



中国石油大学
CHINA UNIVERSITY OF PETROLEUM

专业实践报告

实 践 课 题 人工智能在智能钻井平台上的应用

姓 名 孙延明

学 号 2020216015

专 业 领 域 应用统计

实践基地名称 阿美远东（北京）国际研发中心

专业实践类别 校级基地 院级基地 导师自主安排
学生自行联系

学 校 导 师 许韬 教授

校 外 导 师 李宇鹏

入学时间：2020 年 9 月 报告完成时间：2022 年 11 月

目 录

第 1 章 实践概况	2
1.1 实践单位概况	2
1.2 实践目的	3
第 2 章 实践内容与成果	4
2.1 研究方法与路线	4
2.2 实践内容	5
第 3 章 实践收获	13
第 4 章 总结和建议	14
参考文献	15

第 1 章 实践概况

1.1 实践单位概况

沙特阿拉伯国家石油公司（简称：沙特阿美）是一个有多年历史的综合国际石油公司，是世界最大的石油生产公司和世界第六大石油炼制造商，业务遍及沙特王国和全世界。它主要从事石油勘探、开发、生产、炼制、运输和销售等业务，拥有世界最大的陆上油田和海上油田。

沙特阿美的经营战略为立足本国，实施全国勘探计划。在提高总体原油供应能力和运输灵活性的同时，开始供应高价值超轻型原油；国内炼油工业纵向集成，国外加强与其他公司投资合作。使得整个王国的炼油工业处在整体规划之中，更加符合王国的炼油和产品经销长期需求；积极引进先进技术，提高工作效率及资源利用率。1982 年，沙特阿拉伯石油公司勘探和石油工程中心投入使用。与该中心的密切合作以及建立与其相关的实验室和研究开发中心使沙特阿美公司在上游（勘探和开采）技术后援方面基本不依靠其他石油公司；重视人员培训以及人力资源的开发。持续进步的关键，是要有一支受过良好教育和训练、能够担当重任的劳工队伍。多年来，公司设计了层次广泛的培训计划，包括从人行培训到有学位的专业人士高级管理计划，提高其工作技术，使雇员跟上新技术的发展，并培养世界级的专家。

沙特阿美在中国市场不断开拓业务，其与中国石化、石油业界有着长久的合作关系，并且两国在油气勘探开发及加工领域的合作程度在不断地加深。2010 年 1 月 14 日，中国石化和沙特阿美公司敲定了合资建炼厂的世界性项目，这将是中石化在海外的第一座炼油厂，也是截至目前最大的中国合资项目。2011 年 3 月 20 日，中国石油与沙特阿美公司共同签署了“云南炼油项目谅解备忘录”和“原油销售协议补充协议”。根据协议，双方将共同建设云南石化炼油项目，并从资源、加工、销售等方面进行一体化合作。

沙特阿美亚洲区总部位于北京。阿美亚洲是沙特阿美的全资子公司，也是连接沙特阿美和中国的业务和文化桥梁。其中的阿美北京研发中心（BRC）成立于 2015 年，专注于提供创造性的高影响力解决方案和工具，进而为沙特阿美提供支持，助力后者成为技术创新和发展领域示范领导者的战略目标。

阿美北京研发中心引领地球物理、地质学、油藏工程、炼油化工和战略运输研究等领域的先进技术研发，旨在重塑公司的上游和下游运营，从而为客户提供最先进的解决方案。通过与沙特阿美高级研究中心的科学家以及北京地方高校和科研院所密切合作，该中心已取得了一定显著的成就。比如在最新的油气上游尖端技术的探索，不同类型和尺度的地质资料的地质技术研发以及对油藏工程研究中化学驱油挑战以及先进地震成像的构建建模研究。

在未来，公司将把加强其上游和下游业务在全球的竞争地位作为未来的主要战略。同时在进一步的战略整合以及业务多样化中获取价值。并且在沙特王国和国际上积极扩大天然气活动和扩大公司品牌的全球认可度。在有效分配资金并保持审慎而灵活的资产负债表的基础上，通过原油价格周期提供可持续且不断增长的股息。最终利用先进的技术和富有生命力的创新行为来达到可持续运营的目的。



图 1 阿美公司商标

1.2 实践目的

人工智能技术为石油公司的发展带来了新的发展机遇和契机，阿美北京研发中心正在积极打造人工智能地质研究开发平台。本次实践参与了阿美北京研发中心地质领域中人工智能应用项目的开发工作，具体实践围绕机器学习方法在地质测井领域方面，为测井等相关技术提供新的技术支持。通过对测井钻井特征曲线的预测分析，可以减少开发成本，提高岩性识别准确率。通过实践学习，可以对人工智能在实际工程工作领域有更进一步理解。

