第 24 卷,第 1 期 中国传媒大学学报自然科学版 Vol 24,No 1
2017 年 2 月 JOURNAL OF COMMUNICATION UNIVERSITY OF CHINA(SCIENCE AND TECHNOLOGY) Feb,2017

字典学习的 K - SVD 算法分析

牛秀秀,华敏杰,狄燕飞,相鹏

(中国传媒大学 理工学部,北京 100024)

摘要: 分析了字典学习的 K - SVD 算法,通过引入 K - Means 计算方法,将 K - Means 方法推广到用于字典学习的 K - SVD 计算方法中;分析和描述了 K - SVD 计算过程,指出了 K - SVD 方法与 K - Means 方法之间的关系,最后观察图像数据训练用于稀疏表示的字典。

关键词: K - Means 方法;字典学习;稀疏表示;K - SVD 方法

中图分类号: TN911.73 **文献标识码**: A **文章编号**: 1673 - 4793(2017)01 - 0047 - 04

DOI:10.16196/j.cnki.issn.1673-4793.2017.01.008

The K-SVD Analysis of Dictionary Learning

NIU Xiu-xiu, HUA Ming-jie, DI Yan-fei, XIANG Peng

(Science of School, Communication University of China, Beijing 100024, China)

Abstract: The dictionary learning method i. e. the K - SVD algorithm has been analyzed. We have also ex - tended to K - means to K - SVD method through using some ideas in K - means algorithm. The K - SVD algorithm to solve real problems has been analyzed and given in detailed steps. The differences and similarities between K - SVD and K - Means have been provided. The learned dictionary has been ob- tained by theobserved image data based on the numerical experiments.

Key words: K - Means algorithm; dictionary learning; sparse representations; K - SVD algorithm

1 引言

图像去噪问题是非常重要的,不仅仅是因为在程序上的应用,而且作为最简单的反问题,给图像处理在技术和理念上提供了一个方便的平台。在过去的 50 年左右,许多人有着不同的观点,各种统计估计、空间自适应滤波器、偏微分方程、样条函数等等很多方向都在研究这个问题。在本文中主要专注一个特定的方法来解决图像去噪问题:在稀疏表示下的字典学习。

K - SVD 算法是 2006 年由以色列理工学院 Michal Aharon、Michael Elad 等人[1]提出来的,是一种非常经典的字典训练算法,并且达到了很好的训练效果。其目的是解决下列等式的解:

$$Y = DX , \tag{1}$$

其中 D 是要训练的字典,X 是字典 D 对应的稀疏系数向量。当矩阵的维数过高时,即使在 Matlab 中也很难求得(1)的解。研究表明,K - SVD 算法可以比较简便的求解问题(1)。

2 K - Means 算法

首先明确 K - Means 算法[2 - 4]解决的问题是:求解包含 K 个代码的码本,根据最近邻分配法则,对包含 N 个信号的集合 $Y = \{y_i\}_{i=1}^{N}(N >> K)$

收稿日期: 2016 - 4 - 15

作者简介: 牛秀秀(1991 -),女,(汉族),安徽省淮北人,中国传媒大学硕士研究生. E - mail: 393908086@qq.com

在此码本上进行分类，然后得到最佳分类的问题。在本文中，稀疏编码 X 先利用追踪算法求出，作为已知条件。

在矢量量化(VQ)中，可以通过 K－Means 方法来对码本进行训练，假定码本为 $C = [c_1, c_2, \ldots, c_K]$，代码是 C 中的列 c_i。当码本 C 给定时，每个信号用最近(l^2 范数下)的一个代码表示。我们也可以写作 $y_i = Cx_i$，其中 $x_i = e_j$ 是自然基中的一个向量(除了第 j 个值为1，其余为0)。j 满足

