

K-SVD 算法的目的是为信号设计出适合信号本身的过完备字典,使信号可以用字典中最少原子的线性组合进行逼近或表示. K-SVD 算法通过交替进行稀疏编码和字典更新来实现对信号稀疏逼近,相应的目标函数可以使用下式表示:

$$\min_{D, X} \{ \|Y - DX\|_F^2 \} \text{ s. t. } \forall i, \|x_i\|_0 \leq T \quad (1)$$

式中:被稀疏表示的是 Y , 稀疏表示使用的字典记为 $D \in \mathbf{R}^{m \times K}$, 相应的系数矩阵为 X (其中 x_i 表示 X 的第 i 个列向量), $\|\cdot\|_F$ 表示 Frobenius 范数, $\|\cdot\|_0$ 表示 l_0 范数(非零元素的个数), T 为预先设定的稀疏度.

K-SVD 算法通过迭代最小化目标函数(1). 每次迭代都包含两个阶段,即稀疏编码和字典更新.

(1) 稀疏编码阶段

稀疏编码是根据设定好的初始字典,计算信号 Y 在字典 D 上的稀疏系数矩阵 X ,即:在已知 Y 和 D 的情况下,依据式(2)求解 X .

$$\min_{x_i} \{ \|Y - DX\|_2^2 \} \text{ s. t. } \|x_i\|_0 \leq T \quad (2)$$

for $i = 1, 2, \dots, N$

这个问题可以使用任何一种追踪算法来解决,常用的有 OMP (Orthogonal Matching Pursuit) 等算法.

(2) 字典更新阶段

字典更新阶段的目的是逐个更新字典原子. 更新第 k 个原子 d_k (D 的第 k 列) 时,在固定 Y 、 X 和 D 中除第 k 个原子以外的其余原子的情况下,求解 d_k 及相应的系数 x_T^k (X 的第 k 行). 回顾目标函数(1),字典更新阶段的惩罚项可以改写为:

$$\begin{aligned} \|Y - DX\|_F^2 &= \|Y - \sum_{j=1}^K d_j x_T^j\|_F^2 \\ &= \| (Y - \sum_{j \neq k} d_j x_T^j) - d_k x_T^k \|_F^2 \\ &= \| E_k - d_k x_T^k \|_F^2 \end{aligned} \quad (3)$$

上式中,

$$E_k = Y - \sum_{j \neq k} d_j x_T^j \quad (4)$$

E_k 表示除第 k 个原子以外 Y 的逼近误差. 为了使式(3)取得最小,应当有 $d_k x_T^k = E_k$.

为能进行 SVD 分解,定义集合 ω_k 表示含有原子 d_k 成分的样本 $\{y_i\}$ 的索引:

$$\omega_k = \{i | 1 \leq i \leq K, x_T^k(i) \neq 0\}$$

定义 Ω_k 为 $N \times |\omega_k|$ 矩阵,其中 $(\omega_k(i), i)$ 为 1, 其余为 0. 此时,惩罚项与下式等价:

$$\|E_k \Omega_k - d_k x_T^k \Omega_k\|_F^2 = \|E_k^R - d_k x_R^k\|_F^2 \quad (5)$$

式中, $x_R^k = x_T^k \Omega_k$, $E_k^R = E_k \Omega_k$. 对 E_k^R 进行 SVD 分解, $E_k^R = U \Delta V^T$. 记 $\tilde{d}_k$ 为 U 的第一列,用其替换字典中的第 k 个原子 d_k ,即完成了第 k 个原子的更新. 同时,将 x_T^k 更新为 ΔV^T 第一行 $\tilde{x}_T^k$. 其余原子的更新均与此过程一致,经过 K 次更新迭代,可完成所有原子的更新.

3 基于 BDS 的语音字典构造

K-SVD 算法实际应用于语音信号稀疏表示时,需要设定初始字典的规模,即原子的个数. 选择不当的字典规模会严重影响语音的稀疏表示效果. 本文提出一种为语音信号自适应的选择初始字典规模的模型——最佳字典规模模型 BDS(Best Dictionary Size). 该模型可为稀疏算法选取最佳字典规模,可有效避免语音信号的过表示和欠表示现象.