第 2 章 实践内容与成果

2.1 研究方法与路线

2.2.1 方法选择

文献查阅法：在实践中利用工具书查找文献资料，同时以掌握的文献资料后面所附的文献资料来循环查找提高文献利用率。整理相关资料，进而深入了解深度学习中的 LSTM 等曲线回归预测的方法，从而将文献方法用于实际。了解目前该技术领域的发展现状以及实际应用情况。

比较分析法：在序列数据预测领域中，不仅有 LSTM 等深度学习方法，同时也有传统序列分析统计预测方法。在分类识别领域，不仅有集成学习方法，同时也有神经网络等深度集成学习方法。多个领域具体实践中，可以将两者进行对比分析，最终能够明确各方法在地质领域的应用范围。

实证分析法：将人工智能中深度学习方法用于序列数据回归的预测结果与实际测井曲线特征进行对比分析，可以有效检验已有算法在实际应用中的效果，同时帮助分析人工智能在地质石油领域的未来前景，进而更好服务石油地质领域。

2.2.2 可行性分析

现有机器学习以及人工智能理论发展日渐成熟，部分行业已经将人工智能技术应用到实际生产之中，取得的成果也较传统技术有较大提升。这说明人工智能技术具备实际价值，同时难点也在于如何将人工智能技术与实际生产相结合，这需要不断进行探索和改进。

阿美北京研发中心作为世界领先的石油地质领域技术研发公司，在相关领域具有领先的信息技术，拥有各种软硬件平台，同时可以与下游石油企业进行业务交流，拥有广泛的技术落地场景，这为课题的顺利开展提供了极大的便利条件。

2.2.3 专题分析报告

利用 PyQt5 工具包建立了相关的可视化软件，并初步开发了一个用于记录并处理钻井数据、测井数据、岩性解释数据的算法平台。这一算法平台大大提高了工作效率，并更好地了解目标油藏的属性特征。

完成项目专题指标梳理和人工智能地质应用体系的建立开发，全面、系统性的梳理业务发展流程。针对不同石油地质客户建立人工智能应用方案，结合各用户的实际，完成人工智能技术应用项目实践。

2.2.4 机器学习应用

根据已有数据，学习常规机器学习算法，熟悉算法的原理，优缺点以及适用

范围，并使用 Python 将机器学习算法应用到实际中。了解岩性识别相关背景知识，学习集成学习技术在岩性识别方面的应用，使用 Sklearn 库实现人工智能集成学习技术。

2.2 实践内容

2.2.1 基于人工智能算法的钻井数据分析及钻前预测

利用邻井钻井数据、层位数据、录井数据等对模型进行训练，实现了对当前井的钻前目标层位和岩相以及地层压力等的预测，从而优化钻井工程参数、优化作业。

钻井通常指勘探或开发石油、天然气等液态和气态矿产而钻凿井眼及大直径供水井的工程。钻井在国民经济建设中的应用极为广泛。在石油工业生产的环境下，对当前井的钻前目标层位和岩相以及地层压力等的预测对于优化钻井工程参数、优化作业具有重要意义。充分有效预测可以优化井筒结构并最终实现安全钻井和提质增效。

设备运行过程中获得的邻井钻井数据、层位数据、录井数据其数据量大并且设备自身结构和运行环境的复杂，因此充分理解设备的运行机理并且提取相关的数据特征难度大。所以充分利用深度学习自动提取钻井中的数据信息的隐藏特征，进而实现优化钻井工程成为当前的研究重点之一。

在实践阶段，首先基于异常检测算法孤立森林等对测井曲线进行异常识别，进而能够建立一套完善的应用方案。经过对数据进行预处理、调节模型参数等以达到在钻井平台在钻井工程中的实际应用。进而基于 LSTM 技术对测井等序列化数据，实现钻井数据预测来优化钻井工程任务。同时基于 XGBoost 等集成学习分类技术针对钻井解释预测数据构建岩性分类模型，完成标签分类任务。最终基于 PyQt5 工具建立了上述的相关软件，初步开发一个用于预测钻井数据并完成岩性分类建模的平台。



