求解反问题的一个新方法: 模型求解与范例学习结合

献给李大潜教授 80 华诞

徐宗本*, 杨燕, 孙剑

西安交通大学数学与统计学院, 西安 710049

E-mail: zbxu@mail.xjtu.edu.cn, yangyan92@stu.xjtu.edu.cn, jiansun@mail.xjtu.edu.cn

收稿日期: 2016-10-27; 接受日期: 2017-03-06; 网络出版日期: 2017-05-03; * 通信作者

国家自然科学基金 (批准号: 91330204) 资助项目

摘要 反问题存在于科学、工程与技术等广泛领域, 是数学问题的基本形式之一. 传统的反问题求解基于求解与问题相关的数学模型 (如一个偏微分方程组和变分问题等), 无论是建模还是模型求解都回避不了大量近似, 从而很难获得满意的准确解, 被公认是困难问题. 然而, 我们注意到, 对很多反问题 (如图像处理) 而言, 除模型之外, 还往往存在大量的范例. 本文提出将模型求解与范例学习相结合的反问题求解新思路. 以压缩感知核磁共振成像 (CS-MRI) 为例, 说明如何将压缩感知模型的模型求解与基于范例学习的深度学习相融合, 以形成求解 CS-MRI 问题的一个全新方法. 应用表明, 基于模型与基于数据相结合的这一新方法不仅可行、高效, 而且具有很强的可推广性.

关键词 反问题 范例学习 深度学习

MSC (2010) 主题分类 65M32, 15A29, 65K10, 65T99, 68Q32

1 引言

反问题是基本的数学问题之一, 它特指那些“从外部观测推知其内部结构”的数学物理问题. 最典型的, 例如, 资源勘探: 通过向地层发射人为设计信号 (如地震波和电磁波等)、收集回波和分析回波来推知地层内部的地球物理信息; 医学成像: 通过向人体被探查部位发射不同物理场信号, 实施计算机断层扫描, 重建获得人体骨骼、组织和代谢等体征图像; 天气预报: 通过所收集的各地、各类气象观测数据, 计算分析来对未来时段天气作出预判; 图像处理: 对摄影、摄像和视觉等感知器所获得到的图像进行加工、处理, 以获得更为真实和理想的图像等 (参见文献 [1]). 反问题的一般形式可描述为, 给定一个满足 $f(P, M) = O$ 的物理系统 P , 其中 f 为一个线性或非线性映照, M 为模型参数, 假定已知有关系统 P 的一些定量观测 O , 期望从这些观测中反推引起这些观测的物理系统 P . 最常见 (也是相对容易处理) 的一类反问题是线性反问题. 线性反问题可以用一个抽象的算子方程 $y = Ax + \varepsilon$ ($x \in \mathcal{X}$, $y \in \mathcal{Y}$) 来描述, 其中 $\mathcal{X}$ 和 $\mathcal{Y}$ 为度量空间, A 为一个线性算子, y 是观测值, ε 是观测误差, 我们需要根

英文引用格式: Xu Z B, Yang Y, Sun J. A new approach to solve inverse problems: Combination of model-based solving and example-based learning (in Chinese). Sci Sin Math, 2017, 47: 1345–1354, doi: 10.1360/N012016-00196

据已知的 A 和 y 来反推 x . 线性反问题的典型例子包括医学成像中的 CT (computed tomography) 和 MRI (magnetic resonance imaging), 雷达成像中的 SAR (synthetic aperture radar), 及图像处理问题等.