3.1 BDS 模型原理

基于 BDS 模型的字典构造算法是基于 K-SVD 的一种主要针对有效初始字典规模的字典训练算法. 该算法包含两部分的内容——稀疏分解和原子更新. 其中,本算法选择 OMP 算法作为稀疏分解算法. OMP 算法的特点之一是稀疏度已知. 稀疏度 L 是指稀疏系数的非零个数 L .

本文定义另一个描述信号被稀疏的程度的指标——稀疏比(Sparse Ratio, SR). 样本个数与稀疏系数个数的比值为稀疏比,即:

$$SR = n/L \quad (6)$$

式中, n 为样本的个数. 稀疏比越大,线性表示一个样本所需的原子个数就越少,也就是稀疏系数的非零个数少,即越稀疏.

初始化字典时,字典原子的初始选取不显著影响字典的训练,但是字典规模的初始化显著影响语音信号稀疏逼近的效果. 并且,字典规模一旦在初始化中确定就不会再改变. 字典选取过小时容易出现

信号欠表示,字典选取过大时,容易引起信号过表示.为了说明问题,图 1 列出了不同字典规模时某个语音信号的逼近误差时序图.

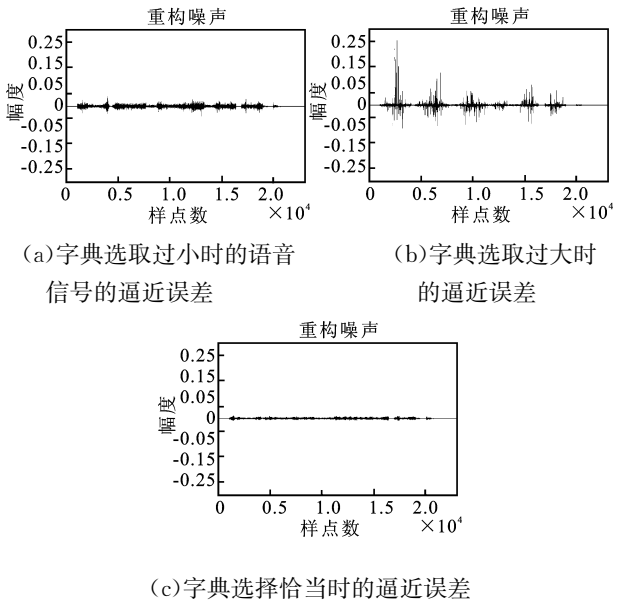


图 1 不同字典规模时逼近误差比较

图 1(c) 表示字典选择恰当时逼近误差, (a) 表示字典选取过小时的语音信号的逼近误差, 与图 (c) 对比可以看出 (a) 中误差较大, 这说明重构出的语音质量较低. 造成这种现象的原因是 OMP 算法无法在可选择的目标较少的情况下追踪出合适的原子. 图 1(b) 表示字典选取过大时的逼近误差, 逼近误差大部分时候很小, 但是存在部分突变点. 这说明重构出的语音信号也出现了突变点, 语音信号的突变有时也被称作“野点”, 听起来像是在语音中夹杂了“啦啦”声. 从时域波形上看, “啦啦”声是一系列随机脉冲在信号波形上的叠加. 引起这种现象的原因是字典更新时不再收敛, 产生了震荡. 无论是信号过表示还是欠表示, 都会引起重构语音的失真, 不利于语音信号的线性逼近.

显然, 对于每段语音信号, 最佳的字典规模是存在的. 实验当中, 字典规模由小至大的变化过程就是语音信号由欠表示到过表示的过程. 此过程中, 最佳的稀疏逼近对应的字典规模即为最佳字典规模.

我们对太原理工大学数字音频与视频实验室录制的语音库中的语音进行测试, 记录测试语句在不同稀疏比情况下最佳原子个数的变化情况. 通过分析得出稀疏比与最佳字典规模之间呈现负相关的关系. 如图 2 所示, 随着稀疏比的增大最佳的字典规模在减小. 这一发现为寻找最佳的字典规模提供了理论基础.

