



波现象与智能反演成像研究组

大数据时代如何进行地震数据分析

报告人：王华忠

波现象与智能反演成像研究组 (WPI)

同济大学海洋与地球科学学院，上海

2021年3月5日

目录

- ◆一、概述
- ◆二、地震数据成像处理存在的问题
- ◆三、非学习类算法及在地震数据分析中的应用
- ◆四、智能学习类算法及在地震数据分析中的应用
- ◆五、地震数据分析中值得研究的问题
- ◆六、总结与讨论



◆一、概述

- ◆目前，人类社会的巨大变革是互联网引起的。
- ◆互联网/物联网、智能传感器等带来的大数据生成及采（收）集能力，与数据传输、存储和高性能并行计算技术的结合，进一步把人类带入大数据时代。
- ◆大数据时代的最大特点是智能化。



◆一、概述

- ◆勘探地震学的发展也深受影响。
- ◆自动与智能的检波器，包括可控的震源，以及各种高效采集方式，催生了“两宽一高”地震数据采集，并逐渐普及。
- ◆地震波反演成像，包括数据预处理，进入到Bayes估计框架下。地震波反演成像进入了波动理论阶段。
- ◆受大数据和AI技术发展的影响，智能学习类算法也迅速渗透进地震波成像处理的各个合适的环节。



◆一、概述

- ◆在这样的情况下，大家对智能学习类算法有太多的想象空间，以为它可以替代FWI+LS_RTM得到更好的地震数据分析结果。
- ◆我认为：对这样的看法要保持高度的警惕！否者，可能会误入歧途！



◆一、概述

- ◆油气勘探地震的目标是：精确地刻画油气藏。
- ◆精确地刻画油气藏的基础是对地下介质进行精确的弹性参数（宽带波阻抗）反演。
- ◆估计系统参数，尤其是巨高维的参数场，并非是生物智能擅长的。当前的智能学习类算法都是借鉴亿万年来发展出的生物智能。
 - ◆但是，生物并不解决系统参数估计问题，也没有进化出高级的智能供我们借鉴。



◆一、概述

- ◆生物智能是适应环境变化不断发展和进化出来的，生物感知环境（变化）的方式是通过感官，接收的信息应该是：声波、可见光波、热场、压力场、气味场等等，并基于这些信息做出最优的反应。
- ◆但是，生物智能不能感知系统内部参数场的分布情况。
- ◆生物智能不是万能的，不同的生物有自己独到的智能，发展智能学习算法不能取代FWI+LS_RTM这种系统参数估计方法，我们可以借鉴生物智能已经发展出的学习方式，而不是机械地仿造生物智能。



◆一、概述

◆探索所有类型生物的通用学习方式非常必要。但是，学习与解决问题的能力不是等价的。学习类算法并不一定是大规模系统参数估计问题的合理求解方法。

◆总而言之，我们要借鉴的是生物智能的“神”，而不是模仿生物智能的“貌”。



◆一、概述

- ◆油气勘探地震的目标是：精确地刻画油气藏。核心是地下介质弹性参数场的估计（反演成像）。
- ◆用AI技术进行油气藏的评价与判断，似乎是可行的，但我不认为有很大的必要性。困难程度远超自动驾驶！自动驾驶是必须发展的技术（有巨大的市场需求，投资有回报），但评价油藏定井位就不是必须的技术（没有市场需求支撑，投资没有汇报）。



◆一、概述

- ◆用智能学习算法改进地震数据分析是必要的，但要做的是选择合适的场景。
- ◆是否有大数据的支持也是智能学习算法能否得到好结果的重要因素。
- ◆对智能学习算法在地震数据分析中的应用要保持冷静的心态，进行客观地评价。



◆一、概述

◆对于WPI这样的大学研究组，要做的是把非学习类算法的精髓把握清楚，在地震数据分析中主要还是基于非学习类算法。

◆关于智能学习类算法，首先要理解要借鉴的是哪种生物智能，这类智能的本质是什么，基于这种智能的本质能抽象出的什么样的算法模式，这样的算法模型适合借鉴什么样的问题，对（大）数据的要求是什么，等等。必须要追根求源！做出自己的东西！