图 2 钻井平台示意图

2.2.1.1. 孤立森林

孤立森林是一种非常高效的异常检测算法，与随机森林算法相似，但孤立森林每次选择划分属性或划分点(值)时都是随机的，而不是根据信息增益或基尼指数来选择。

孤立森林算法通过随机选择一个特征来“隔离”观察结果，然后在最大值和最小值之间随机选择一个分割值，所选特征的值，由于递归分区可以用树结构表示，所以分离样本所需的分裂次数等于路径从根节点到终止节点的长度，这个路径长度在这些随机树的森林上平均，用于度量常态和实现我们的决策功能。当随机树组成的森林共同产生更短的路径长度时，它将很可能是异常。

孤立森林算法适用于连续数据的异常检测，将异常定义为“容易被孤立的离群点”。可理解为分布稀疏且离密度高的群体较远的点，在数据空间里面，分布稀疏的区域表示数据发生在此区域的概率很低，因此可认为落在这些区域里的数据是异常的。这就实现了实时数据的异常筛查。

孤立森林不定义数学模型也不对数据进行标记，而是去查找哪些点容易被“孤立”，它使用了一套非常高效的策略，例如我们用一个随机超平面来切割数据空间，切一次可以生成两个子空间，之后我们再继续用一个随机超平面来切割每个子空间，循环下去，直到每个子空间里面只有一个数据点为止。直观上来讲，我们可以发现那些密度很高的簇是被切分很多次才会停止切割，但是那些密度很低的点很容易很早就停到一个子空间里了。孤立森林算法得益于随机森林的思想，与随机森林由大量决策树组成一样，孤立森林也由大量的“树”组成，但和决策树不太一样，其构建过程也比决策树简单，是一个完全随机的过程。

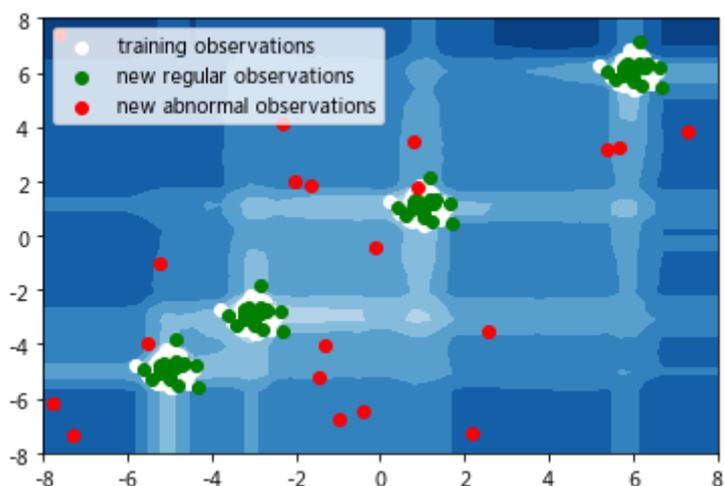


图3 孤立森林方法数据检测示意图

2.2.1.2 LSTM 模型介绍

长短期记忆模型（long-short term memory）是一种特殊的 RNN 模型，是为了解决 RNN 模型梯度弥散的问题而提出的；在传统的 RNN 中，训练算法使用的是 BPTT，当时间比较长时，需要回传的残差会指数下降，导致网络权重更新缓慢，无法体现出 RNN 的长期记忆的效果，因此需要一个存储单元来存储记忆，因此 LSTM 模型被提出；

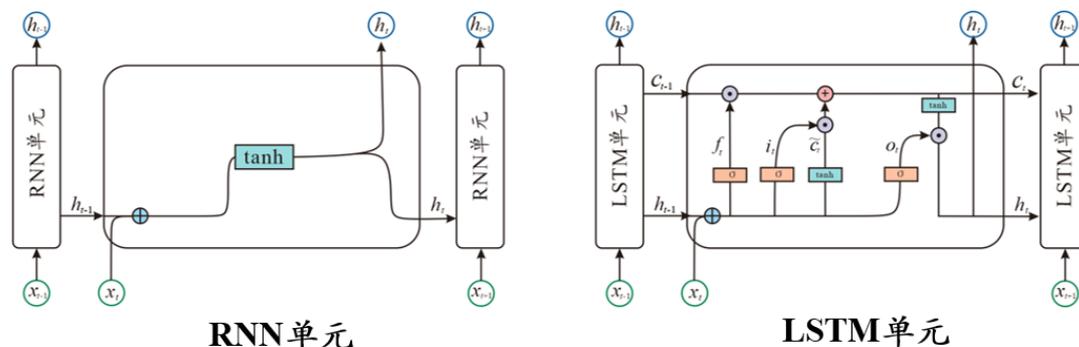


图 4 LSTM 与 RNN 对比图

LSTM 加入 3 个门结构来控制信息在序列中的传递，使信息能够长期传递下去。LSTM 的第一个门是遗忘门(forget gate)，确定丢弃哪些信息。第二个门是输入门(input gate)，用来确定存储哪些新信息。第三个是输出门(output gate)，确定输出的内容。在 LSTM 网络中，记忆单元 c_t 可以在某个时刻捕捉到某个关键信息，并有能力将此关键信息保存一定的时间间隔，

