



波现象与智能反演成像研究组

Bayes决策理论下的地震数据分析


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4月20日

目录

- ◆ 一、概述
 - ◆ 二、Bayes决策理论
 - ◆ 三、地震数据分析中根本问题
 - ◆ 四、Bayes决策理论下的地震数据分析
 - ◆ 五、总结与讨论
- 



◆一、概述

◆为什么Bayes决策理论应该是当前所有智能学习类算法的根本理论基础？

◆首先，要给当前所有智能学习类算法寻找一个合理的理论基础。

◆方便归纳和理解问题！

◆第二，智能学习类算法存在的核心共同点是**自主与智能地决策**。

◆第三、当前的机器学习算法是**数据科学**的当前发展阶段的成果。

数据科学的理论基础就是数理统计学，数理统计学的重要内容之一就是统计推断、或统计决策。**Bayes统计推断或统计决策**是（我认为是）最完美的理论框架。




◆一、概述

◆为什么一定需要一个Bayes决策理论的框架？

- ◆首先，人类的一切活动，都是基于“数据”中的信息并形成认识后，在“最佳”决策基础上才能开展的！数据是随机的，信息是不完整的，决策是需要先验认识的。引入Bayes决策理论框架是必然的，也是顺理成章的。
- ◆第二，决策本质上是个**信息综合利用和寻求决策代价最小**的行为。
- ◆第三，看起来当前的机器学习算法多种多样，分析它们的本质，无非在实现“所谓的最佳决策”。

目录

- ◆一、概述
 - ◆二、Bayes决策理论
 - ◆三、地震数据分析中根本问题
 - ◆四、Bayes决策理论下的地震数据分析
 - ◆五、总结与讨论
- 



◆二、Bayes决策理论

◆Bayes决策理论与Bayes估计理论有什么区别？

- ◆Bayes估计理论的重点是估计“自然”系统参数。
- ◆Bayes决策理论的重点是期望基于“自然”系统参数的决策行为付出的代价最小。
- ◆二者的联系当然是很密切的。
- ◆但是，二者的关注点是不同的。 Bayes估计理论的关注点当然是期望更准确的估计“自然”系统参数； Bayes决策理论的关注点是期望决策代价最小！



◆二、Bayes决策理论

- ◆ (Bayes) 统计决策理论：
 - ◆ 依据样本数据（中蕴含的统计规律），在既定的决策代价函数下，基于最优决策的要求，结合先验信息，产生最佳决策行为和结果。
- ◆ 样本空间；参数空间；决策行为空间是 (Bayes) 统计决策的域。参数的先验信息和决策代价函数是 (Bayes) 统计决策的另外两个问题。



◆二、 Bayes决策理论

◆ (Bayes) 统计决策是：依据数据样本产生的参数，并依据决策代价最小进行决策的。

◆决策函数：设 $X \sim P_\theta$ $\theta \in \Theta$ 是随机变量（数据X）满足的统计模型（概率密度函数，是参数 θ 的函数）。取值于决策（行动）空间 \mathcal{A} 的函数 $\delta(X)$ 称为决策函数。

◆决策风险函数：设 $X \sim P_\theta$ $\theta \in \Theta$ 是随机变量（数据X）满足的统计模型， $L(\theta, a)$ 是损失函数。称决策函数 $\delta(X)$ 的平均损失 $R(\theta, \delta) = E_\theta L(\theta, \delta(X))$ 为的风险函数。

◆ (Bayes) 统计决策就是风险函数最小的决策。



◆二、 Bayes决策理论


- ◆笼统地看，所有的机器学习算法，我认为都仅仅是Bayes统计决策的智能化实现而已。
- ◆在采样数据 X 的统计信息不足（原因很复杂，譬如采样不充分、噪声太大等），且决策参数与采样数据 X 的关系过于非线性时，无论构造什么样的网络构型和算法，也不可能训练它并做出正确的决策。无论什么样的机器学习算法也很难给出正确的决策结果。
- ◆样本数据 X 与决策目标之间的强非线性性关系依然是做出正确/合理决策的最大障碍。做出正确/合理决策的根本基础：数据样本 X （或其特征）与决策目标 Y 之间存在较为线性的关系（或二者之间存在连续的关系）！