◆一、概述

- ◆对已有的人工深度神经网络，要用批判性的眼光看待，它是如何体现出普遍的学习能力的？这种学习能力的本质是什么？能否用数学利用来解释这种学习能力？
- ◆简单地应用别人构建的网络构型和学习（训练）算法在某个实例数据上，尽管也有必要，但这不是科学研究！



◆一、概述

◆我自己的感觉：

- ◆非学习类算法的边界已经逐渐探触到了，就是在非高斯、非线性、非平稳下，感知数据中的结构（相关）信息。然后基于这些信息解决问题。
- ◆参数估计问题稍微复杂，它利用了数据（中的信号或波现象）与参数之间的因果关系。
- ◆还有很多具体问题要解决，但核心工作就是非高斯、非线性、非平稳下的建模。



◆一、概述

◆我自己的感觉：

- ◆智能学习类算法的最高目标应该是尽可能地仿真人脑的学习能力。
 - ◆实际上，不可能也不必要。AI算法一定会在很多方面超越人脑，但是在某些方面永远也达不到人脑的某些智能。
 - ◆即便AI计算机出现，也与人脑这个信息处理机差异极大的，因此不可能指望在所有方面超越人脑。
- ◆智能是多方面的。智能学习算法真正要做的就是抽象出某种智能的本质，设计合适的算法来实现。



◆一、概述

◆我自己的感觉：

- ◆我给智能的定义：行为体依“环境”变化自适应地做出最优决策、最佳判断及最合理响应的能力。
- ◆大数据与Von Neumann计算时代（即便是今后的AI计算机时代），计算机智能（即所谓的人工智能）一定是来自于数据驱动的智能。数据中信息（相关结构、模式）的提取一定是核心问题，建立这些信息与决策目标之间的联系一定是核心问题。
- ◆这又与非学习类算法联系在一起了！



◆一、概述

◆我自己的感觉：

- ◆另一方面， Von Nomann计算时代，算法和计算机结构脱节（对绝大多数人而言是脱节的），导致我们认识不到算法和计算机结构应该是一体的。
- ◆人脑是典型的算法和计算机结构融为一体的实例。因此，算法可以随输入不断调整，计算机也跟着调整。目前认为是脑细胞突触电位水平被修改。
- ◆今后的学习类算法应该也是这样的。所谓AI计算机的发展是必须的。目前主要是神经网络计算机。今后AI计算机会是什么样？



◆一、概述

◆我的预感是：

◆智能的产生建立在数据中信息与决策目标之间的联系上，然后在“决策代价”的控制下产生了所谓的“智能”。

◆把本质问题思考清楚，打通非学习类算法和学习类之间的联系，就进入数据分析的自由王国了！

目录

- ◆一、概述
- ◆二、地震数据成像处理存在的问题
- ◆三、非学习类算法及在地震数据分析中的应用
- ◆四、智能学习类算法及在地震数据分析中的应用
- ◆五、地震数据分析中值得研究的问题
- ◆六、总结与讨论



◆二、地震数据成像处理存在的问题

◆为了发展非学习类算法和学习类算法解决地震数据分析中的问题，有必要把各环节存在的问题罗列一下。

◆地震数据分析包括地震波场正演模拟、数据采集、数据管理、数据预处理、地震波反演成像、成像结果的后处理及图像处理、地震地质解释、油藏描述及含油气性识别。



◆二、地震数据成像处理存在的问题

◆地震波场正演模拟：

◆非水平地表下弹性波场模拟

◆时间空间域

◆频率空间域

◆射线坐标系下非水平地表下弹性波场模拟

◆并行网络模型与弹性波场模拟

◆.....



◆二、地震数据成像处理存在的问题

◆数据采集：

- ◆压缩感知采样理论下的随机观测系统的布设
- ◆频率域可控震源激发与高效采集
- ◆用户定制接收子波形态的地震数据采集
- ◆数据压缩与传输
- ◆.....