其中相应的机构遗忘门 f_t ，输入门 i_t ，输出门 o_t ，记忆单元 c_t ，隐状态 h_t 计算过程分别为如下：

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{2.1}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{2.2}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{2.3}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{2.4}$$

$$h_t = f(o_t \odot \tanh(c_t)) \tag{2.5}$$

其中 $\tanh(x)$ 为 tanh 激活函数， $\sigma(x)$ 为 sigmoid 函数， $f(x)$ 为根据实际任务（分类、回归等）需要所选择的应用函数。

2.2.1.3 LSTM 改进模型介绍

堆叠循环神经网络可以增强循环神经网络的预测能力。一种常见的增加循环神经网络深度的做法是将多个循环网络堆叠起来。

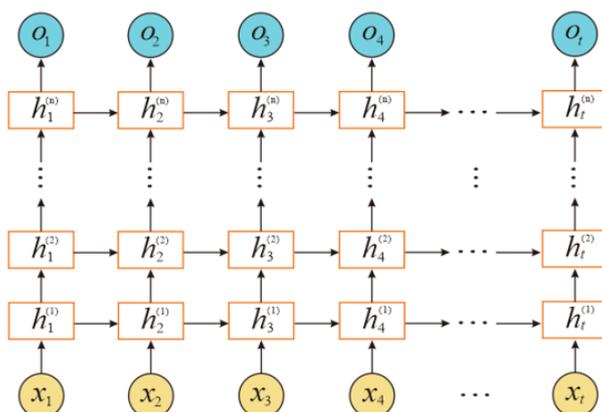


图 5 堆叠循环神经网络结构图

$h_t^{(n)}$ 为在时刻 t 时第 n 层的隐状态如下,其中 $f(x)$ 为根据实际目的需要所选中的应用函数:

$$h_t^{(n)} = f(V^{(n)}h_{t-1}^{(n)} + W^{(n)}h_t^{(n-1)} + b^{(n)}) \quad (2.6)$$

双向循环神经网络由两层循环神经网络组成,它们的输入相同,只是信息传递的方向不同。

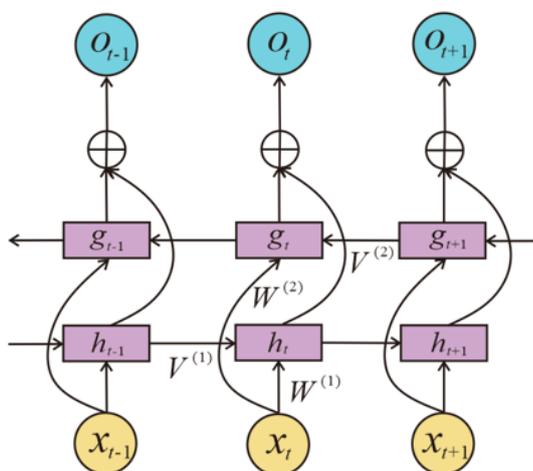


图 6 双向循环神经网络结构图

顺序输入的结果如下,其中 $f(x)$ 为根据实际目的需要所选中的应用函数:

$$h_t = f(V^{(1)}h_{t-1}^{(1)} + W^{(1)}x_t + b^{(1)}) \quad (2.7)$$

逆序输入的结果如下,其中 $f(x)$ 为根据实际目的需要所选中的应用函数:

$$g_t = f(V^{(2)}g_{t-1}^{(2)} + W^{(2)}x_t + b^{(2)}) \quad (2.8)$$

隐藏层输出:

$$o_t = h_t \oplus g_t \quad (2.9)$$

2.2.1.4. Gradient Boosting 思想

梯度增强方法（Gradient Boosting）是序列化集成的一种方法。当将其用在回归问题上，集成学习分类器设为 $F_m(x)$ ， m 为迭代次数也是集成的轮数，真实值为 y ，那么解决问题的损失函数我们设定为 $L(y, F_m(x))$ 。最普通的梯度增强方法就是将损失函数作为我们的目标函数，然后对目标函数进行处理改进。