◆二、 Bayes决策理论

- ◆人类擅长的是（主要是）基于视觉模式进行抽象并形成概念认识。
- ◆人类很不擅长基于大数据中蕴含的信息进行正确/合理的决策，这正是发展AI、ML、DL的根本动因。ML算法借助高性能计算机可以从大数据中提炼出信息在Bayes决策理论指导下进行决策。这体现出的是人造的智能，它与人脑的智能是由本质区别的。
- ◆人脑的智能是自然（进化）智能的最高代表。借鉴人脑智能，强化人工智能是必然的发展路线。

目录

- ◆一、概述
 - ◆二、Bayes决策理论
 - ◆三、地震数据分析中根本问题
 - ◆四、Bayes决策理论下的地震数据分析
 - ◆五、总结与讨论
- 



◆三、地震数据分析中根本问题

◆勘探地震学发展到现在，其中的核心问题逐渐聚焦，就是聚焦到：

◆把三维油气探区视为一个弹性介质系统，人工震源作为激励该系统的输入，地表观测的波场作为该系统的输出（理论上讲，应该在三维油气探区的所有边界上布设地震波震源和接收器，才能采集到系统的完整激励和输出），该实际物理系统被认为可以由人为建立的数学物理方程来描述，然后在上述正过程描述的基础上，基于Bayes估计理论对实际物理系统的弹性参数（主要是纵波速度、横波速度和密度，实际上更多的是角度反射系数和纵波阻抗）进行估计。

◆在弹性参数估计（通常称地震波反演成像）结果的基础上，结合岩石物理关系，进行岩石物性参数的估计，进而对储层含油气性进行评价，最后进入油气开发阶段。

◆这是当今油气地震勘探的基本逻辑！



◆三、地震数据分析中根本问题

- ◆ 无论是否进入大数据和人工智能阶段，油气地震勘探的核心目标依然是**精确地刻画和评价油气藏**。
- ◆ “**两宽一高**” **数据采集技术** + **广义的地震波反演成像技术**（推进到**宽带波阻抗成像**） + **多种信息融合的油藏描述与评价技术**，始终是三个核心问题。
- ◆ 但是，发展替代人的智能的方法技术，替代人的部分智能劳动，提高工作效率、改善工作效果，是人类发展的必然阶段（即当前所谓的大数据和人工智能阶段！）。
- ◆ 决策是智能劳动的核心任务！发展能自主决策的算法就是十分必要的！
- ◆ 用智能学习算法在**所有合适的应用场景**改进当前地震勘探技术也是十分必要的。



◆三、地震数据分析中根本问题

◆勘探地震全过程涉及到的智能算法应用场景分析：

◆数据采集阶段

- ◆无人接收设备---智能的海上和陆上Node检波器系统
- ◆智能的可控震源系统---客户定制反射子波的可控震源系统
- ◆智能的高效采集方案实施控制系统
- ◆.....

◆数据预处理阶段

- ◆智能的去噪、规则化、Deblending等方法
- ◆智能的地表一致性处理方法
- ◆智能的保真预处理质量控制系统
- ◆.....



◆三、地震数据分析中根本问题

◆勘探地震全过程涉及到的智能算法应用场景分析：

◆自动与智能成像阶段

- ◆（高光谱）电磁波场与地震波场特征变化与地表岩性变化关系的智能辨识及应用
- ◆智能的（类）CMP道集速度谱解释与建模方法技术
- ◆自动与智能的透射波层析反演方法技术
 - ◆智能算法+初至波辨识
- ◆自动与智能的波动理论特征反射波旅行时层析反演方法技术
 - ◆智能算法+特征反射波辨识
 - ◆智能算法+地质体形态识别
- ◆自动与智能的波动理论特征反射波波形层析反演方法技术
 - ◆智能算法+特征反射波辨识
 - ◆智能算法+地质体形态识别
- ◆盲反演与超级分辨率反演的智能算法实现
 - ◆PSF的智能估算
- ◆.....



◆三、地震数据分析中根本问题

◆勘探地震全过程涉及到的智能算法应用场景分析：

◆地震地质解释阶段

- ◆层位与断层自动追踪与解释技术

- ◆五维地震数据中，随方位和张角变化的AVAZ、VVAZ、FVAZ信息的自动与智能提取方法技术

- ◆.....