◆二、地震数据成像处理存在的问题

◆数据管理：

- ◆PB 级数据的高效I/O
- ◆勘探地震各种类型数据的管理
- ◆.....



◆二、地震数据成像处理存在的问题

◆数据预处理：

- ◆陆上地表相关噪音的压制
- ◆海上地表相关多次波的压制
- ◆基于波场特征变化的地表岩性分区
- ◆特征波场提取
 - ◆初至波提取
 - ◆反射波提取
- ◆特征波场传播方向提取
- ◆震源子波估计/提取



◆二、地震数据成像处理存在的问题

◆数据预处理：

- ◆不同炮检对带限子波幅值、形态不一致性的消除
- ◆剩余静校正量的消除
- ◆连片探区数据的带限子波特征统一化处理
- ◆地震数据规则化
- ◆混叠数据的解混叠（盲源分解问题）
 - ◆连续源激发数据的处理
- ◆.....



◆二、地震数据成像处理存在的问题

◆地震波反演成像---背景速度部分：

- ◆基于折射波和Diving Wave及（类）CMP道集的扫描速度建模
- ◆基于反射波及（类）CMP道集的自动与智能化扫描速度建模
 - ◆Q值、各向异性速度
- ◆透射波波动理论层析反演与建模
 - ◆Q值、各向异性速度
- ◆基于CIG道集的层析速度建模
 - ◆Q值、各向异性速度
- ◆特征反射波波动理论层析反演与建模
 - ◆Q值、各向异性速度
- ◆



◆二、地震数据成像处理存在的问题

◆地震波反演成像---高波数估计部分:

- ◆沿特征波传播方向的散射波成像方法
- ◆数据域最小二乘PSDM
- ◆像域最小二乘PSDM
 - ◆叠加成像后的图像反褶积
 - ◆共角度成像剖面中的图像反褶积
- ◆宽带波阻抗成像
- ◆超级分辨率成像
- ◆.....



◆二、地震数据成像处理存在的问题

◆成像结果的后处理及图像处理：

- ◆CIG道集剩余时差及波形差异的消除

- ◆成像叠加剖面的提高分辨率处理

- ◆.....



◆二、地震数据成像处理存在的问题

◆地震地质解释：

- ◆三维成像叠加剖面的属性提取
- ◆三维成像叠加剖面的层位追踪
- ◆三维成像叠加剖面的断层体系追踪
- ◆三维成像叠加剖面的异常体雕刻
- ◆.....



◆二、地震数据成像处理存在的问题

◆与模型相关的空间数据融合：

- ◆三维成像叠加剖面、速度场、井数据的有机融合

- ◆不同源信息的有机融合

- ◆.....



◆二、地震数据成像处理存在的问题

◆油藏描述及含油气性识别：

◆AI方法的应用

◆.....



◆二、地震数据成像处理存在的问题

◆解决问题的基本思想与方法：

- ◆上述所有问题的解决，都基于各种模型。描述数据的模型是最基础的。
- ◆该模型表达了数据与潜变量之间的关系。
- ◆该模型可以是概率密度函数，也可以是半解析的表达式（数据科学中，它应该是多项式或Fourier级数形式，否则计算机上无法实现）。这些是信号分析与图像分析中的模型。
- ◆描述物理系统的模型一般是偏微分方程。往往可以退化为褶积方程。
- ◆对模型的理解和运用是最基本的。



◆二、地震数据成像处理存在的问题

◆解决问题的基本思想与方法：

- ◆模型中包含的潜变量是数据分析所要求解的！
- ◆它的估计可以统一在Bayes估计理论框架下。

◆在上述抽象框架下，可以解决在信号分析、图像分析和参数估计中遇到的大部分问题。

◆当前的智能学习算法的内核核心，我相信，也是上述问题。

◆总之，非高斯、非线性、非平稳情形下的潜变量的估计是当前数据科学中的真正难题。

目 录

- ◆一、概述
- ◆二、地震数据成像处理存在的问题
- ◆三、非学习类算法及在地震数据分析中的应用
- ◆四、智能学习类算法及在地震数据分析中的应用
- ◆五、地震数据分析中值得研究的问题
- ◆六、总结与讨论