我们引入泰勒公式来近似和简化目标函数，泰勒公式是一个用函数某点的信息描述其附近取值的公式。引入泰勒公式进行近似的本质为采用的数值优化的思维，用最速下降法去求解损失函数的最优解。

泰勒公式的一阶展开式为：

$$f(x) \cong f(x_0) + f'(x_0)(x - x_0) \tag{2.10}$$

那么对损失函数用一阶泰勒函数展开：

$$L(y, F_m(x)) \cong L(y, F_{m-1}(x)) + \frac{\partial L(y, F_{m-1}(x))}{\partial F_{m-1}(x)} (F_m(x) - F_{m-1}(x)) \tag{2.11}$$

更新后的集成学习分类器损失函数按照要求应当是不断降低，即 $L(y, F_m(x)) - L(y, F_{m-1}(x)) \leq 0$ ，那么就 $\frac{\partial L(y, F_{m-1}(x))}{\partial F_{m-1}(x)} (F_m(x) - F_{m-1}(x)) \leq 0$ ，所以当 $F_m(x) - F_{m-1}(x) = -\frac{\partial L(y, F_{m-1}(x))}{\partial F_{m-1}(x)}$ 时，满足基本条件，那么我们得到了更新集成学习分类器的基本公式：

$$F_m(x) = F_{m-1}(x) - \frac{\partial L(y, F_{m-1}(x))}{\partial F_{m-1}(x)} \tag{2.12}$$

上式符合优化分类器的梯度下降法，偏导表示损失函数的梯度方向，再取负表示负梯度方向，通过负梯度方向逐步优化模型。

2.2.1.5. XGBoost 模型

XGBoost 应用了 Gradient Tree Boosting 的思想，在目标函数和树模型两个方面进行了优化，以提高算法的计算效率和精度。

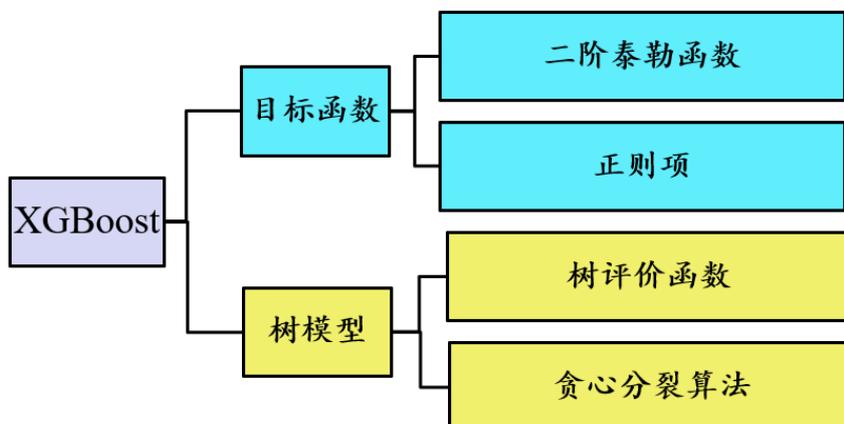


图 7 XGBoost 优化策略示意图

第 s 轮目标函数：

$$Obj^{(s)} = \sum_{i=1}^n L(y_i, \hat{y}_i^{(s-1)} + f_s(x_i)) + \Omega(f_s) \quad (2.13)$$

目标函数由两部分组成：

第一部分为损失函数，用来评估模型预测值和真实值之间的损失和误差，并在后续用二阶泰勒公式展开；第二项为正则化项，用来控制模型的复杂程度。

此处 $L(\cdot, \cdot)$ 是损失函数， $\hat{y}_i^{(s-1)}$ 是样本 x_i 在第 $(s-1)$ 轮迭代的预测值， $f_s(x_i)$ 指第 s 轮的新子分类器。

2.2.1.6. XGBoost 目标函数优化

XGBoost 也引入了泰勒公式来近似和简化目标函数，此处泰勒公式展开到二阶即 $f(x) \cong f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2$ ，那么目标函数就可以转换为

$$Obj^{(s)} = \sum_{i=1}^n \left[L(y_i, \hat{y}_i^{(s-1)}) + g_i f_s(x_i) + \frac{1}{2} h_i f_s^2(x_i) \right] + \Omega(f_s) \quad (2.14)$$

式中， g_i 为损失函数的一阶梯度统计， h_i 为二阶梯度统计。其分别为 $g_i =$

$$\frac{\partial L(y_i, \hat{y}_i^{(s-1)})}{\partial \hat{y}_i^{(s-1)}}, \quad h_i = \frac{\partial^2 L(y_i, \hat{y}_i^{(s-1)})}{\partial \hat{y}_i^{(s-1)^2}}.$$