反问题的求解传统上是以模型为基础的. 首先必须运用领域知识对观测过程 $f(P, M)$ 或算子 A 进行细致建模; 其次需要对确定模型下的观测 $f(P, M) = O$ 发展一个精确高效的模拟 (正演) 方法; 最后结合各种先验 (如解应有的性态和观测误差先验) 选择合适的变分形式, 以从确定模型中反演问题的解. 这样的基于模型的反问题求解方法充满着困难性和不确定性. 首先, 建模做到精确基本上不可能, 即使在各种各样的近似及假设条件下, 近似建模成为可能但通常也有相当大的困难性; 其次, 发展高精度的正演模拟 (通常表现为一个偏微分方程数值解) 难免引进新的近似; 最后, 由于观测误差 ε 的不确定性和解的先验刻画的困难性, 选择合适的变分形式, 以确保所导致的问题解存在、唯一也相当不易且具多样性. 由于本文限于讨论反问题的反演方法, 我们把这最后一个不确定性解释得更清楚些. 以下以线性反问题为例. 给定一个线性观测系统 $y = Ax + \varepsilon$, 并假定观测算子 A 是已经精确建模的. 由于观测系统和观测过程不可避免地出现误差 ε , 选择什么样的度量 l 以从 $l(y, Ax)$ 中反演 x 成为反问题求解的首要问题. 通常最为普遍的选择是取 $l(y, Ax) = \|Ax - y\|_2^2$, 即通过最小二乘反演, 但很显然, 这种选择当且仅当误差 ε 服从正态分布时为最佳. 容易证明 (参见文献 [2]), 当 ε 服从 p 阶 Laplace 分布时, 取 $l(y, Ax) = \|Ax - y\|_p^p$ 最好; 而当 ε 服从混合 Gauss 分布时, 加权 l_2 度量 $l(y, Ax) = \sum_{l=1}^L \lambda_l \|Ax - y\|_2^2$ 最佳, 这里 L 是混合 Gauss 分布的混合成份数. 所以, 在无从知晓观测误差的情况下 (注意, 任意分布可以用 Gauss 混合分布逼近), 最优的度量选择应是加权 l_2 度量, 但这里加权向量 λ_l 未知, 其成份个数 L 也未知. 另一方面, 即使正确选择了反演度量 l , 变分问题 $\min_x l(y, Ax)$ 也未必是良定的, 我们还必须根据解的先验信息来限定上述变分问题解的搜索空间. 正则化方法是实现这种限制的基本框架, 它用一簇与原问题相“邻近”的良定问题的解去逼近原问题的真解, 一般形式为

$$\min_x \{l(y, Ax) + \lambda P(x)\}, \quad (1.1)$$

其中 λ 是正则化参数, $P(x)$ 为正则子 (一个非负的实函数), 反映解的先验和对所求解的期望. 例如, $P(x) = \|x\|_2^2$ 建模解的光滑性先验, 相应方法即熟知的 Tykhonov 正则化方法, 它所产生的反问题解具有光滑性. 稀疏性是近年来备受关注的另一类重要先验 (参见文献 [3]), 它描述了解 x 具有大量近似零分量的特征 (换言之, 它的非零分量数 $\|x\|_0 < n$). 由于零和非零能够自然描述“感兴趣”和“不感兴趣”的分量, 稀疏的 x 描述了现实中的一大类对象, 如稀疏场景/图像 (对于一个应用问题, 如目标识别, 一个场景中令人感兴趣的像素只是很少的一部分). 因此, 如果已知所感知的对象 x 是稀疏的, 则观测量就可能大大减少, 这即是近年来所发展起来的压缩感知 (compressed sensing, CS) 方法 (参见文献 [4]). CS 方法的经典重构过程为 $\min_x \{l(y, Ax) + \lambda \|x\|_0\}$, 它已被广泛应用于雷达成像、医学成像和图像处理等领域, 其中 l_0 度量 $\|x\|_0$ 在计算中常用 l_q 度量 $\|x\|_q$ ($q < 0 \leq 1$), 或更一般地, 一个能够诱导稀疏性的非凸正则化函数 $g(x)$ 来近似. 应用中, x 常常未必是自然稀疏的, 但却可能存在某个变换 (线性或非线性) D , 或对 x 分量的某种分组 $x = (x_1, x_2, \dots, x_L)$, 使其变换或分组后的向量具有稀疏性 (参见文献 [5, 6]). 于是, 一个一般的稀疏先验正则子可表示为 $P(x) = \sum_{l=1}^L \lambda_l g(D_l x)$, 其中 D_l 为一个变换, $g(\cdot)$ 为正则化函数, λ_l 为权系数. 注意到, 当 $L = n$, $x = x_i$, $g(x_i) = |x_i|$, $P(x)$ 正是加权 l_2 度量, 这一正则子也自然包含平滑正则子. 所以, 一个一般的线性反问题正则化方法可描述为如下变分问题:

$$\min_x \left\{ l(y, Ax) + \sum_{l=1}^L \lambda_l g(D_l x) \right\}. \quad (1.2)$$

值得注意的是, 由上述 (1.2) 所描述的模型其实是一个包含各种参数 (如 $g(\cdot)$ 、 D_l 、 λ_l 和 L 等) 的模型

族. 当我们基于模型求解线性反问题时, 其实总是在选定一组参数下求解该模型 (此过程常称为模型选择). 这说明求解反问题即使在选择变分形式时 also 具有很强的多样性, 从而反演结果也具有相当的不确定性.