◆油藏描述阶段

- ◆井震数据的自动与智能匹配

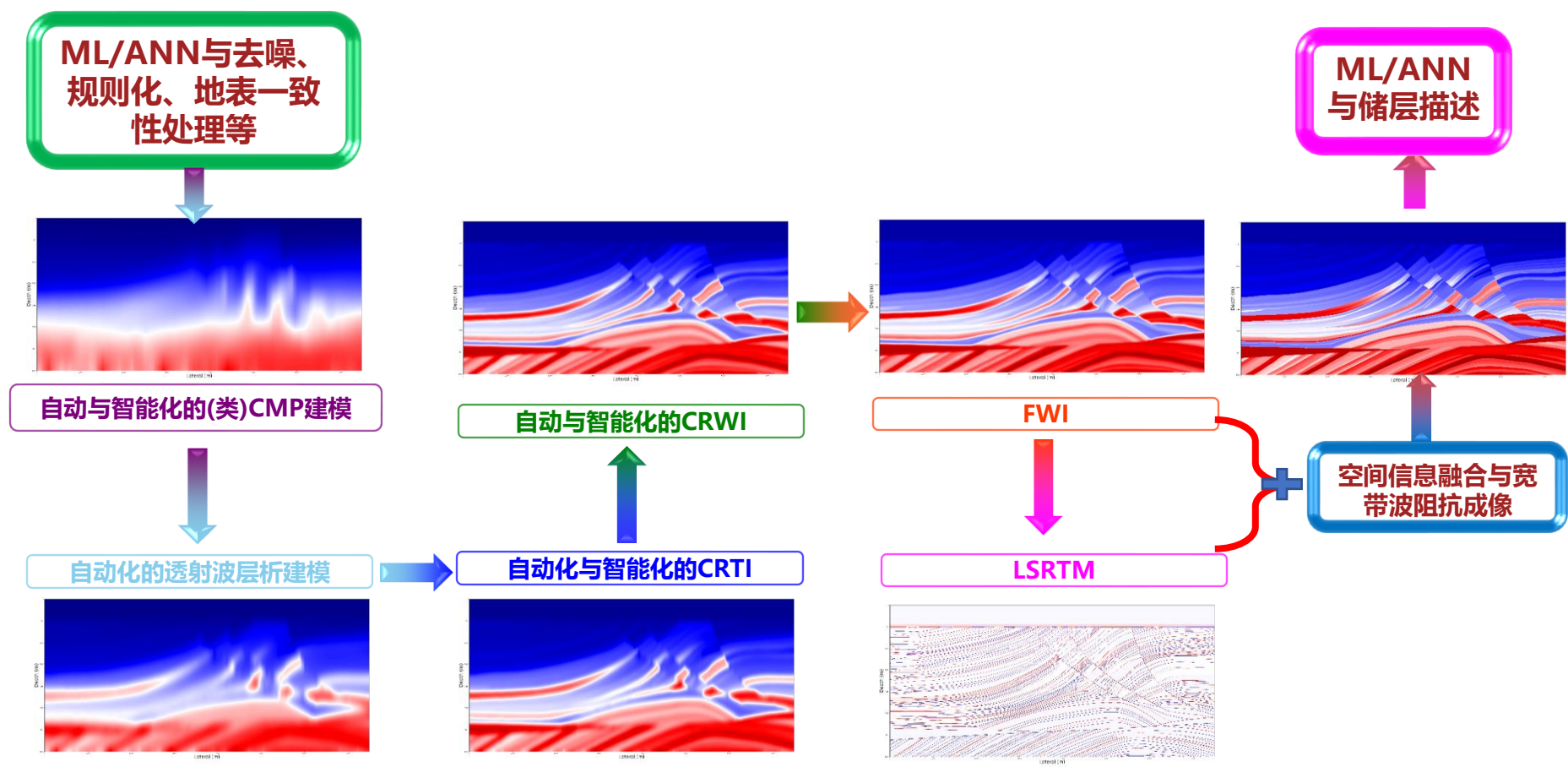
- ◆成像剖面或成像道集属性信息及含油气性辨识

- ◆.....