◆三、非学习类算法及在地震数据分析中的应用

- ◆非学习类算法的一个典型特征是：
 - ◆没有泛化应用能力，不同的观测数据，算法要重复执行。
- ◆可以看出，非学习类算法的规律性、统一性是很强的。核心就是求解一个反问题：在Bayes估计理论下，求解物理数学方程中包含的系统参数的估计问题，或求解对信号或图像的最佳建模问题（基函数选择和叠加系数最佳估计问题）。
- ◆目前，正在非线性、非高斯、非平稳假设的方向上不断深化，但发展势头已经不太强劲。
- ◆目前研究和应用的重点，似乎正在向学习类算法转移。
- ◆但是，我认为WPI当前还是要在此方向上深入下去，这是我们当前的立脚点。



◆三、非学习类算法及在地震数据分析中的应用

◆非学习类算法的三种典型反问题模型：

◆图像处理中ROF模型： $J(u) = \frac{1}{2} \|Au - g\|^2 + \alpha TV(u)$ 。对于图像去噪问题， $A=I$ 单位矩阵。

◆信号处理中去噪或反褶积模型： $J(\beta) = \frac{1}{2} \|\Psi\beta - d^{obs}\|^2 + \alpha \|\beta\|_1$

◆物理系统参数估计模型： $J(m) = \frac{1}{2} \|A(m) - d^{obs}\|^2 + \alpha \Omega(m)$ 。 $\Omega(m)$ 代表不同形式的正则化方式。

◆在Bayes估计理论下，把微分方程/积分方程中的（物理）参数估计出来，与把基函数叠加模型中的系数估计出来，算法本质上是没有任何差别的。

◆不过，一般地，我们对（物理系统）参数会有更多的先验认识，能提出更多的、基于物理的正则化约束。

◆而对于基函数叠加系数，基本上仅能施加稀疏性约束。这是信号分析、图像分析与物理系统参数估计很不一样的地方。



◆三、非学习类算法及在地震数据分析中的应用

◆当前，地震数据分析中绝大部分的应用问题都可归结如下

◆正问题模型：

◆ 2D褶积模型： $g(x, y) = \iint f(x', y') h(x - x', y - y') dx' dy' + \varepsilon(x, y)$

◆高维褶积模型的抽象表示： $g(s) = [Hf(r)](s) + \varepsilon(s), r \in R, s \in S$

◆反问题模型： $J(f) = \|g - H(f)\|^2 + \lambda\Omega(f)$

目 录

- ◆一、概述
- ◆二、地震数据成像处理存在的问题
- ◆三、非学习类算法及在地震数据分析中的应用
- ◆四、智能学习类算法及在地震数据分析中的应用
- ◆五、地震数据分析中值得研究的问题
- ◆六、总结与讨论



◆四、智能学习类算法及在地震数据分析中的应用

◆我给智能的定义：行为体依“环境”变化自适应地做出最优决策、最佳判断及最合理响应的能力。

◆智能的产生建立在数据中信息与决策目标之间的联系上，然后在“决策代价”的控制下产生了所谓的“智能”。

◆智能学习类算法的定义：从数据中提取信息，并能据此进行自主决策的一类算法。“决策代价”最小化作为算法构建的依据。



◆四、智能学习类算法及在地震数据分析中的应用

◆地震数据分析中，构建油藏评价及钻井决策系统应该是AI在油气地震勘探的终极目标。

◆但我不认为这是一个好的选择。能否做到是个问题，最关键投入产出比太低，没有人愿意投入资金做这件十分困难的事。

◆我认为它比自动驾驶更困难。

◆我认为AI的算法理论研究是大学研究组应该做，也是可以做的事情。AI的应用不该是大学研究组关心的，不掌握大数据，无法解决应用问题。



◆四、智能学习类算法及在地震数据分析中的应用

- ◆AI和机器学习（ML）是有重要区别的。
- ◆我认为：
 - ◆机器学习（统计学习）是为AI提供算法的，提供的是**Learning From Data**的算法。
 - ◆AI是实现对生物智能的仿生的。
- ◆这是两个不同层次的东西！
- ◆我们应该关注的是ML算法在地震数据分析中的应用，而不是AI在地震数据分析中的应用。



