

信号稀疏表示方法的研究进展

朱亚菲

2014 年 10 月 30 日

1 背景

1.1 稀疏表示的视觉神经学基础

人类视觉系统如何对外界信息进行感知与处理一直是脑科学、视觉神经科学及心理学研究的重要问题。目前的研究成果证实，人类视觉系统能够用有限的视觉神经元感知复杂多变的外界输入模式。对此 Barlow 认为视觉信息神经处理过程的一个重要原则是信息编码的有效性,Olshausen 和 Field 进一步提出了稀疏编码原则。稀疏编码的基本原理可以描述为: 对于一个给定的自然图像刺激, 仅有一小部分视觉皮层简单细胞被激活; 或者说在不同的刺激下, 一个特定的视觉皮层简单细胞是很少被激活的。

最早关于稀疏编码的文献是 1996 年, Cornell 大学心理学院的 Bruno 在 Nature 上发表了一篇题名为 “Emergence of simple-cell receptive field properties by learning a sparse code for nature images” 的文章, 大意是讲哺乳动物的初级视觉的简单细胞的感受野具有空域局部性、方向性和带通性(在不同尺度下, 对不同结构具有选择性), 和小波变换的基函数具有一定的相似性。当时描述这些性质主要从自然图像编码的统计结构上来理解这些视觉细胞的特性, 但是大部分都没有成功, 接着 Bruno 在文章中提出通过最大化稀疏编码假说成功描述了上述细胞的性质, 然后稀疏编码就火了。

信号的稀疏表示之所以诱人，可以从以下几方面解释：

(1) 稀疏表示有较强的数据压缩能力，而且似乎是人类视觉系统提取信息的特点。人类视觉系统可以把外周视觉通道获得的约 10^7 的数据量压缩成视觉中枢所感受的约 50b/s。其数据压缩能力远高于人工的数据压缩算法。

(2) 与独立分量分析相比较，稀疏性是更稳健的建模假设。

(3) 从数学上看，系稀疏表示建立在过完备性基础上。而从生物背景上看，过完备性比独立性更自然。“正交”、“独立”这些性质数学上虽是可行的，但在生物系统中未必能严格实现。

(4) 信号处理技术发展中发现，采用过完备系统的自适应表示比采用单一完备基的系统更容易取得稀疏性。

1.2 压缩感知理论

压缩感知理论建立在信号稀疏表示理论的基础上，是一种充分利用信号可压缩性或稀疏性的全新信号获取及处理理论。根据信号采样定理——奈奎斯特定理，只有当采样频率大于信号中最高频率的 2 倍时，采样之后的数字信号才能完整地保留原始信号中的信息，这样势必会造成资源的巨大浪费。一方面，信号需要通过大于信号带宽两倍的高速率进行采样观测，随着信号带宽的不断增加，对传感系统造成了更大的采样速率方面的压力；另一方面，采样得到的样本太多，对我们存储和传输都带来了巨大的不便，因此我们需要对采样得到的样本数据再进行一定的压缩。这样一来，先以高速率采集得到大量样本然后再压缩就造成了很大的浪费。

于是考虑是不是存在一种采样方式可以直接采样得到适量的信息，并且利用这些信息可以足够好地恢复原始信号？

2004 年，Donoho 和 Candes 等人提出了压缩感知理论，它是信号采样时联系模拟信源和数字信息的桥梁，该理论指出：只要信号是可压缩的或在某个变换域是稀疏的，那么就可以利用不相关的观测矩阵直接将这样一

个高位信号投影到低维空间上，然后利用少量的投影解一个优化问题，就可以高概率重构原信号。压缩感知理论是一种新的在采样的同时实现压缩的理论框架。

压缩感知和稀疏表示其实是有些不同的。压缩感知的字典是固定的，在压缩感知的术语里面其字典叫做测量矩阵。但压缩感知的恢复算法和稀疏表示是同一个问题。他们都可以归结为带约束条件的 $L1$ 范数最小化问题。