因为常数项并不影响优化结果，因此可以对式(2.14)进一步简化，去掉常数项 $L(y_i, \hat{y}_i^{(s-1)})$ ，并将 $\Omega(f_s)$ 表达式代入公式，式(2.14)可以转化为

$$Obj^{(s)} = \sum_{i=1}^n \left[g_i f_s(x_i) + \frac{1}{2} h_i f_s^2(x_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \quad (2.15)$$

2.2.1.6. XGBoost 树模型优化

上式前半部分的 $f_s(x_i)$ 表示为一个子模型， ω_j 表示叶子节点的权重。为了后续对目标函数更好的推导，两者需要进行统一。 $f_s(x_i)$ 实质上是一个树模型，那么每个样本必定会被划分到该树模型的某一叶子节点上，因此对上式进行改写

$$Obj^{(s)} = \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right] + \gamma T \quad (2.16)$$

上式中， I_j 为叶子节点 j 的样本集，即落在叶子节点 j 上的所有样本。 $f_s(x_i)$ 将样本划分到叶子节点上，就可以计算到该叶子节点上的分数 ω ，因此当 $i \in I_j$ ，可以用 ω_j 代替 $f_s(x_i)$ 。所以上式两个公式是等价的。

将上式看作一个自变量为 ω_j ，因变量为 $Obj^{(s)}$ 的一元二次方程，根据其最值公式，对于固定的树结构，叶子节点 j 的最优 ω_j^* 为

$$\omega_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (2.17)$$

由上式可以看出，叶子权重不仅取决于一阶，二阶统计信息，还和 λ 值有关。得到了叶子节点的最优 ω_j^* ，对于固定的树结构，可以求得最优的目标函数值

$$Obj^{(s)} = -\frac{1}{2} \sum_{j=1}^T \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (2.18)$$

可用 G_j 表示 $\sum_{i \in I_j} g_i$ ，用 H_j 表示 $\sum_{i \in I_j} h_i$ ，则有

$$\omega_j^* = -\frac{G_j}{H_j + \lambda} \quad (2.19)$$

$$Obj^{(s)} = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (2.20)$$

上式可以作为评价一个树模型的评分函数，评分越小，表明该树模型越好。

利用评分函数即可对一个确定的树模型进行评价。在每一轮训练过程中，只要对所有候选树模型分别计算评价得分即可，从中选出最优的即可。但候选树的个数是无穷的，我们不可能得到所有候选树的评价得分。

XGBoost 采用贪心算法，先从树的根节点开始，计算节点分裂后比分裂前目标函数值是否减少，假设当前分裂前节点为 j ，其对目标函数的贡献为 $Obj_j =$

$$-\frac{1}{2} \frac{G_j^2}{H_j + \lambda} + \gamma$$

该节点分裂后，两个子节点的目标函数贡献为 $Obj_s = -\frac{1}{2} \left(\frac{G_{jL}^2}{H_{jL} + \lambda} + \frac{G_{jR}^2}{H_{jR} + \lambda} \right) + 2\gamma$ 。可以得到节点分裂的目标函数变化，值为 $Obj_{split} = Obj_j - Obj_s$ 。

在一个叶子节点的分离过程中，计算所有特征及其切分点的 Obj_{split} ，选取 Obj_{split} 最大的特征及其切分点作为最优特征和最优切分点，使叶子节点按照最优特征和最优切分点进行分裂。

