

文章编号: 1001-0920(2013)07-1109-04

基于主成分分析的字典学习

余付平^{1,2}, 冯有前¹, 范成礼¹, 沈 堤¹

(1. 空军工程大学 防空反导学院, 西安 710051; 2. 中国人民解放军 94559部队, 江苏 徐州 221000)

摘要: 在 K 奇异值字典学习方法的基础上, 结合主成分分析方法提出了 K 主成分分析字典学习方法. 该方法取代了 K 奇异值分解(KSVD)方法中对误差项直接进行SVD分解来更新原子, 取而代之的是通过对误差项进行PCA分解, 提取其主成分作为字典中原子的更新. 仿真结果表明, 与KSVD字典学习方法相比, 所提出的方法字典学习效果更好, 对训练样本的表达误差更小, 学习字典更能表达训练样本的特征.

关键词: K -主成分分析; K 奇异值分解; 字典学习; 稀疏表示

中图分类号: TN957.51

文献标志码: A

Dictionary learning based on principle component analysis

YU Fu-ping^{1,2}, FENG You-qian¹, FAN Cheng-li¹, SHEN Di¹

(1. Institute of Air Force and Anti-Missile, Air Force Engineering University, Xi'an 710051, China; 2. Troops 94559, People's Liberation Army, Xuzhou 221000, China. Correspondent: YU Fu-ping, E-mail: junjingj@163.com)

Abstract: The K -principle component analysis dictionary learning method is proposed based on the K -singular value decomposition(KSVD) method and the principle component analysis(PCA) method. Instead of the SVD decomposition to the error in the KSVD method, the atoms of the dictionary of the method are updated by distilling the principle component of the PCA decomposition. Simulation results show that, compared with the KSVD method, the better learning effect is achieved, the representation error is small, and the learned dictionary reflects the features of the training data much better with the KPCA method.

Key words: K -principle component analysis(KPCA); K -singular value decomposition(KSVD); dictionary learning; sparse representation

0 引言

近年来, 信号稀疏分解已被广泛应用于图像压缩^[1]、图像重构^[2]、图像去模糊和超分辨^[3]、信号去噪^[4-7]、人脸特征识别^[8]等领域, 并取得了很好的效果. 然而, 字典构建是稀疏分解中的关键环节, 关系到信号能否实现稀疏分解. 关于字典构建, 主要分为以下两大类: 1) 根据一定的先验知识构造固定的字典; 2) 在没有先验知识或先验知识不足时, 通过样本训练的方法学习得到自适应的学习字典^[7,9-12]. 由于通过字典学习得到的字典具有良好的自适应性和对训练样本的学习能力, 字典学习已成为国内外很多学者和研究机构在稀疏分解领域研究的一个热点.

目前, 字典学习方法主要包括 K 奇异值分解方法(KSVD)^[2,9]、递归最小二乘(RLS)字典学习方法^[10]、最优方向法(MOD)^[11]和最大后验概率逼近(MAP)^[5]

方法等. KSVD方法已被广泛应用于图像去噪、图像去模糊和图像超分辨等领域, 并取得了很好的效果; RLS方法的应用领域比较少; MOD方法和MAP方法主要用于框架压缩以及框架设计等方面. 为此, 本文结合KSVD字典学习方法和PCA方法的思想, 提出一种基于主成分分析的字典学习方法, 即KPCA字典学习方法.

1 稀疏分解理论

给定一个集合 $D = \{d_j\}_{j=1}^K \in R^N$, 其元素是张成整个 N 维 Hilbert 空间 $H = R^N$ 的单位矢量, 如果其元素个数 $K > N$, 则集合 D 构成一个过完备基, 也称为字典, 其列向量 d_j 称为原子, 且其范数为 1. 通过过完备字典 D , 信号 $y \in R^N$ 能够被表示为其列向量的稀疏线性组合, 即信号具有确切的表达 $y = Dx$, 或者信号具有近似表达 $y \approx Dx$, 且满足 $\|y - Dx\|_p \leq \xi$, 其

收稿日期: 2012-02-10; 修回日期: 2012-11-10.

基金项目: 国家自然科学基金项目(61003148).