2 发展

图像的稀疏表示作为一种图像模型，能够用尽可能简洁的方式表示图像，不仅揭示了图像的内在结构与本质属性，同时能够降低噪声与误差，从而有利于后续的图像处理。

2.1 基于正交基的稀疏表示

传统的信号表示方法是基于变换空间展开的，这种基称为“变换基”。如经典的傅里叶变换、DCT 变换、小波变换等，都是将信号进行某个正交基空间的投影，表示为 $y = \Psi x$ 。其中， x 为原始信号， Ψ 为某个正交基空间，上式其实是将空间域的信号 x 变换到了另外一个空间 Ψ ，在这个空间中信号能得到更好的稀疏表达。但是这种建立在正交基上的信号分解算法也存在一定局限性，因为不同的正交基往往具有不同的特性，如傅里叶变换对振荡信号表达的效果比较好，但对点状奇异性的信号的表示并不有效；而二维小波对图像的点状和板状奇异性表示却非常有效；DCT、Brushlet 等变换对图像的纹理特征表达效果更好。不同信号所具有的特点可能相差很大，即便是同一个信号，信号本身的构成也可能错综复杂，单用某一空间正交基对信号进行系数表达，往往不总能得到较好的稀疏表达效果。因此，一些学者开始考虑能否将信号投影到几个不同的组合正交基上，以获得更好的稀疏表达。

2.2 基于组合正交基的稀疏表示

2001 年 Donoho、Candes 等人发表的文章即基于组合正交基的思想，文中通过联合小波和曲波变换对图像进行了重构，也是因为考虑到自然现象中的混合信号用单一的正交变换基往往得不到非常有效的表现。2002 年 Elad、Bruckstein 又基于组合正交基提出了不确定性准则和最稀疏解的唯一性结论。基于组合正交基的稀疏分解算法，在构造原子时需要花费一定的时间。但是实验表明，基于组合正交基的稀疏分解方法在总体上计算速度一般能够得到 5-10 倍的提高。

2.3 基于多尺度几何分析理论的稀疏表示

Curvelet 变换

Bandelet 变换

Contourlet 变换

2.4 基于冗余字典的稀疏表达研究

超完备信号稀疏表示方法肇始于 20 世纪 90 年代，1993 年 Mallat 和 Zhang 首次提出了应用超完备冗余字典对信号进行稀疏分解的思想，并引入了匹配追踪算法。这是一种全新的信号表示理论：用超完备的冗余函数库取代基函数，称之为冗余字典，字典中的元素被称为原子。字典的选择应尽可能好地符合被逼近信号的结构，其构成可以没有任何限制。从冗余字典中找到具有最佳线性组合的 k 项原子来表示一个信号，称作信号的稀疏逼近或高度非线性逼近。

当前信号在冗余字典下的稀疏表示研究主要集中在两个方面：

- 1) 如何构造适合信号特点的冗余字典，以及如何提高冗余字典的普适性；
- 2) 用怎样的方法快速设计构造字典，并求得信号的最稀疏表示。

2.4.1 字典设计方法

1. 从已知的变换基中选取，比如 DCT、小波基等，这种方法很通用，但是不能自适应于信号。

2. 学习字典，即通过训练和学习大量的与目标数据相似的数据来获得。

字典的学习目前主要有两类：一类是由一组参数和一套选取的含有参数的若干函数构成，用其近似表达信号。这类字典不需要存储整个字典，只需要存储相关参数信息即可，因此大大降低了存储量，但由于涉及的字典与原始信号无关，所以不具有适应性；另一类字典学习方法则是根据信号或图像特点进行训练学习得到的自适应字典。目前用得较多的方法有基于 K-均值聚类的思想的 K-SVD 算法、MOD 字典学习算法等。

2.4.2 字典学习方法

1. MOD

流程：读入图像（测试图像）→ load 预选的字典（一般是训练样本集）
→ 使用 OMP 计算稀疏表示系数

2. K-SVD

流程：读入图像 → load 预选的字典（初始值）→ 使用 K-SVD 生成新的字典与新的表示系数 → 利用新的字典进行下一步研究。

3 应用

3.1 图像压缩

信号的稀疏表示并不是新的东西。我们很早就一直在利用这一特性。例如，最简单的 JPEG 图像压缩算法。原始的图像信号经过 DCT 变换之后，只有极少数元素的绝对值比较大的，而大部分元素都等于零或者说接近于零。这些绝对值较大的重要系数保持了信号的绝大部分信息。在有损压缩