有意思的是, 上述讨论说明, 基于模型求解反问题常常能够自然地产生参数化的模型族 (像 (1.2)), 而我们传统的求解方法却只是从中选择使用了之一模型来求解, 所获解的质量因而严重依赖于模型选择策略. 这应是传统反问题求解性能的关键要素之一. 一个自然的问题是, 我们能否同时求解模型族而不是其中之一模型呢? 如果能, 这样既能减轻传统反问题求解模型选择的困难 (甚至减弱对建模精确性的过度依赖), 也能提升基于模型求解反问题的性能. 本文目的是说明这样求解模型族的想法其实是可能的. 之所以可能, 是由于在模型之外, 存在着大量的范例 (例如, 给定一个线性观测算子之后, (Ax, x) 对任意 x 就构成无穷无尽的范例, 更不用说, 图像处理中存在大量现实的“被污染”和“未被污染”的图像对), 这些范例构成了可以用机器学习原理学习模型族中参数的样本. 我们将说明, 基于这些范例样本, 将传统的模型求解方法与近年来蓬勃发展起来的深度学习 (deep learning) 技术结合 (参见文献 [7]), 能实现求解模型族的功能. 我们将在第 2 节给出这样的一般结合框架, 而在第 3 节, 以 CS-MRI 为例, 说明这种结合方法的可行性和有效性. 第 4 节给出本文的结论.

2 模型求解与范例学习结合的反问题求解框架

对于一个反问题 (以线性反问题为例), 假定我们已经有了一个模型求解方法, 例如, (1) 观测算子 A 和观测值 y 已知, 并且它定义了一个可供选择的反演模型族 $\mathcal{F} = \{\hat{f} \mid X = \hat{f}_\theta(Y)\}$, 其中 $\hat{\theta}$ 为模型族参数集 (例如, 由 (1.2) 定义, 模型族相当于机器学习中的假设空间); (2) 反演方法是一个迭代法 (参见文献 [8–10]), 例如,

$$x^{k+1} = f(x^k, y, \theta), \quad (2.1)$$

其中 θ 为迭代法中的参数集; (3) 除此之外, 假定我们也有一组范例 $\mathcal{D} = (y_1, x_1), (y_2, x_2), \dots, (y_N, x_N)$, 范例服从未知分布 $P(\mathcal{Y}, \mathcal{X})$ 且是 iid (独立同分布) 的. 我们提出将模型求解与范例学习相结合的反问题求解框架如下:

第 1 步 将反演方法 (2.1) 展开成 T 层深度神经网络 DeepN, 将范例集 $\mathcal{D}$ 随机分拆为训练集 $\mathcal{D}_{\text{train}}$ 和测试集 $\mathcal{D}_{\text{test}}$ 两部分;

第 2 步 基于 $\mathcal{D}_{\text{train}}$ 运用经验风险极小化方法训练 DeepN, 基于 $\mathcal{D}_{\text{test}}$ 测试 DeepN;

第 3 步 输入观测 y , 取 DeepN(y) 为反问题的最终解.

在以上框架中, DeepN 由模型求解方法完全唯一决定. 更具体地, 模型族决定了学习的假设空间, 反演方法决定了 DeepN 的拓扑结构 (深度、每层的神经元数及每层的非线性传输机制). 形式上, DeepN 是一个从观测空间 $\mathcal{Y}$ 到特征空间 $\mathcal{X}$ 的一个映照, 对任何 $y \in \mathcal{Y}$, DeepN(y) $\in \mathcal{X}$ 可显式表示为

$$\text{DeepN}(y) = f_T(\cdots f_i(\cdots f_1(x^0, y, \theta_1) \cdots, y, \theta_i) \cdots, y, \theta_T), \quad (2.2)$$

其中 θ_i 是 DeepN 第 i 层的连接权 (可供学习的参数), 总层数 T (深度) 由对近似解的精度要求决定 (当反演方法 (2.1) 已知有理论上的收敛速度估计时, 这样的 T 显然很容易事先确定, 否则需要用交叉验证等启发式方法确定). 一个典型的 DeepN 如图 1 所示.

图 1 典型 DeepN 示意图

上述框架中的第 2 步是运用所给定的范例 $\mathcal{D}$ 来训练 DeepN, 换言之, 确定各层的最优连接权值 θ_i . 这一过程在机器学习中称作学习. 学习的目的不仅在于调整权值以拟合所给定的范例, 更重要的是学习范例集所蕴含的分布. 所以, 这一过程的效果是通过学习结束后 DeepN 的泛化性来检验的, 实践上则用 DeepN 在测试集 $\mathcal{D}_{\text{test}}$ 上的表现来度量. 根据统计学习理论 (参见文献 [11]), DeepN 的学习过程可通过它在训练集 $\mathcal{D}_{\text{train}}$ 上的经验风险来督导, 这里经验风险定义为

$$R(\theta) = \frac{1}{n} \sum_{i=1}^n l(x_i, \text{DeepN}(y_i), \theta), \quad (2.3)$$