上述算法是一种启发式算法，因为在节点分裂的时候，只选择了当前最优的分裂策略，而非全局最优的分裂策略。

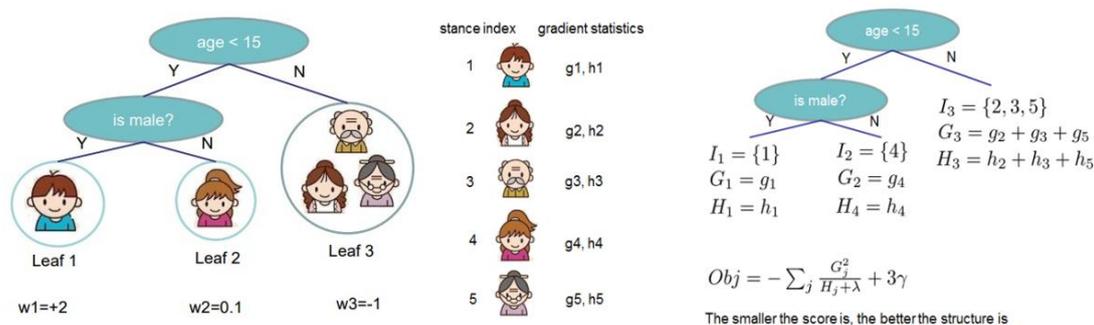


图 8 XGBoost 树构建过程示意图

2.2.2 其他工作

2.2.2.1 异常检测及数据探索工具箱研究

针对异常测井曲线识别任务，基于 Isolation Forest、one-class SVM 等异常检测算法，并在数据出现异常时及时反馈给业务，进行针对性策略的实进而完成智能离群检测任务。同时基于热力图、数据分布图等传统手段，针对油藏数据实现了智能数据研究和分析。同时为公司搭建科学的指标体系，并基于 PyQt5 工具建立了上述的相关软件，初步开发了一个用于分析油藏数据的平台。平台的构建使得公司人员更加清晰明了地看到数据，并获得相关结论。

2.2.2.2 基于自然语言处理的人工智能地质建模方法研究

针对地质沉积相数据特征提取任务，基于 NLP 技术对类别序列数据完成词嵌入表构建。针对增强地质数据集多样性和提高地质模型适应性任务，基于 SeqGAN 技术模拟生成新的地质沉积相特征。

同时基于项目成果，提出一种基于对应对数曲线信息的人工智能相解释模型。人工智能方法可以通过考虑训练井井日志曲线和标签的形状和值来解释相。在传统和通常的做法中，用于相解释的对数曲线通常作为单个值输入机器学习模型，无法考虑曲线形状信息。在所提出的解决方案中，根据对数曲线计算形状因子，将测井曲线形状信息集成到人工智能模型中。自然语言聚类技术在垂直相序列模式检测中实现，该图也反馈到人工智能模型中，根据案例研究表明，将测井曲线形状和相垂直堆垛模式集成到机器学习训练过程中大大提高了准确性，并且可以在训练和验证过程中再现更真实的地质堆垛模式。在最后的讨论中强调，成功实施机器学习技术最关键的部分是了解地球科学原理并找到合适的特征工程工作以将其集成到机器学习架构中。

第3章 实践收获

沙特阿美作为全球领先石油企业，拥有完善的公司管理体系，为员工提供了便捷高效的制度安排。通过本次实习，丰富了自己的工作经验，提高了应对业务的能力。作为全球石油行业优秀的公司代表，公司为全球石油行业智能化转型贡献了自己的力量。作为公司一员，我也感受到将自身发展紧密与世界能源战略结合在一起的自豪。

同时我在公司与同事高效的交流以及与公司内同事和外部运营商的对接人之间的高效率合作，让我感受到了沙特阿美公司中的简单与高效，在业务中也体会到了石油行业的数据驱动与快速迭代。

通过专业实践，我熟悉了算法工程师的日常工作，并学习了更多公司业务分析方法和统计学与机器学习分析方法。如 LSTM 模型，树模型。在知识方面，也不断意识到统计学知识在石油行业的重要性，对统计学知识的应用更加熟练。如常见的平均数、中位数、异常值检测、时间序列分析、参数估计、假设检验、多元线性回归等知识的应用。

近半年的实习经历，让我将实习项目与毕业论文紧密结合在一起，切实践行了专业硕士培养要求，为下一步毕业论文的顺利完成奠定了基础。

第4章 总结和建议

通过本次实践，将在本专业所学的统计知识充分与实际生产活动相结合，充分发挥专业特长为企业解决了若干实际问题，这是我踏入社会前的一段重要经历。在基于人工智能算法的钻井数据分析及钻前预测项目上已经取得初步进展，成功将 LSTM 模型以及 XGBoost 等模型应用于钻井设备中的序列数据处理上，并基于此建立一套完整钻井优化流程方案；同时基于 PyQT 等技术，成功开发了一套钻井算法平台，并更好地了解目标钻井数据的属性特征。并且在此基础之上我也参加了异常检测及数据探索工具箱研究和基于自然语言处理的人工智能地质建模方法研究等工作的研究，并且取得了一定的进展。

通过实习的过程，我加强了个人的实际动手操作与应用能力，同时我也了解到算法工程师在企业中的价值体现，并让我有了更加清晰的职业方向。在实践过程我也不断地通过努力加强优势与弥补不足。在未来的职业生涯中，我也会不断学习去扩大自己的知识面，不断精进自己各方面的能力。