作者简介: 余付平(1983-), 女, 工程师, 博士生, 从事信号稀疏分解、雷达信号处理的研究; 冯有前(1960-), 男, 教授, 博士生导师, 从事信号处理等研究.

中 x 为信号 y 的表示系数. 在近似表示中, p 的取值一般为 1, 2 或 ∞ , 本文取 $p = 2$. 由于 $K > N$, D 字典为满秩矩阵, 信号 y 具有无穷多个表示系数. 如何找到非零元素个数最少的表示系数是信号稀疏分解中的重要问题. 信号稀疏分解的实现^[2,13-15]可以表示为

$$\min_x \|x\|_0 \text{ 满足 } y = Dx, \quad (1)$$

或者

$$\min_x \|x\|_0 \text{ 满足 } \|y - Dx\|_2 \leq \xi, \quad (2)$$

其中 $\|\cdot\|_0$ 表示向量中非零元素的个数. 然而, 上述问题的求解是一个 NP 难题. 为了解决这个问题, 人们提出了许多稀疏分解算法^[2,13], 如匹配追踪 (MP) 算法、基追踪 (BP) 算法、框架方法 (MOF) 和最优正交基 (BOB) 算法等.

2 KPCA 字典学习方法

2.1 PCA 方法^[3,6]

PCA 是一种典型的消除信号相关的方法, 已广泛应用于信号去噪、信号降维、模式识别和数据压缩等方面. 信号 Y 进行 PCA 分解的目的是寻找信号的主成分分量, 消除信号的相关性. 信号 PCA 分解的过程如下:

定义向量 $y = [y_1 \ y_2 \ \cdots \ y_m]^T$, 信号

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} y_1^1 & y_1^2 & \cdots & y_1^n \\ y_2^1 & y_2^2 & \cdots & y_2^n \\ \vdots & \vdots & \ddots & \vdots \\ y_m^1 & y_m^2 & \cdots & y_m^n \end{bmatrix}$$

为 $m \times n$ 的矩阵, 均值矩阵

$$\bar{Y} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_m \end{bmatrix} = \begin{bmatrix} \bar{y}_1^1 & \bar{y}_1^2 & \cdots & \bar{y}_1^n \\ \bar{y}_2^1 & \bar{y}_2^2 & \cdots & \bar{y}_2^n \\ \vdots & \vdots & \ddots & \vdots \\ \bar{y}_m^1 & \bar{y}_m^2 & \cdots & \bar{y}_m^n \end{bmatrix}.$$

其中

$$\bar{y}_i = y_i - \mu_i, \quad \mu = [\mu_1 \ \mu_2 \ \cdots \ \mu_m]^T, \quad \mu_i = \frac{1}{n} \sum_{j=1}^n y_i^j.$$

则信号 Y 进行 PCA 分解的过程如下: 首先求解协方差矩阵 $\Omega = E(\bar{y}\bar{y}^T) \approx \frac{1}{n} \bar{Y}\bar{Y}^T$; 然后对协方差矩阵 Ω 进行奇异值分解, 得到 $\Omega = \Phi \Lambda \Phi^T$. 因此, 信号 Y 的估计值 $\hat{Y} = PY$, 其中 $P = \Phi^T$.

2.2 KPCA 方法

给定一系列训练信号 $\{y_i, i = 1, 2, \dots, N\}$ 组成矩阵 Y , 训练信号 y_i 在字典 D 上的分解系数表示为 x_i , 则稀疏表示系数矩阵 $X = \{x_i, i = 1, 2, \dots, N\}$. 因此, Y 在字典 D 上的稀疏分解问题可以表示为

$$\min_X \sum_i \|x_i\|_0 \text{ 满足 } \|Y - DX\|_F^2 \leq \varepsilon, \quad (3)$$

或者

$$\min_X \|Y - DX\|_F^2 \text{ 满足 } \forall i, \|x_i\|_0 \leq T_0. \quad (4)$$

字典学习通过一系列训练数据学习训练一个字典, 使得训练信号在该字典上总体表示误差最小, 从而使得训练得到的字典能够更好地适应所要解决的问题. 因此, 对于训练信号 Y , 寻找可能的最优字典的问题可以表示为

$$\min_{D, X} \sum_i \|x_i\|_0 \text{ 满足 } \|Y - DX\|_F^2 \leq \varepsilon, \quad (5)$$