◆四、智能学习类算法及在地震数据分析中的应用

◆深度学习算法（人工深度神经网络算法）是当前AI的核心驱动力量。但它是地震数据分析的核心驱动力量吗？

◆我对此表示怀疑！

◆我的观点：按照行为体依“环境”变化自适应地做出最优决策、最佳判断及最合理响应的能力这个对智能定义，发展相应的智能学习类算法，应用于地震数据分析的适当场景中。这是WPI团队应该做的事情。



◆四、智能学习类算法及在地震数据分析中的应用

◆研发智能学习类算法的思想依据来自于如下观点：

- ◆智能的产生建立在数据中信息（数据中的相关性或结构或模式）与决策目标之间的联系上，然后在“决策代价”的控制下产生了所谓的“智能”。



◆四、智能学习类算法及在地震数据分析中的应用

◆地震数据分析中，智能学习类算法的应用场景：

- ◆1、去噪、插值、Deblending、数据规则化这类问题，可以由智能学习类算法来完成，但必要性不是很大。
- ◆2、拾取初至、速度谱解释与建模这种需要决策的问题，发展智能学习类算法是必要的。
- ◆3、地震剖面上层位、断层、特殊地质体的识别，发展智能学习类算法是必要的。
- ◆4、凡是符合计算机自主决策的场景都可发展学习类算法。

目录

- ◆ 一、概述
- ◆ 二、地震数据成像处理存在的问题
- ◆ 三、非学习类算法及在地震数据分析中的应用
- ◆ 四、智能学习类算法及在地震数据分析中的应用
- ◆ 五、地震数据分析中值得研究的问题
- ◆ 六、总结与讨论



◆五、地震数据分析中值得研究的问题

◆地震反演成像部分：

- ◆1、地震波+电磁波波场特征分析与地表岩性区域分划；
- ◆2、针对初至波的CMP道集速度扫描及初始速度建模；
- ◆3、高维多属性+进化优化算法+Markov最佳演化的初至识别
- ◆4、初至波到达时Beam层析；到达时波动理论层析。
 - ◆各向同性速度估计与建模
 - ◆各向异性速度估计与建模
 - ◆Q值估计与建模
- ◆5、基于CMP道集与类CMP道集（PSTM及PSDM CIG转化来的CMP道集）自动与智能速度建模



◆五、地震数据分析中值得研究的问题

◆地震反演成像部分：

- ◆6、完善CIG道集射线层析速度建模
- ◆7、特征反射波波动理论层析速度估计与建模
 - ◆最后过渡到反射波FWI建模。
- ◆8、最小二乘保真成像方法
 - ◆保真的方位角度反射系数估计
- ◆9、自由表面多次波成像
 - ◆最后发展到层间多次波成像
- ◆10，最终统一到宽带波阻抗成像
 - ◆提供给油藏描述阶段



◆五、地震数据分析中值得研究的问题

◆WPI的研究如何由非学习类的算法逐渐转移到二者并重，最后逐渐发展到以学习类算法为主？

- ◆1、进一步强化非学习类算法，重点是发展非线性、非高斯、非平稳假设下的数据分析算法；
- ◆2、改造非学习类算法，引入合理的决策机制，发展出新的智能学习类算法。
- ◆3、针对模拟人脑的人工神经网络（DL），在网络构型和学习算法上提出自己的新理念和新做法。在地震数据分析和一般数据分析的应用中，获得好的应用效果。