在实践过程中，校内导师许韬教授给予了大力的支持，经常关心我在企业的学习生活情况，并给予了我很多实用建议。企业导师李宇鹏高级研究员在实践过程中不断给予指导，在学习实践方面给予了重要建议，并在日常工作中不断鼓励我，不仅让我在知识上得到了提升，而且锻炼了我在工作中的沟通交流能力。特别感谢两位导师的谆谆教导！

同时我也建议在以后的校企合作中，企业应当更加合理地匹配学生所学专业特长与工作岗位的需求，进一步提升实践效果。同时专业实践课题与项目分配尽可能与学生毕业论文课题相结合，在保障顺利完成论文的情况下，尽可能对企业项目有所帮助。

参考文献

- [1] 王燕, 郭元凯. 改进的 XGBoost 模型在股票预测中的应用[J]. 计算机工程与应用, 2019, 55(20): 202-207.
- [2] 杨贵军, 徐雪, 赵富强. 基于 XGBoost 算法的用户评分预测模型及应用[J]. 数据分析与知识发现, 2019, 3(01): 118-126.
- [3] Chen, T.Q., Guestrin, C., XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining., 2016.,:785-794.
- [4] Friedman, J.H., Hall, P., 2006. On bagging and nonlinear estimation. J Stat Plan Infer. 137(3), 669-683. <https://doi.org/10.1016/j.jspi.2006.06.002>
- [5] Friedman, J.H., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). Ann Stat. 28(2), 337-407. <https://doi.org/10.1214/aos/1016218223>
- [6] Giacinto, G., Roli, F., 2001. Design of effective neural network ensembles for image classification purposes. Image Vision Comput. 19, 699-707. [https://doi.org/10.1016/s0262-8856\(01\)00045-2](https://doi.org/10.1016/s0262-8856(01)00045-2)
- [7] Gorai, A.K., Raval, S., Patel, A.K., et al., 2021. Design and development of a machine vision system using artificial neural network-based algorithm for automated coal characterization. Int J Coal Sci Tech. 8(4), 737-755. <https://doi.org/10.1007/s40789-020-00370-9>
- [8] Gu, Y.F., Zhang, D.Y., Lin, Y.B., et al., 2021. Data-driven lithology prediction for tight sandstone reservoirs based on new ensemble learning of conventional logs: A demonstration of a Yanchang member, Ordos Basin. J Petrol Sci Eng. 207, 109292. <https://doi.org/10.1016/j.petrol.2021.109292>
- [9] Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second ed. Springer..
- [11] He, L., Wen, K., Wu, C.C., et al., 2019. A corroded natural gas pipeline reliability evaluation method based on multiple intelligent algorithms. Petrol Sci Bull. 4(3), 310-322. <https://doi.org/10.3969/j.issn.2096-1693.2019.03.028> (in Chinese)
- [12] Hou, E.K., Wen, Q., Ye, Z.N., et al., 2020. Height prediction of water-flowing fracture zone with a genetic-algorithm support-vector-machine method. Int J Coal Sci Tech. 7(4), 740-751. <https://doi.org/10.1007/s40789-020-00363-8>
- [13] Ji, X.X., Zhang, G. SAR Image Target Recognition with Increasing Sub-classifier Diversity Based on Adaptive Boosting. 2014 Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics; 2014: IEEE; 2014. p. 54-57.
- [14] Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. Front Neurobot. 7, 21. <https://doi.org/10.3389/fnbot.2013.00021>
- [15] Wang, Y.M., Li, Y.B., Li, X.P., et al., 2020. Recent progress on ANN-based pipeline erosion predictions. Petrol Sci Bull. 5(1), 114-121. <https://doi.org/10.3969/j.issn.2096-1693.2020.01.011> (in Chinese)
- [16] Yang, L.Y., 2011. Classifiers selection for ensemble learning based on accuracy and diversity. Procedia Eng. 15, 4266-4270. <https://doi.org/10.1016/j.proeng.2011.08.800>
- [17] Zhou, Z. When semi-supervised learning meets ensemble learning. International Workshop on Multiple Classifier Systems; 2009: Springer; 2009. p. 529-538.

- [18] Zhou, Z., Li, N. Multi-information ensemble diversity. International Workshop on Multiple Classifier Systems; 2010: Springer; 2010. p. 134-144.
- [19] Zhou, Z.H., 2011. When semi-supervised learning meets ensemble learning. Front Electr Electron Eng China. 6, 6-16. <https://doi.org/10.1007/s11460-011-0126-2>