或者

$$\min_{D, X} \|Y - DX\|_F^2 \text{ 满足 } \forall i, \|x_i\|_0 \leq T_0. \quad (6)$$

式 (6) 的目标函数可以表示为

$$\begin{aligned} \|Y - DX\|_F^2 &= \left\| Y - \sum_{j=1}^N d_j x_T^j \right\|_F^2 = \\ &= \left\| \left(Y - \sum_{j \neq k} d_j x_T^j \right) - d_k x_T^k \right\|_F^2 = \\ &= \|E_k - d_k x_T^k\|_F^2. \end{aligned} \quad (7)$$

KPCA 方法结合了 KSVD 方法和 PCA 方法, 它通过对式 (7) 中的 E_k 利用 PCA 方法进行分解, 得到

$$\Omega \approx \frac{1}{N} (\bar{E}_k \bar{E}_k^T) = PAP^T, \quad (8)$$

其中 P 为主成分矩阵. 将 Λ 中最大的特征值对应的 P 中的特征向量作为 d_k 的更新. 从式 (8) 可看出, KPCA 方法与 KSVD 方法的计算复杂度相同.

3 仿真实验

3.1 随机信号的字典学习

为了研究 KPCA 字典学习方法字典学习效果, 随机生成大小为 20×50 的生成字典, 且字典的每一列均被归一化. 根据生成字典生成 1500 个训练样本信号, 且每个样本信号由初始字典中 3 个不同原子线性组合生成, 每个样本信号的采样点数为 20. 令 $E = Y - DX$, 则定义样本总误差 = $\sum_{j=1}^n \sum_{i=1}^m E(i, j)^2$. 原子比率为学习字典与生成字典原子之间乘积大于某一值的个数; 相似度为学习字典与生成字典之间的列向量点乘绝对值之和. 通过 100 次迭代, 得到图 1 所示两种字典学习方法样本总误差随着迭代次数的变化曲线. 从图 1 可以看出, KPCA 字典学习方法得到的样本总误差小于 KSVD 字典学习方法获得的样本总误差, 前者的字典学习效果更好, 通过学习得到的字典更好地反映了训练样本的信息.

在训练样本中抽取第 1000 个样本信号, 利用两种字典学习方法得到的学习字典和相关分解系数对第 1000 个样本信号进行重构. 两种方法对应的重构信号和原始信号如图 2 所示. 从图 2 可以看出, 两种字典学习方法对信号重构的效果的差别很小, 但均能够较好地重构原始信号, 且信号误差均较小.

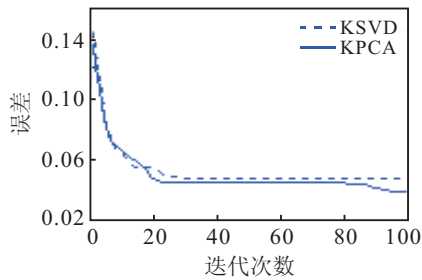


图1 样本表达总误差随迭代次数的变化

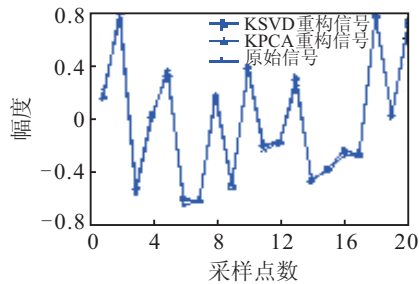


图2 第1000个样本信号的重构信号

两种字典学习方法在100次迭代后总的运行时间、原子比率、相似度以及第100次迭代后的总样本误差等指标如表1所示. 从表1可以看出, KPCA字典学习方法在原子比率、相似度以及总误差方面均优于KSVD字典学习方法, 但KSVD字典学习方法运行时间优于KPCA方法. 所以与KSVD字典学习方法相比, KPCA字典学习方法获得了较好的字典学习效果, 但牺牲了计算速度.