其中 n 为训练集 $\mathcal{D}_{\text{train}}$ 所含范例个数, $\theta = (\theta_1, \theta_2, \dots, \theta_T)$ 为 DeepN 所有层的权值, l 是损失度量, 通常取为最小二乘度量 $l(x, y) = \|x - y\|_2^2$. 于是, 学习 DeepN 即求解下述优化问题:

$$\theta^* = \arg \min_{\theta} R(\theta). \quad (2.4)$$

求解这个学习问题有各种方法, 最基本的是应用梯度下降类方法, 这其中最困难的是如何计算 $R(\theta)$ 对 $\theta = (\theta_1, \theta_2, \dots, \theta_T)$ 的梯度. 仿照神经网络中的 BP (back-propagation) 格式, 该梯度可反向逐层计算如下:

$$\frac{\partial R}{\partial \theta} = \left\{ \frac{\partial R}{\partial \theta_k} \right\}_{k=1}^T, \quad \frac{\partial R}{\partial \theta_k} = \frac{\partial R}{\partial f_T} \frac{\partial f_T}{\partial f_{T-1}} \cdots \frac{\partial f_k}{\partial \theta_k}. \quad (2.5)$$

以上所给出的结合模型求解与范例学习的反问题求解方法有如下明显优点: (1) 与纯粹的模型求解方法相比, 由于允许在模型族中基于范例寻优, 不仅可容忍建模与反演中的更大不确定性 (特别是不精确性), 而且也能带来更高质量的反演结果; (2) 与纯粹的范例学习 (如深度学习) 相比, 由于其拓扑结构由反问题的模型和基于模型的反演方法唯一确定, 回避了机器学习方法中选择网络拓扑的公认难题. 这个新方法本质上给合了基于模型与基于数据解反问题的两种不同思路, 既充分利用了基于模型的精确宏观指导, 又充分利用了深度学习的强大建模能力.

3 应用实例: 基于压缩感知的 MRI 成像

本节将上述解反问题的新框架用于基于压缩感知的 MRI 成像, 以展示其可行性和有效性 (具体细节参见文献 [12]).

3.1 CS-MRI 和 ADMM 算法

MRI (magnetic resonance imaging) 是核磁共振成像的简称, 是一种对软组织病灶十分有效的非电离性临床检测手段. MRI 应用十分广泛, 但公认的挑战问题是如何让成像快速以减轻受检者的不舒适感和减少伪影的产生. 基于压缩感知 (compressed sensing) 的 MRI 重构 (CS-MRI) 是解决这一问题的

有效途径, 它利用压缩感知原理从低于 Nyquist 采样频率的采样数据中重构高质量图像, 是一个典型的线性反问题.

令 $x \in C^N$ 是一个被重构的 MRI 图像, $y \in C^{N'}$ ($N' < N$) 是在 K - 空间 (Fourier 空间) 的欠采样数据, 根据压缩感知理论, CS-MRI 重构模型可以一般地建模为如下优化问题:

$$\hat{x} = \arg \min_x \left\{ \frac{1}{2} \|Ax - y\|_2^2 + \sum_{l=1}^L \lambda_l g(D_l x) \right\}, \quad (3.1)$$

其中 $A = PF$ 是测量矩阵, P 是欠采样矩阵, F 是 Fourier 变换矩阵, D_l 表示一个线性滤波操作对应的变换, $g(\cdot)$ 表示一个正则化函数, λ_l 是一个正则化参数, L 是可能的变量分组参数. 所有参数 (D_l, g, λ_l, L) 都是不确定和可供选择的 (反映建模的不精确性和不确定性), 从而, (3.1) 定义了一个 CS-MRI 模型族. 易观察到, 在这个模型族中, $L = N$, $D_l x = x_l$, $g(\cdot) = |\cdot|_p^p$, 对应正则化子 $\sum_l \lambda_l |x_l|_p^p$, 它刻画混合噪声情形下的重构; $L \leq N$, $D_l x = \mathcal{P}_{x_l} x$, $g(\cdot) = |\cdot|_2^2$, 对应正则化子 $\sum_l \lambda_l |\mathcal{P}_{x_l} x|_2^2$, 它刻画组稀疏先验下的重构; $L \leq N$, D_l 为第 l 个滤波器, $g(\cdot) = |\cdot|_p^p$, 对应基于滤波的稀疏重构 (参见文献 [5, 6, 13]).