编码中，通过保存重要系数及其位置信息就可以实现信号压缩。

3.2 稀疏表示在人脸识别中的应用

任何模型都有建模的假设条件。压缩感知正是利用信号的稀疏性这个假设。对于我们处理的信号，时域上本身就具有稀疏性的信号是很少的。但是，我们总能找到某种变换，使得在某个变换域之后信号具有稀疏性。这种变换是很多的，最常见的就是 DCT 变换、小波变换、gabor 变换等。

然而，这种正交变换是传统视频图像处理采用的方法。目前所采用的一般不是正交变换，而是基于样本采样的，或者说是通过大量图像数据学习得到的，其结果称作字典，字典中的每一个元素称作原子。相关的学习算法称作字典学习。学习的目标函数是找到所有样本在这些原子的线性组合表示下是稀疏的，即同时估计字典和稀疏表示的稀疏这两个目标。

人脸的稀疏表示是基于光照模型。即一张人脸图像，可以用数据库中同一个人所有的人脸图像的线性组合表示。而对于数据库中其它人的脸，其线性组合的系数理论上为零。由于数据库中一般有很多个不同的人脸的多张图像，如果把数据库中所有的图像的线性组合来表示这张给定的测试人脸，其系数向量是稀疏的。因为除了这张和同一个人的人脸的图像组合系数不为零外，其它的系数都为零。

上述模型导出了基于稀疏表示的另外一个很强的假设条件：所有的人脸图像必须是事先严格对齐的。否则，稀疏性很难满足。换言之，对于表情变化，姿态角度变化的人脸都不满足稀疏性这个假设。所以，经典的稀疏脸方法很难用于真实的应用场景。

稀疏脸很强的地方在于对噪声相当鲁棒，相关文献表明，即使人脸图像被 80% 的随机噪声干扰，仍然能够得到很高的识别率。稀疏脸另外一个很强的地方在于对于部分遮挡的情况，例如戴围巾，戴眼镜等，仍然能够保持较高的识别性能。上述两点，是其它任何传统的人脸识别方法所不具有的。

一谈到识别问题，大家都会想到要用机器学习的方法。先进行训练，把训练的结果以模板的形式存储到数据库上；真实应用环境的时候，把测试样本经过特征提取之后，和数据库中的模板进行比对，查询得到一个最相似的类别作为识别结果。往往，机器训练的时间都超级长，几天，几个礼拜乃至几个月，那是常见的事情；识别的时间一般是很小的。典型的例如人脸检测问题。这是可以接受的，因为训练一般都是离线的。

然而，基于稀疏分解的人脸识别是不需要训练的，或者说训练及其简单。基于稀疏表示的人脸识别，其稀疏表示用的字典直接由训练所用的全部图像构成，而不需要经过字典学习（也有一些改进算法，针对字典进行学习的）。当然，一般是经过简单的特征提取。由于稀疏表示的方法对使用什么特征并不敏感。故而，其训练过程只需要把原始图像数据经过简单的处理之后排列成一个很大的三维矩阵存储到数据库里面就可以了。

关键的问题在于，当实际环境中来了一张人脸图像之后，去求解这张人脸图像在数据库所有图像上的稀疏表示，这个求解算法，一般比较耗时。尽管有很多的方法被提出，但是对于实时应用问题，依然没法满足。所以，问题的关键还是归结于 $L1$ 范数最小化问题上来。

$L1$ 范数最小化问题已经有很多种快速求解方法，这里主要包括有梯度投影 Gradient Projection, 同伦算法，迭代阈值收缩，领域梯度 Proximal Gradient, 增广拉格朗日方法，这几种方法都比正交匹配追踪算法 OMP 要高效的多。上述几种快速算法中，采用增广拉格朗日的对偶实现相比其它的快速算法要更好。最近流行的 Split Bregman 算法也是不错的选择。