◆三、地震数据分析中根本问题

◆地震波成像处理流程及关键技术环节的自动与智能化实现



目 录

- ◆一、概述
 - ◆二、Bayes决策理论
 - ◆三、地震数据分析中根本问题
 - ◆四、Bayes决策理论下的地震数据分析
 - ◆五、总结与讨论
- 



◆四、Bayes决策理论下的地震数据分析

◆聚类与分类，并给出类别的具体含义，是AI/ML的核心目标。或判决/决策是AI/ML的核心目标。

◆地震波成像的目标是实现由数据到系统参数的映射—参数估计。这当然也可以归结为一个判决问题。

$$d^{obs} \Rightarrow m$$

$$Lm = d^{obs}$$

$$\hat{m} = L^T L^{-1} L^T d^{obs}$$

信号分析、图像处理和参数估计都可以抽象为这样的问题！
Bayes框架下解释上述问题产生更为本质的认识！
也可以从泛函分析和函数逼近的角度解释这样的问题。
数据与参数之间的关系居于核心地位！

◆这类问题我们已经理解得比较清楚了。尤其是反问题已经很清楚了。

◆宏观的物理规律，基本上也比较清楚了。但微观的、细节的、复杂的物理过程我们并不很清楚。好在很多工程技术问题并不关心微观的、细节的、复杂的物理过程。



◆四、Bayes决策理论下的地震数据分析

◆机器学习算法大致分为：

◆有监督的学习算法；无监督的学习算法；半监督学习算法。

◆有监督学习的基本逻辑：基于（训练和测试）数据包含的统计信息（结构性信息），建立输入和输出（导师）之间的映射关系。然后，泛化应用这种映射关系，实现某种判决（或称决策）！

◆根本的理论问题并没有解决！不完全数据、含噪数据、有误差的导师、数据和导师之间的关系非常复杂等等情形下的映射关系，为什么能泛化应用？



◆四、Bayes决策理论下的地震数据分析

◆机器学习算法大致分为：

◆有监督的学习算法；无监督的学习算法；半监督学习算法。

◆无监督学习的基本逻辑：数据中包含有**不同类别的统计特征**的子数据，依据**不同类别的统计特征**，把这些子类数据分开。

◆子类数据的含义解释，还是要利用先验信息。或者自主地赋予这些子类数据特定的名称或意义等！

◆半监督学习的基本逻辑：数据中包含有**不同类别的统计特征**的子数据，依据**不同类别的统计特征**，把这些子类数据分开。基于部分已有的标签数据，对子类数据的含义进行标定，或更精准地进行子类数据的划分！



◆四、Bayes决策理论下的地震数据分析

◆机器学习中的算法为什么都要借鉴类脑智能？或为什么借鉴类脑智能的ML算法有更强大的能力？

- ◆总体上看，人脑是自然界中最强大的智能处理器。人脑是一个分布（存储）并行计算的处理器，是一个改变硬件结构实现智能信息处理的“机器”、人脑处理信息是分层次的和信息稀疏表达的、人脑是一个非线性的信息处理器。
- ◆人类发展AI的目的本身就是要实现对人脑部分智能的替代或解放。发展类脑智能是当然的选择！
- ◆目前，我们对人脑进行信息处理的深层认识很不够。譬如不知道人脑如何抽象出事物的模式，语言（概念）和抽象的模式之间是如何关联的等等。



◆四、Bayes决策理论下的地震数据分析

◆机器学习中的算法为什么都要借鉴类脑智能？或为什么借鉴类脑智能的ML算法有更强大的能力？

- ◆AI和机器学习的发展一定是在对人脑本质功能更深入理解的基础上，建立合适的数学模型抽象地表达这些功能，据此提出更有效的算法，发展出更强大的人工智能机器或学习机器。
- ◆从数学上看，这就是一个如何能更好地实现Bayes决策的问题。



◆四、Bayes决策理论下的地震数据分析

◆机器学习中的算法为什么都要借鉴类脑智能？或为什么借鉴类脑智能的ML算法有更强大的能力？

- ◆类脑学习方法是累积式的、渐进式的，通过有标签或无标签的数据样本，不断地训练学习机器。理论上，这样的学习方式应该好于仅仅利用一次数据观测进行的估计（ $\hat{\mathbf{m}} = \mathbf{L}^T \mathbf{L}^{-1} \mathbf{L}^T \mathbf{d}^{obs}$ ）。这样的学习方式似乎能更好地挖掘数据中的信息，并据此做出更为可靠的判决！
- ◆但是，分析目前的ML算法，并没有体现出很强大的决策能力。问题出在什么地方了？如何发展出功能强大的、能做出合理决策的ML算法？