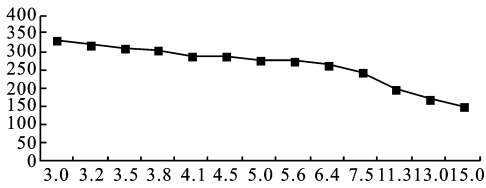


图 2 sp01 语音信号稀疏比与最佳字典规模关系

为了科学、准确的得到它们之间的关系, 通过对数据进行回归分析, 结果指出, 随着稀疏比的增大, 最佳的字典规模呈对数衰减. 由此得到一个满足此变化规律的如下数学关系:

$$y=441.4-107 \cdot \ln(x) \tag{7}$$

我们根据此函数关系为语音库搭建了模型, 该语音库中的语句可以根据这个模型预测恰当的字典规模. 此模型为预测非测试语音的初始字典规模提供了理论支持, 避免了过表示或欠表示现象, 保证了语音稀疏表示的质量. 由于避免了不必要的训练过程, 进一步提高了字典训练的效率.

3.2 基于 BDS 模型的字典构造实现

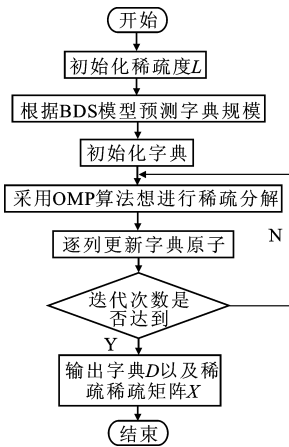


图 3 基于 BDS 模型的字典训练算法流程图

4 实验与分析

为了验证本文算法的有效性, 本文设置分别从最佳字典存在性、稀疏性与最佳字典规模关系和模型预测有效性三个方面进行实验和分析. 对于每段语音的每个实验, 都重复 5 次, 结果取其平均值. 实验所用的仿真软件为 MATLAB.

实验一: 最佳字典存在性的考察

字典规模过小和过大都不利于语音信号稀疏表示. 图 4 给出了稀疏比为 12 时 SNR 与 PESQ 值随字典规模的变化规律图.

从图 4 可以看出, 信噪比与 PESQ 值起初较低, 但随着原子个数的增加, 信噪比与 PESQ 值逐渐增

大,在原子个数为 340 的时候达到最大值,此时的信噪比可以达到 27 dB,PESQ 值可达到 3.7. 随着原子个数的继续增加,语音质量迅速陡然下降,当原子个数为 380 的时候,信噪比下降至 15 dB 以下.

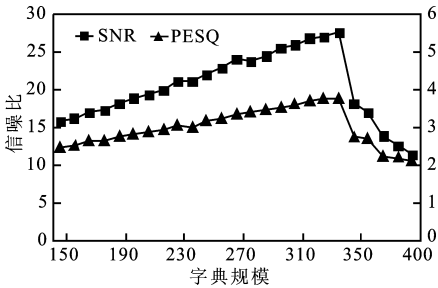


图 4 稀疏比为 12 时 SNR 与 PESQ 值随字典规模变化

字典规模的变化过程也是语音信号经历由欠表示到最佳稀疏表示再到过表示的过程. 字典规模过小会造成欠表示,欠表示时逼近误差较大. 误差会随着字典规模的增大逐渐减弱,直至达到最佳的状态. 后再增加字典规模就会造成语音信号的过表示,导致稀疏表示质量迅速降低.

本实验说明:稀疏比一定时,存在最佳的字典规模值使得语音质量最优.

实验二:考察稀疏比与字典规模的关系

从语音库中抽取 20 组语音信号进行实验,记录不同稀疏比情况下的变化趋势. 表 1 表示字典规模随稀疏比增加的变化趋势.