模型族 (3.1) 可以运用 ADMM (alternating direction method of multipliers) 方法有效求解 (参见文献 [9]). 更具体地, 通过引入辅助变量 $z = \{z_1, z_2, \dots, z_L\}$, 我们可以重写 (3.1) 为

$$\min_{x, z} \left\{ \frac{1}{2} \|Ax - y\|_2^2 + \sum_{l=1}^L \lambda_l g(z_l) \right\} \quad \text{s.t.} \quad z_l = D_l x, \quad \forall l \in [1, 2, \dots, L]. \quad (3.2)$$

与该问题对应的增广 Lagrange 函数为

$$L_\rho(x, z, \alpha) = \frac{1}{2} \|Ax - y\|_2^2 + \sum_{l=1}^L \lambda_l g(z_l) - \sum_{l=1}^L \langle \alpha_l, z_l - D_l x \rangle + \sum_{l=1}^L \frac{\rho_l}{2} \|z_l - D_l x\|_2^2, \quad (3.3)$$

其中 $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_L\}$ 是 Lagrange 乘子, $\rho = \{\rho_1, \rho_2, \dots, \rho_L\}$ 为惩罚系数. 求解 (3.2) 的 ADMM 算法是交替求解如下的三个子问题的一个迭代过程:

$$\begin{cases} x^{(n+1)} = \arg \min_x \frac{1}{2} \|Ax - y\|_2^2 - \sum_{l=1}^L \langle \alpha_l^{(n)}, z_l^{(n)} - D_l x \rangle + \sum_{l=1}^L \frac{\rho_l}{2} \|z_l^{(n)} - D_l x\|_2^2, \\ z^{(n+1)} = \arg \min_z \sum_{l=1}^L \lambda_l g(z_l) - \sum_{l=1}^L \langle \alpha_l^{(n)}, z_l - D_l x^{(n+1)} \rangle + \sum_{l=1}^L \frac{\rho_l}{2} \|z_l - D_l x^{(n+1)}\|_2^2, \\ \alpha^{(n+1)} = \arg \min_\alpha \sum_{l=1}^L \langle \alpha_l, D_l x^{(n+1)} - z_l^{(n+1)} \rangle, \end{cases} \quad (3.4)$$

其中 n 表示第 n 步迭代. 求解这三个子问题, 令 $\beta_l = \frac{\alpha_l}{\rho_l}$, 并且把 $A = PF$ 代入 (3.4) 得

$$\begin{cases} x^{(n)} = F^T \left[P^T P + \sum_{l=1}^L \rho_l F D_l^T D_l F^T \right]^{-1} \left[P^T y + \sum_{l=1}^L \rho_l F D_l^T (z_l^{(n-1)} - \beta_l^{(n-1)}) \right], \\ z_l^{(n)} = S \left(D_l x^{(n)} + \beta_l^{(n-1)}; \frac{\lambda_l}{\rho_l} \right), \\ \beta_l^{(n)} = \beta_l^{(n-1)} + \eta_l (D_l x^{(n)} - z_l^{(n)}), \end{cases} \quad (3.5)$$

其中 $S(\cdot)$ 为一个与 $g(\cdot)$ 有关的非线性投影函数, η_l 为更新率. (3.5) 即是求解模型族 (3.1) 的 ADMM 迭代过程, 它构成了一个算法族.

注意到, 当使用传统的基于模型方法求解时, 所有的线性变换 D_l 、非线性投影函数 $S(\cdot)$ 和算法参数 λ_l 及 ρ_l 都必须事先给定. 另外, 我们也注意到, 从 (3.5) 可看出, $x^{(n)}$ 可以通过快速 Fourier 变换 (FFT) 来快速计算.

3.2 ADMM-Net: ADMM 与深度学习混合方法

我们按照第 2 节所提出的模型求解与范例学习结合的反问题求解框架来设计一个无需事先确定线性变换 D_l 、非线性投影函数 $S(\cdot)$ 和算法参数 λ_l 及 ρ_l 等而直接求解模型族 (3.1) 的一个新方法.

首先, 将算法族 (3.5) 展开成 (映射到) 一个 T 层深度神经网络 (ADMM-Net). 根据 (3.5), ADMM-Net 的拓扑如图 2 所示, 它由 T 层操作单元 (stage) 串联组成, 每个操作单元包含四个网络层: 重构层、卷积层、非线性变换层和乘子更新层.