◆五、地震数据分析中值得研究的问题

◆在信号分析、图像分析、空间数据分析中希望研究的问题：

◆1、信号分析，理论上，继续探索复杂情形下的**信号建模**问题。这是个长期方向。

- ◆高维情形下的模式组合，不是一维基函数组合，表达高维信号要引起关注。

- ◆突破二阶统计量是必须的。

◆2、信号分析，应用上，解决复杂**噪音压制、数据压缩、压缩感知随机采样、解混叠/盲源分解、数据规则化、连续源数据的处理、展宽频带、子波估计**等等问题。

- ◆测量波场传播方向；测量波场之间的差异也是很值得探索研究的。

◆3、图像分析、理论上、对**高维图像的建模**是核心问题。PDE、变分法、高维小波变换、Bayes方法的综合应用没有达到信号处理的水平，主要是实践不够。同样地、图像建模是个核心问题！



◆五、地震数据分析中值得研究的问题

◆在信号分析、图像分析、空间数据分析中希望研究的问题：

- ◆4、图像分析，应用上，基本level中，**图像反褶积**展宽频带的问题没有得到很好的解决，PSF的构造，PSF反褶积的问题性等缺乏研究。中等Level中，**图像增强**，尤其是基于统计方法（相关方法）的而不是用微分方法的图像增强，做得不够。**图像属性提取**，做得更是不够，这限制了我们发展更好的层位追踪、断层识别等方法的精度。图像属性的提取是一个十分核心的事情！图像属性是图像处理中很多方法技术的基础信息。高级Level中，图像智能识别，我认为，可以人工神经网络，也可以用PDE、变分法、高维小波变换、Bayes方法综合的做法设计智能算法解决特定的问题。



◆五、地震数据分析中值得研究的问题

◆在信号分析、图像分析、空间数据分析中希望研究的问题：

◆5、空间数据分析，理论上，我希望提出一种我们自己的表达空间数据规律的方法。本质上，这也是对空间数据的建模问题。或散乱数据的逼近问题（函数逼近的特例！）。

◆Kriging方法、径向基函数方法都是经典的方法。

◆对于统计规律相同的随机过程，它的规则采样和散乱采样如何影响对它的建模过程与结果？

◆6、空间数据分析，应用上，首先解决散乱的空间数据的融合插值问题，为各种建模问题提供基础手段。希望从空间散乱数据变化预测的角度提出新的解决散乱数据插值的新方法。希望分析空间变化的带限于波统计特征，消除震源激发和检波值接收引起的振幅和相位畸变。

目录

- ◆一、概述
- ◆二、地震数据成像处理存在的问题
- ◆三、非学习类算法及在地震数据分析中的应用
- ◆四、智能学习类算法及在地震数据分析中的应用
- ◆五、地震数据分析中值得研究的问题
- ◆六、总结与讨论



◆六、总结与讨论

- ◆基于数据产生的计算机智能可以描述为：**智能的产生建立在数据中信息与决策目标之间的联系上，然后在“决策代价”的控制下产生了所谓的“智能”。**
- ◆这种判断实际上描述了当前计算机智能的基本状况，即提取数据中的信息（统计相关信息或高维数据中的结构信息或图像中的结构模式信息），然后建立与决策目标之间的关系，最后在“决策代价”的控制下进行最佳决策。



◆六、总结与讨论

- ◆这样的判断打通了非学习类算法和智能学习类算法之间的内在联系。
- ◆提取数据中的信息，建立与决策目标之间的关系是非学习算法的核心。
- ◆在“决策代价”的控制下进行最佳决策，应该是智能学习算法的核心。这一点也是二者之间的差异所在。
- ◆因此，对数据的理解、对数据的最佳建模是其中共同的根本问题。参数估计问题也是在对数据进行最佳建模，即找到一组最佳参数实现对数据的误差最小的解释！



◆六、总结与讨论

◆基于上述判断，进行地震数据分析的应用研究就不会偏离主题，迷失方向。



谢谢
欢迎批评指正