表1 两种字典学习方法结果对比

字典学习方法	运行时间/s	原子比率	相似度	误差
KSVD	54.480 57	3 632	7.953 1	0.047 1
KPCA	80.578 26	3 726	8.868 0	0.038 7

3.2 雷达信号训练样本的字典学习

为了进一步研究KPCA字典学习方法对实际信号的字典学习效果, 以雷达信号为训练样本, 样本个数为4000, 采样点数为16, 生成字典大小为 16×32 . 通过Matlab仿真, 得到的样本总表达误差曲线如图3所示. 从图3可以看出, 利用KPCA字典学习方法得到的学习字典使得样本表达总误差较小, 即与KSVD学习字典相比, KPCA学习字典更好地表达了训练样本, 更好地学习了样本的信息.

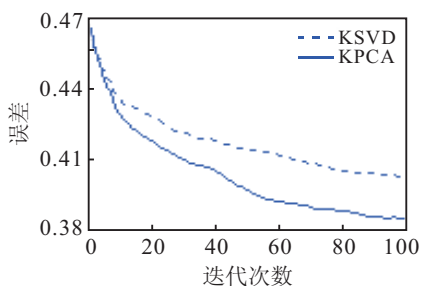


图3 样本表达总误差随迭代次数的变化

图4为通过两种字典学习方法对第3000个雷达信号训练样本的重构信号. 从图4可以看出: 通过KPCA字典学习方法重构的信号与原始信号更相似, 差别较小, 大致反映了原始信号的变化规律和特征; 而KSVD字典学习方法重构的信号与原始信号之间的差别较大, 丢失了原始信号的部分特征.

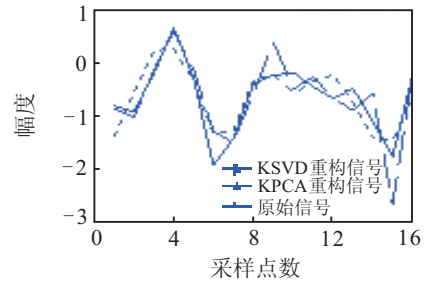


图4 第3000个样本信号的重构信号

表2为两种字典学习方法针对雷达信号训练样本得到的结果对比. 从表2可以看出, KPCA字典学习方法在原子比率、相似度以及总误差方面优于KSVD字典学习方法, 但运行时间较长.

表2 两种字典学习方法结果对比

字典学习方法	运行时间/s	原子比率	相似度	误差
KSVD	180.715 23	806	8.552 2	0.402 31
KPCA	369.547 87	812	9.144 4	0.385 502

4 结 论

本文在结合KSVD字典学习方法和PCA方法的基础上, 提出了KPCA字典学习方法. 通过仿真实验验证可以得到如下结论: 与KSVD字典学习方法相比, KPCA字典学习方法具有较好的字典学习效果, 学习得到的字典与生成字典更相似, 更能反映训练样本的信息. 但是, KPCA字典学习方法在获得较好的字典学习效果的同时牺牲了效率, 计算速度较慢. 如何在保证字典学习效果的同时改进计算速度尚待进一步研究.

参考文献(References)

- [1] Skretting K, Engan K. Image compression using learned dictionaries by RLS-DLA and compared with K -SVD[C]. Proc of IEEE Int Conf on Acoust, Speech, Signal Process(ICASSP). Pargue, 2011: 1517-1520.
- [2] Elad M. Sparse and redundant representations from theory to applications in signal and image processing[M]. New York: Springer, 2010: 286-294.
- [3] Weisheng Dong, Lei Zhang, Guangming Shi, et al. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization[J]. IEEE Trans on Image Processing, 2011, 20(7): 1838-1856.