$$\forall_{k \neq j} \| y_i - Ce_j \|_2^2 \leq \| y_i - Ce_k \|_2^2$$

这种表示方法中，y_i 的方差为 $e_i^2 = \| y_i - Cx_i \|_2^2$，则 Y 的量化误差由 $E = \sum_{i=1}^{N} e_i^2 = \| Y - CX \|_F^2$ 来确定。从而 K－Means 的目标函数为

$$\min_C \{ \| Y - CX \|_F^2 \} \qquad \forall i, x_i = e_k$$

我们可以发现 K－Means 算法就是一个对码本 C 进行更新迭代的过程。

3 K－SVD 算法——广义 K－Means 算法

在讲述 K－SVD 算法之前，我们首先要了解奇异值分解。奇异值分解就是假设 M 是一个 $m \times n$ 阶矩阵，其中的元素全部属于域 K(实数域或复数域)。如此则存在一个分解使得 $M = U\Delta V^*$，其中 U 是 $m \times m$ 阶酉矩阵；Δ 是半正定 $m \times n$ 阶对角矩阵；而 V^*，即 V 的共轭转置，是 $n \times n$ 阶酉矩阵。这样的分解就称作 M 的奇异值分解。Δ 对角线上的元素 Δ_i，Δi 即为 M 的奇异值。

本文我们研究方程

$$\min_x \| x \|_0 \quad \text{s.t} \quad y = Dx \tag{2}$$

其中，令 $D \in R^{n \times K}$ 代表字典，$y \in R^n$ 代表训练信号，$x \in R^k$ 代表训练信号的稀疏表示系数向量，给出 N 个训练信号的集合 $\{y_i\}_{i=1}^N$，其解向量的集合为 $\{x_i\}_{i=1}^N$。原问题(2)从线性组合角度来看，可以化成

$$\min_D \{ \| Y - DX \|_F^2 \} \quad \text{s.t} \quad \forall i, \| x_i \|_0 \leq T_0 \tag{3}$$

我们可以发现求解(3)的过程就是一个迭代过程，具体迭代过程如下：

将字典 D 逐列进行更新：首先假设字典 D 和稀疏矩阵 X 都是固定的，下面要更新字典的第 k 列 d_k。记稀疏矩阵 X 与 d_k 相乘的那行为 x_T^k(不同于 X 的第 k 列 x_k 的转置)，则(3)式中惩罚项可记为

$$\| Y - DX \|_F^2 = \| Y - \sum_{j=1}^{K} d_j x_T^j \|_F^2$$
$$= \| (Y - \sum_{j \neq k} d_j x_T^j) - d_k x_T^k \|_F^2 = \| E_k - d_k x_T^k \|_F^2 \tag{4}$$

其想法是将 DX 分解为 k 个秩为1的矩阵的和，假设其中的 $k-1$ 项都是固定的，剩下的1列就是要更新的，矩阵 E_k 表示的是去掉原子 d_k 的成分在所有 N 个样本中造成的误差。如果此时利用奇异值分解(SVD)来更新 d_k 和 x_T^k，SVD 可以找到距离 E_k 最近的且秩为1的矩阵。但是其中的 x_T^k 不一定稀疏，所以更新的 d_k 不满足稀疏条件。也就是说用 SVD 得到的 x_T^k 更新向量中的非零值的位置和数量与原 x_T^k 不同，会出现“发散”现象。

因此，为了消除其“发散”现象，也就是先将 x_T^k 中所有的零元素缩减，仅保留非零值，然后再用 SVD 更新 d_k 和 x_T^k，便不会出现“发散”现象。从而，定义集合 $\omega_k = \{ i | 1 \leq i \leq N, x_T^k(i) \neq 0 \}$，即 ω_k 为用到 d_k 所有集合 $\{y_i\}$ 的索引所构成的集合，也就是 $x_T^k \neq 0$ 的点的索引值。定义 Ω_k 是一个 $N \times |\omega_k|$ 的矩阵，其中在点 $(\omega_k(i), i)$ 处为1，其他处为0。定义 $x_R^k = x_T^k \Omega_k$，$Y_R^k = Y\Omega_k$，$E_R^k = E_k \Omega_k$，从而三者分别是 x_T^k，Y，E_k 去掉0后的收缩结果，Y_R^k 为当前用到原子 d_k 样本的集合，E_R^k 为去掉不受原子 d_k 影响的样本，如果不考虑 d_k 在受其影响的样本中的成分时带来的误差。此时，可将式(4)化为

$$\| E_k \Omega_k - d_k x_T^k \Omega_k \|_F^2 = \| E_R^k - d_k x_R^k \|_F^2 \tag{5}$$