重构层 ($\mathbf{X}^{(n)}$): 该层实现公式 (3.5) 中的重构运算, 经过这层操作, 可以重构出一幅 MR (magnetic resonance) 图像. 给定输入 $z_l^{(n-1)}$ 和 $\beta_l^{(n-1)}$, 则该层的输出为

$$x^{(n)} = F^T \left(P^T P + \sum_{l=1}^L \rho_l^{(n)} F H_l^{(n)T} H_l^{(n)} F^T \right)^{-1} \left[P^T y + \sum_{l=1}^L \rho_l^{(n)} F H_l^{(n)T} (z_l^{(n-1)} - \beta_l^{(n-1)}) \right], \quad (3.6)$$

其中 $H_l^{(n)}$ 表示第 l 个线性滤波变换, $\rho_l^{(n)}$ 表示第 l 个惩罚系数. 在第一个操作单元 ($n=1$) 中, 可取 $z_l^{(0)}$ 和 $\beta_l^{(0)}$ 为零.

卷积层 ($\mathbf{C}^{(n)}$): 该层执行卷积操作. 给定图像 $x^{(n)}$, 该层的输出为

$$c_l^{(n)} = D_l^{(n)} x^{(n)}, \quad (3.7)$$

其中 $D_l^{(n)}$ 表示第 l 个可学习的线性滤波变换, 为了进一步拓展网络的性能, 不限制每层的各种参数保持一致.

非线性变换层 ($\mathbf{Z}^{(n)}$): 该网络层来源于 (3.5) 中的非线性投影函数操作. 为了参数化, 假定这里的非线性函数 S 具有分段线性形式, 如 $S = S_{\text{PLF}}$, 而 $S_{\text{PLF}}(\cdot)$ 是由控制点 $\{p_i, q_{l,i}^{(n)}\}_{i=1}^{N_c}$ 所确定的分段线性函数 (如图 3), 这里 $\{p_i\}_{i=1}^{N_c}$ 假定事先选定, 但 $\{q_{l,i}^{(n)}\}_{i=1}^{N_c}$ 通过学习给出. 给定输入 $c_l^{(n)}$ 和 $\beta_l^{(n-1)}$, 该层的输出为

$$z_l^{(n)} = S_{\text{PLF}}(c_l^{(n)} + \beta_l^{(n-1)}; \{p_i, q_{l,i}^{(n)}\}_{i=1}^{N_c}), \quad (3.8)$$

图 2 ADMM-Net 拓扑: 输入 K -空间的欠采样数据, 经过 T 层处理后输出重构的 MRI 图像

图 3 由控制点确定的非线性变换

乘子更新层 ($\mathbf{M}^{(n)}$): 该网络层实现 (3.5) 中的乘子更新操作. 给定输入 $\beta_l^{(n-1)}$ 、 $c_l^{(n)}$ 和 $z_l^{(n)}$, 该层的输出为

$$\beta_l^{(n)} = \beta_l^{(n-1)} + \eta_l^{(n)}(c_l^{(n)} - z_l^{(n)}), \quad (3.9)$$

其中 $\eta_l^{(n)}$ 表示第 l 个可学习的参数.

ADMM-Net 中需要学习的参数包括重构层的 $H_l^{(n)}$ 和 $\rho_l^{(n)}$ 、卷积层的 $D_l^{(n)}$ 、非线性变换层的 $\{q_{l,i}^{(n)}\}_{i=1}^{N_c}$ 和乘子更新层的 $\eta_l^{(n)}$, 其中 $l \in 1, 2, \dots, L$ 和 $n \in 1, 2, \dots, N_s$ 表示滤波器和操作单元的指标 (注意, 层数 T 可由 ADMM 的收敛估计理论事先确定). 我们将基于范例, 用深度学习方法来学习确定 ADMM-Net 中的参数. 对于 CS-MRI 问题, 全采样下的重构图像 x^{gt} 与其相对应的 K -空间中欠采样数据 y 构成一对范例, 所以, 大量已知的 MRI 图像可用作构造范例集

$$D = D_{\text{train}} \cup D_{\text{test}}.$$

训练 ADMM-Net 需要首先对网络参数初始化. 我们建议取初始线性变换 D_l 为 DCT (discrete cosine transform) 变换, 初始化正则子为 l_1 正则子, 于是对应的非线性投影函数取为软阈值函数 (参见文献 [13]), 其他参数随机初始化. 现在, 我们取定一个损失函数 (例如, 取规范化均方误差损失), 对应的经验风险为

$$R(\theta) = \frac{1}{\Gamma} \sum_{(y, x^{gt}) \in D_{\text{train}}} \frac{\|\hat{x}(y, \theta) - x^{gt}\|_2}{\|x^{gt}\|_2}, \quad (3.10)$$

在经验风险极小化的原则下学习确定 ADMM-Net 中的参数, 这里 Γ 是 D_{train} 所含范例个数, $\hat{x}(y, \theta)$ 为网络输出, 它是由输入 y 经由带参数

$$\theta = \{(q_{l,i}^{(n)})_{i=1}^{N_c}, D_l^{(n)}, H_l^{(n)}, \rho_l^{(n)}, \eta_l^{(n)}\}_{n=1}^{N_s} \cup \{H_l^{(N_s+1)}, \rho_l^{(N_s+1)}\}, \quad l = 1, \dots, L$$

的深度网络 ADMM-Net 作用后的结果. 如果 ADMM-Net 是学习后的深度网络, 则对任何 K -空间中欠采样数据 y , ADMM-Net(y) 给出重构 MRI 图像.