从表 1 可以看出,随着稀疏比的增大,所需的字典规模在减小. 这说明稀疏比与字典的规模呈负相关. 这是因为语音越稀疏,稀疏后的语音信号携带的信息越少,因此,线性表示其所需的原子个数也会相应减少,原子所属的字典规模也会相应的缩小. 从语音质量上也可以得出相同结论,随着语音携带信息的丢失,语音信噪比(SNR)与主观语音质量评估值(PESQ)均降低了. 对全部数据进行统计分析,可得到稀疏比与最佳字典规模的关系. 图中圆圈代表测

得数据. 为了准确描述上述关系,采用统计方法分析观测值,并对数据进行曲线拟合,如图 5 所示. 假设用 y 表示最佳字典规模,用 x 表示稀疏比. 它们的关系可以用式(7)表示.

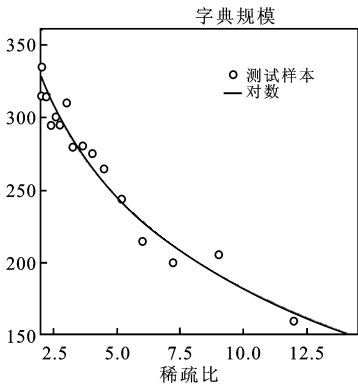


图 5 稀疏比与字典规模之间的关系

实验三:模型有效性验证

为了验证本文提出模型的有效性,我们用另外的 10 句语音信号当做测试样本. 利用模型预测出这十句语音信号在确定稀疏比情况下的最佳字典规模. 利用得到的最佳字典规模初始化学字典,考察此字典是否恰当,考察标准为是否出现过匹配与欠匹配.

利用本文提出的模型得到字典规模,对 sp25 测试语音进行稀疏重构. 稀疏比为 3.8 时,图 6 给出了采用本文算法得到的重构语音波形及逼近误差波形. 可以看出,原始语音和稀疏逼近的语音包络几乎一致,从图(c)中可以看出,逼近误差较小且没有出现突变点,这说明初始字典规模的设置是恰当的. 表 2 给出了 10 句语音在稀疏度为 12 时的 SNR 与 PESQ 值. 从表中可以看出,所有语音信号重构语音质量均较高. 实验表明模型可以准确预测最佳的字典规模. 本文提出的模型保证了语音信号稀疏表示的效果,并提高了训练效率.

表 1 字典规模与稀疏比的关系

稀疏比	2. 01	2. 13	2. 26	2. 41	2. 60	2. 77	3. 00	3. 27
字典规模	335	315	315	295	300	295	310	280
稀疏比	3. 60	4. 02	4. 51	5. 16	6. 01	7. 24	9. 04	12. 01
字典规模	280	275	265	243	215	200	205	160

表 2 基于 BDS 模型的 10 句语音重构质量

	sp21	sp22	sp23	sp24	sp25	sp26	sp27	sp28	sp29	sp30
SNR	26. 55	28. 17	21. 65	25. 33	28. 64	23. 04	31. 30	25. 11	29. 40	28. 71
PESQ	3. 56	3. 55	3. 40	3. 37	3. 72	3. 55	3. 86	3. 68	3. 86	4. 05