- [4] Protter M, Elad M. Image sequence denoising via sparse and redundant representations[J]. IEEE Trans on Image Processing, 2009, 18(1): 27-36.
- [5] Javier S Turek, Irad Yavneh, Michael Elad. On MMSE and MAP denoising under sparse representation modeling over a unitary dictionary[J]. IEEE Trans on Signal Processing, 2011, 59(8): 3526-3535.
- [6] Lei Zhang, Weisheng Dong, David Zhang, et al. Two-stage image denoising by principal component analysis with local pixel grouping[J]. Pattern Recognition, 2010, 43(4): 1531-1549.
- [7] Cherian A, Sra S, Papanikolopoulos N. Denoising sparse noise via online dictionary learning[C]. Proc of IEEE Int Conf on Acoust, Speech, Signal Process(ICASSP). Pargue, 2011: 2060-2063.
- [8] Yong Xu, David Zhang, Jian Yang, et al. A two-phase test sample sparse representation method for use with face recognition[J]. IEEE Trans on Circuits and Systems for Video Technology, 2011, 21(9): 1255-1262.
- [9] Michal Aharon, Michael Elad, Alfred Bruckstein. K -SVD: An algorithm for designing overcomplete dictionaries for sparse representation[J]. IEEE Trans on Signal Processing, 2006, 54(11): 4311-4322.
- [10] Karl Skretting, Kjersti Engan. Recursive least squares dictionary learning algorithm[J]. IEEE Trans on Signal Processing, 2010, 58(4): 2121-2130.
- [11] Engan K, Aase S O, Husoy J H. Method of optimal directions for frame design[C]. Proc of IEEE Int Conf on Acoust, Speech, Signal Process(ICASSP). Phoenix, 1999: 2443-2446.
- [12] Ivana Tomic, Pascal Frossard. Dictionary learning for stereo image representation[J]. IEEE Trans on Image Processing, 2011, 20(4): 921-934.
- [13] Gribonval R, Schnass K. Dictionary identification: Sparse matrix-factorization via l_1 minimisation[J]. IEEE Trans on Information Theory, 2010, 56(7): 3523-3539.
- [14] Maria G Jafari, Mark D Plumbley. Fast dictionary learning for sparse representations of speech signals[J]. IEEE J of Selected Topics in Signal Processing, 2011, 5(5): 1025-1031.
- [15] Kai Labusch, Erhardt Barth, Thomas Martinetz. Robust and fast learning of sparse codes with stochastic gradient descent[J]. IEEE J of Selected Topics in Signal Processing, 2011, 5(5): 1048-1060.

(上接第1108页)

- [6] Gebauer H, 王春芝. 制造企业服务业务扩展及其认知因素研究[J]. 中国管理科学, 2006, 14(1): 69-75.
(Gebauer H, Wang C Z. Study on extending service business and the cognitive factors for manufacturing company[J]. Chinese J of Management Science, 2006, 14(1): 69-75.)
- [7] Chun Y H, Tang K. Determining the optimal warranty price based on the producer's and customer's risk preferences[J]. European J of Operational Research, 1995, 85(1): 97-110.
- [8] Decroix G A. Optimal warranties, reliabilities and prices for durable goods in an oligopoly[J]. European J of Operational Research, 1999, 112(3): 554-569.
- [9] Wu C C, Lin P C, Chou C Y. Determination of price and warranty length for a normal lifetime distributed product[J]. Int J of Production Economics, 2006, 102(1): 95-107.
- [10] 沈铁松, 熊中楷. 考虑厂商承诺行为的产品延伸服务市场竞争分析[J]. 中国管理科学, 2010, 18(4): 93-100.
(Shen T S, Xiong Z K. Analysis of market competition on product supplementary service by considering firms commitment behavior[J]. Chinese J of Management Science, 2010, 18(4): 93-100.)
- [11] Iyer G. Coordination channel under price and non price competition[J]. Marketing Science, 1998, 17(4): 338-355.
- [12] Taylor T A. Supply chain coordination under channel rebates with sales effort effects[J]. Management Science, 2002, 48(8): 992-1007.
- [13] Krishnan H, Kapuscinski R, Butz D A. Coordinating contracts for decentralized supply chains with retailer promotional effort[J]. Management Science, 2004, 50(1): 48-63.
- [14] 田厚平, 郭亚军, 杨耀东. 分销系统中多委托人及多委托人可能合作的委托代理模型[J]. 系统工程理论方法应用, 2004, 13(4): 361-366.
(Tian H P, Guo Y J, Yang Y D. The analysis on principal-agent theory in distribution system with multi-principal and the cooperation between principals[J]. Systems Engineering-Theory Methodology Applications, 2004, 13(4): 361-366.)
- [15] 鲁其辉. 基于成本共担策略的服务供应链协调研究[J]. 控制与决策, 2011, 26(11): 1649-1653.
(Lu Q H. Coordinate with cost sharing strategy for service supply chain[J]. Control and Decision, 2011, 26(11): 1649-1653.)
- [16] Heese S. Supply chain dynamics under extended warranty sales[R]. Bloomington: Indiana University, 2008.