从而将 E_R^k 进行 SVD 分解，则 $E_R^k = U\Delta V^T$，即可对 d_k 进行更新，令 $\tilde{d}_k$ 是 U 的第一列，则 $\tilde{d}_k$ 是 d_k 更新的结果。同时，用 V 的第一列和 $\Delta(1,1)$ 的乘积对 x_R^k 进行更新。之后用新字典做稀疏分解，达到停止条件(达到迭代次数或者原信号与重构信号的误差率)则停止，否则迭代继续。此解决方案必须满足两个条件：1) D 的列保持规范化；2) 所有表示的支持度保持不变或者变小。

总结下来得到 K－SVD 算法过程：

1. 在已知稀疏矩阵 X 以及数据样本 $\{y_i\}_{i=1}^N$ 的情况下，给出要解决的问题：

$$\min_D \{ \| Y - DX \|_F^2 \} \quad s.t \quad \forall i, \| x_i \|_0 \leq T_0$$

2. 给出初始字典 $D^{(0)} \in R^{n\times K}$，其中的列向量都是 l^2 范数下的标准形式。给定 $J = 1$。

3. 对 $D^{(J-1)}$ 中的每列 $k = 1,2,\ldots K$ 进行迭代：

(1) 通过 $E_k = Y - \sum_{j\neq k} d_j x_T^j$ 来计算误差；

(2) 由 E_k 通过 ω_k 的限定得到 E_R^k；

(3) 对 E_R^k 进行奇异值分解得到 $E_R^k = U\Delta V^T$，令 $\tilde{d}_k$ 是 U 的第一列，则 $\tilde{d}_k$ 是 d_k 更新的结果。同时，用 V 的第一列和 $\Delta(1,1)$ 的乘积对 x_R^k 进行更新。

最后，$J = J + 1$ 继续重复迭代过程，直到满足停止条件。

我们看到 K－SVD 可以看做 K－Means 的一种泛化形式，K－Means 算法中每个信号只能用一个原子来近似表示，而 K－SVD 算法可看做广义的矢量量化(VQ)，其中每个信号可以用多个原子的线性组合来表示。因此，我们可以发现当 K－SVD 算法中要求的每个信号只用一个原子来近似时，K－SVD 算法就退化为 K－Means 算法。

4　数值实验

我们用 Matlab 对 K－SVD 算法进行了编程，从脸图像数据库中找到训练数据，其由 11000 例像素为 8×8 的小块构成，按照他们的方差随机抽 500 个构成训练的图像如图 1。