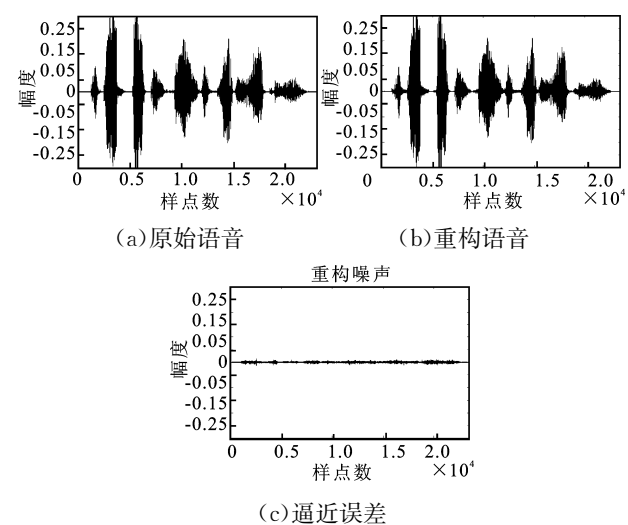


图 6 利用 BDS 模型初始化字典后的
重构语音波形及逼近误差

5 结束语

字典大小的变化严重影响重构语言质量,导致语音过表示或欠表示现象.本文以语音质量为评价标准,统计得出稀疏比与字典大小呈负相关关系,并得出二者之间的数学关系.据此关系建立了 BDS 模型,为字典规模的选择提供了理论依据.实验结果表明,在利用 K-SVD 训练语音字典的过程中,此模型可以有效预测字典规模,保证了稀疏表示语音的重构质量.由于避免了不必要的训练过程,此模型还提高了字典训练效率.综上所述,采用 BDS 模型的字典训练算法实现了自适应的为语音选择最佳字典规模,并确保重构语音性能达到最优.

参考文献:

[1] Mallat S. A wavelet tour of signal processing[M]. USA, Salt Lake City, Academic press, 1999.

(上接第 29 页)

Kerala, IEEE, 2013; 223-228.

[3] Hsieh C Y, Fan M L, Hu V P H, et al. Independently-controlled-gate FinFET Schmitt trigger sub-threshold SRAMs[J]. Very Large Scale Integration (VLSI) Systems, IEEE Transactions on, 2012, 20 (7): 1201-1210.

[4] Seevinck E, List F J, Lohstroh J. Static-noise margin analysis of MOS SRAM cells[J]. Solid-State Circuits, IEEE Journal of, 1987, 22(5): 748-754.

[5] Lim W, Chin H C, Lim C S, et al. Performance evaluation of 14 nm FinFET-based 6T SRAM cell functionality for DC and transient circuit analysis[J]. Jour-

[2] Candès E J, Donoho D L. Curvelets—a surprisingly effective nonadaptive representation for objects with edges[J]. Vanderbilt University Press, 2000, 283(3): 1051-1057.

[3] LePennec E, Mallat S. Sparse geometric image representations with bandelets[J]. IEEE Trans. Image Process, 2005, 14 (4): 423-438.

[4] Aharon M, Elad M, Bruckstein A M. The k-svd: an algorithm for designing of overcomplete dictionaries for sparse representation [J]. IEEE Trans. Signal Process, 2006, 54 (11): 4311-4322.

[5] 黄晓生, 黄萍, 曹义亲, 等. 一种改进的基于 K-SVD 字典学习的运动目标检测算法[J]. 微电子学与计算机, 2014, 31(3): 5-9.

[6] Xu Yangfei, Bao Guangzhao, Xu Xu, et al. Single-channel speech separation using sequential discriminative dictionary learning[J]. Signal Processing, 2015, 106(C): 134-140.

[7] Zhou Yan, Zhao Hening, Shang Li, et al. Immune K-SVD algorithm for dictionary learning in speech denoising[J]. Neurocomputing, 2014, 137(5): 223-233.

[8] Yan Zhou, Heming Zhao, Xinyu Pan, et al. Lie detection from speech analysis based on K-SVD deep belief network model[J]. Springer International Publishing Switzerland, 2015, 9225(1): 189-196.

作者简介:

张宏乐 女, (1989-), 硕士研究生. 研究方向为语音压缩感知. E-mail: 358941469@qq. com.

李凤莲 女, (1972-), 博士, 副教授, 硕士生导师. 研究方向为语音信号处理、大数据及数据挖掘等.

张雪英 女, (1964-), 博士, 教授, 博士生导师. 研究方向为语音信号处理.

nal of Nanomaterials, 2014; ID:802736.

[6] Kou L, Robinson W H. Impact of process variations on reliability and performance of 32-nm 6T SRAM at near threshold voltage [C] // VLSI (ISVLSI), 2014 IEEE Computer Society Annual Symposium on. France, Montpellier, IEEE, 2014; 214-219.

作者简介:

于雨情 女, (1991-), 硕士研究生. 研究方向为集成电路抗辐射设计. E-mail: yuqingyu1211@163. com.

肖立伊 女, (1961-), 博士, 教授, 博士生导师. 研究方向为集成电路可靠性设计.