3.3 实验验证

我们应用上述方法完成了一系列应用验证. 例如, 我们随机挑选了 150 张脑部 MRI 全采样重构图像, 按不同采样率对其在 K -空间进行下采样, 由此获得了 150 对脑部 MRI 重构范例. 我们用其中的 100 对范例作为训练数据, 而 50 对作为测试数据. K -空间采样模式选择为拟径向采样, 采样率分别取 20%、30%、40% 和 50%. 为了客观, 将我们的方法与各种传统的和最新的 CS-MRI 方法在不同的采样率下进行了比较, 这些传统方法包括 zero-filling、TV [14]、RecPF [15] 和 SIDWT [16], 而最新方法包括 PANO [17] 和 FDLCP [18]. 另外, 比较准则取为在测试集上平均的规范化均方误差 (NMSE) 和峰值信噪比 (PSNR). 我们取 ADMM-Net 的深度 $n = 15$ (ADMM-Net₁₅) 来执行我们的算法, 所有实验在处理器为 i7-4790k (CPU) 的个人计算机上运行.

如表 1 所示, 我们的方法无论在各种准则下, 还是在不同的采样率下, 都达到了最好的重构精度, 而且重构速度显著加速 (提高了重建速度几十倍以上). 图 4 是重构 MRI 图像的可视化比较, 可以看出 ADMM-Net 方法相较于其他方法很好地保持了边界并且没有明显的人工产物 (artifacts). 所有实验验证了新方法的可行性和高效性.

表 1 不同方法不同采样率在脑部数据的比较结果

方法	20%		30%		40%		50%		测试时间 (s)
	NMSE	PSNR	NMSE	PSNR	NMSE	PSNR	NMSE	PSNR	
zero-filling	0.1700	29.96	0.1247	32.59	0.0968	34.76	0.0770	36.73	0.0013
TV	0.0929	35.20	0.0673	37.99	0.0534	40.00	0.0440	41.69	0.7391
RecPF	0.0917	35.32	0.0668	38.06	0.0533	40.03	0.0440	41.71	0.3105
SIDWT	0.0885	35.66	0.0620	38.72	0.0484	40.88	0.0393	42.67	7.8637
PANO	0.0800	36.52	0.0592	39.13	0.0477	41.01	0.0390	42.76	53.4776
FDLCP	0.0759	36.95	0.0592	39.13	0.0500	40.62	0.0428	42.00	52.2220
ADMM-Net ₁₅	0.0739	37.17	0.0544	39.84	0.0447	41.56	0.0379	43.00	0.7911

图 4 脑部图像重构结果示例: 第一行是 20% 采样率的比较, 第二行是 30% 采样率的比较. 从左到右分别比较 ADMM-Net₁₅、RecPF、PANO 和 FDLCP 方法

4 结论

传统的基于模型的反问题求解方法虽然有精确建模难和精确反演 (特别是模型选择) 难等瓶颈困难, 但它有 “能够从物理机制出发界定解的搜索空间、从数学理论出发导出求解过程” 的巨大优势. 另一方面, 基于学习的反问题求解方法虽然不依赖模型 (只依赖数据, 可用一个标准的机器学习程序实现), 但其有效性却本质地依赖网络拓扑和假设空间. 本文提出了将范例学习与模型求解相结合的反问题求解新思路, 其基本出发点在于, 利用机器学习的强大学习能力来解决传统模型方法精确性和模型选择难的困难, 而利用模型方法解决机器学习网络拓扑和假设空间确定的困难. 正如第 2 和第 3 节所述, 新方法通过将模型方法所诱导的模型族取为机器学习的假设空间、将求解模型族的算法族展开成深度学习网络、利用深度学习标准过程从范例中学习来完成模型选择并给出反问题的解. 压缩感知核磁共振成像 (CS-MRI) 的应用表明, 这种新的反问题求解思路不仅可行, 而且高效, 常能带来目前已知最好的结果.

事实上, 本文所提出的将模型求解与范例学习相结合的反问题求解思路具有一般性, 可以推广到有范例存在的各类反问题以及参数识别问题中. 另外, 本文思想的先前探索参见文献 [19–22].