◆四、Bayes决策理论下的地震数据分析

◆Bayes决策理论下的地震数据分析方法应该是什么样的？

- ◆在非学习类算法下， $Lm=d^{obs}$ 这个正问题起到核心作用，与 Bayes参数估计理论结合，产生了当前绝大部分的数据分析思想和方法。目前的地震数据分析基本上就是在这个模式下进行的。
- ◆大数据与AI时代，人的部分智力工作（要求人做决策的工作）被替代成为了时代的要求。
- ◆地震数据分析中，哪些环节是需要人做出决策的？
 - ◆而且是人擅长做出这些决策的！




◆四、Bayes决策理论下的地震数据分析

◆Bayes决策理论下的地震数据分析方法应该是什么样的？

- ◆Bayes决策理论下的地震数据分析就是要在所有可能的环节中，发展出替代人做出合理决策的算法。

目录

- ◆一、概述
 - ◆二、Bayes决策理论
 - ◆三、地震数据分析中根本问题
 - ◆四、Bayes决策理论下的地震数据分析
 - ◆五、总结与讨论
- 



◆五、总结与讨论

- ◆ Bayes决策理论是数据科学中发展各种具体算法的指导思想。
- ◆ 类（尤其是类人）脑智能是人工智能（AI）的发展方向。ML的主要任务是提供实现人工智能（AI）需要的、在（目前是在Von Neumann构型的）计算机上运行的算法。
- ◆ （尤其是人）脑科学的发展（即对生物大脑的认知）是类脑智能发展的基础。抽象出实现类脑智能的数学模型及各种算法是类脑智能的关键。
- ◆ 无论如何，把Bayes决策理论作为有高度指导性的框架是非常有意义的！



◆五、总结与讨论

- ◆数据空间、决策空间、先验信息是Bayes决策理论的三个基本要素。
- ◆基于数据样本和（描述数据样本统计特征的）先验信息，实现决策代价最小的决策是Bayes决策的目标，也是人类智能的目标。
- ◆把AI和ML放在Bayes决策理论框架下展开认识和讨论，会产生更深入的认识，可以把应用做得更好！



◆五、总结与讨论

- ◆地震数据分析的最终目的是实现高精度的油藏描述。用人工智能（AI）实现高精度的油藏描述是不合适的（这不是个纯粹的决策问题，但可以用在各种具体环节上）。
- ◆用AI进行信息综合决定是否钻井是合适的（这是个纯粹的决策问题）。但是，我很怀疑AI做出的钻井决策的代价是否会比人做出的钻井决策的代价小！
- ◆当前，我们的重点依然是在地震数据分析中，利用ML和AI的新思想和新做法，提升老问题的解决效率和解决精度。



◆五、总结与讨论

◆这样的问题包括：

- ◆1、特征波场的识别与提取，尤其是初至波的识别与提取。
- ◆2、速度谱的智能解释与初始背景速度建模
- ◆3、电磁波场的特征变化与地表岩性空间变化之间关系。
- ◆4、近地表介质变化与波场变化之间的关系及地表一致性问题新解法
- ◆5、地质体（层位、断层、火山体、特殊地质体）图像识别及在地震波反演成像中的应用
- ◆6、PSF的识别与盲反褶积问题
- ◆7、



◆五、总结与讨论

◆从科学的角度看，有非常多的问题值得思考：

- ◆1、人工神经网络能建立输出与输入的非线性映射关系。但并不代表这样的网络具有可泛化应用的能力。网络可以泛化应用的本质是什么？
 - ◆我的观点是：输出与输入间存在比较线性的关系！或比较连续的关系！
- ◆2、我们应构建什么样的网络才能自动实现输入与输出间比较线性的关系之间的映射，以牺牲决策精度为代价实现稳定的泛化？
- ◆3、人类决策是基于模式（和概念）的定性的决策！基于大数据和高性能计算的AI和ML如何参考人类的决策模式，达到更好的定量决策！
- ◆4、分布式并行计算是否具有高度抽象性的、统一的计算模式？
- ◆.....



谢谢
欢迎批评指正