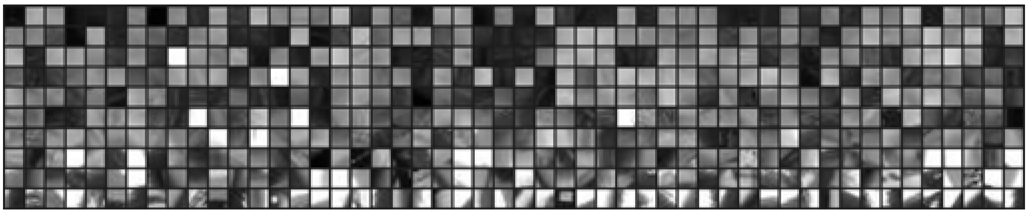

图 1

为了运行 K－SVD 算法，我们还要给出要字典的大小为 64×256，得到训练字典的图像如图 2。

然后，利用 K－SVD 算法得到的字典对观察图像进行去噪，此时选取两个图像的大小为 512×512，从而得到去噪后的图像如图 3。

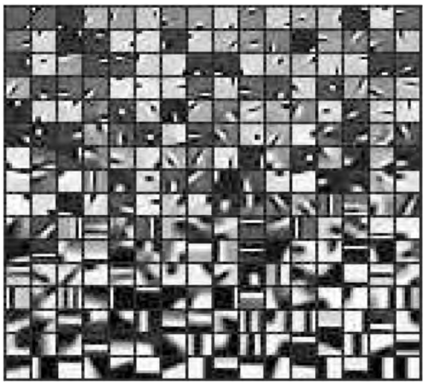

图 2

图 3

5 结论

本文重点分析了 K－SVD 算法的计算过程，由于 K－SVD 算法针对不同的图像均有较好的适应性，并且能获得更好的恢复效果，因此，在图像学习中得到普遍运用。

参考文献

[1] Aharon M，Elad M，Bruckstein A M. The K－SVD: An algorithmfor designing of overcomplete dictionaries for sparse representation [J]. IEEE Trans Signal Process，2006，54(11): 4311－4322.

[2] Pati Y C，Rezaiifar R，Krishnaprasad P S. Orthogonal matching pursuit: Recursive function approximation with applications to waveletdecomposition [C]. 27th Annu Asilomar Conf Signals，Systems，and Computers，1993.

[3] Mallat S，Zhang Z. Matching pursuits with time－frequency dictionaries [J]. IEEE Trans Signal Process，1993，41(12): 3397－3415.

[4] Gersho A，Gray R M. Vector Quantization and Signal Compression [M]. New York: Springer，1991.

(责任编辑: 王谦)

(上接第 40 页)

[38] Whissell C，Fournier M，Pelland R. A dictionary of affect in language: Reliability，validity，and applications [J]. Perceptual and Motor Skills，1986，62(3): 875－888.

[39] Mao X，Xue Y L，Cheng L L. Harmonious graphics generating based on the 1/f function theory [J]. Chaos Solitons & Fractals，2007，32(2): 521－525.

[40] Sato K，Mitsukura Y. Analysis of the interaction between image and music focused on physical features [J]. Sice Conference，IEEE，2012: 2179－2184.

[41] Sato K，Mitsukura Y. Effects of music on image impression and relationship between impression and physical feature [J]. Human System Interactions(HSI)，2011 4th International Conference，IEEE，2011: 369－373.

[42] Collignon O，Girard S，Gosselin F. Audio－visual integration of emotion expression [J]. Brain Research，2008，1242(4): 126－135.

[43] Hinton G E，Osindero S，Teh Y W. A Fast Learning Algorithm for Deep Belief Nets [J]. Neural Computation，2006，18(7): 1527－1554.

[44] Salakhutdinov R，Hinton G. An efficient learning procedure for deep Boltzmann machines [J]. Neural Computation，2012，24(8): 1967－2006.

[45] Hinton G E，Salakhutdinov R R. Reducing the Dimensionality of Data with Neural Networks [J]. Science，2006，313(5786): 504－507.

[46] Srivastava N，Salakhutdinov R. Multimodal Learning with Deep Boltzmann Machines [J]. Journal of Machine Learning Research，2014，15(8): 1967 － 2006.

[47] Srivastava，N Salakhutdinov R. Learning representations for multimodal data with deep belief nets [J]. International Conference on Machine Learning Workshop(2012).

[48] Rasiwasia N，Costa Pereira J，Coviello E. A new approach to cross－modal multimedia retrieval [J]. International Conference on Multimedea 2010，Firenze，Italy，2010: 251－260.

[49] Zhang M L，Zhou Z H. A Review on Multi－Label Learning Algorithms [J]. IEEE Transactions on Knowledge & Data Engineering，2014，26(8): 1819－1837.

(责任编辑: 宋金宝)