参考文献

- 1 Groetsch C W. Inverse problems in the mathematical sciences. Vieweg, 1993, 52: 799–811
- 2 Xu Z B, Chang X Y, Xu F M, et al. $L_{1/2}$ regularization: A thresholding representation theory and a fast solver. IEEE Trans Neural Netw Learn Syst, 2012, 23: 1013–1027
- 3 Tropp J A, Wright S J. Computational methods for sparse solution of linear inverse problems. Proc IEEE, 2010, 98: 948–958
- 4 Donoho D L. Compressed sensing. IEEE Trans Inform Theory, 2006, 52: 1289–1306
- 5 Dupé F X, Fadili M J, Starck J L. Deconvolution under Poisson noise using exact data fidelity and synthesis or analysis sparsity priors. Statist Methodol, 2012, 9: 4–18
- 6 Huang J, Zhang T. The benefit of group sparsity. Ann Statist, 2010, 38: 1978–2004
- 7 LeCun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521: 436–444
- 8 Yang A Y, Sastry S S, Ganesh A, et al. Fast l_1 -minimization algorithms and an application in robust face recognition: A review. In: Proceeding of the 17th IEEE International Conference on Image Processing (ICIP). Hong Kong: IEEE, 2010, 1849–1852
- 9 Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found Trends Mach Learn, 2011, 3: 1–122
- 10 Blumensath T, Davies M E. Iterative hard thresholding for compressed sensing. Appl Comput Harmonic Anal, 2009, 27: 265–274
- 11 Friedman J. The Elements of Statistical Learning. Berlin: Springer, 2001
- 12 Yang Y, Sun J, Li H B, et al. A deep ADMM-Net for compressive sensing MRI. Adv Neural Inf Process Syst, 2016, 29: 10–18
- 13 Bach F, Jenatton R, Mairal J, et al. Optimization with sparsity-inducing penalties. Found Trends Mach Learn, 2012, 4: 1–106
- 14 Michael L, David D, John M P. Sparse mri: The application of compressed sensing for rapid mr imaging. Magnet Reson Med, 2007, 58: 1182–1195
- 15 Yang J F, Zhang Y, Yin W. A fast alternating direction method for TVL1-L2 signal reconstruction from partial fourier data. IEEE J Selected Topics Signal Process, 2010, 4: 288–297
- 16 Nowak R D. Wavelet-based Rician noise removal for magnetic resonance imaging. IEEE Trans Image Process, 1999, 8: 1408–1419
- 17 Qu X B, Hou Y K, Lam F, et al. Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator. Medical Image Anal, 2014, 18: 843–856
- 18 Zhan Z F, Cai J F, Guo D, et al. Fast multi-classdictionaries learning with geometrical directions in mri reconstruction. IEEE Trans Biomed Eng, 2016, 63: 1850–1861
- 19 Sun J, Tappen M. Learning non-local range Markov random field for image restoration. In: Advances in IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs: IEEE, 2011, 2745–2752
- 20 Sun J, Tappen M. Separable Markov random field and its application in low level vision. IEEE Trans Image Process, 2013, 22: 402–408
- 21 Sun J, Sun J, Xu Z B. Color image denoising via discriminatively learned iterative shrinkage. IEEE Trans Image Process, 2015, 24: 4148–4159
- 22 Yang H R, Sun J, Li H, et al. Deep fusion net for multi-atlas segmentation: Application to cardiac MR images. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016. Lecture Notes in Computer Science, vol. 9901. New York: Springer, 2016, 521–528

A new approach to solve inverse problems: Combination of model-based solving and example-based learning

XU ZongBen, YANG Yan & SUN Jian

Abstract Inverse problem, which is one of the basic forms of mathematical problems, exists in the science, engineering and technology extensively. Traditional inverse problems are resolved through solving mathematical

models related to problems (such as partial differential equations, and variation problems). A great deal of approximations are unavoidably employed during both the modeling and model solving processes, so it is hard to achieve an exact solution. Owing to this point, these inverse problems are universally recognized as hard problems. However, it is noticed that, besides models, a large number of examples exist for many inverse problems. Inspired by this condition, a new approach for solving the inverse problems which combines model solving and example-based learning is proposed in this work. Compressed sensing magnetic resonance imaging (CS-MRI) is taken as an example to illustrate how to incorporate compressed sensing model solving and deep learning based on example-based learning, which forms a totally new method to solve CS-MRI problem. Applications demonstrate that the new method is not only feasible and efficient, but also has high generalization.

Keywords inverse problem, example-based learning, deep learning

MSC(2010) 65M32, 15A29, 65K10, 65T99, 68Q32

doi: 10.1360/N012016-00196