

基于大数据分析技术的油气 录井软件开发实践与应用实例及价值

杨承禹（中国石油集团长城钻探工程有限公司录井公司，辽宁 盘锦 124010）

摘要：随着我国科学技术的发展和进步，石油勘探领域积极应用新型技术，其中录井技术是油气勘探的关键环节，通过对地下岩性和油气进行识别，确定油气层所处位置。以往开展油气录井时大多需要建立油气矿区物理模型或依照录井资料解释人员既往经验进行确定，工作复杂程度高，会受到诸多因素影响而产生误差问题。发展至今，单一应用传统录井手段已经难以适应当前飞速发展的石油勘探以及生产需求，基于此，本文结合大数据分析技术，深入探讨油气录井软件的开发要点，希望能够对当前录井业务的自动化改革和精细化水平提升提供参考，全面增强复杂地质条件下录井技术应用水平和油气勘探效益。

关键词：大数据分析技术；油气；录井；软件开发

当前我国石油勘探领域将焦点放置于复杂储层和中深部地层，因此前期井场勘探工作的重要作用越发凸显，单一应用测定手段难以针对当前石油储层、流体性质、地质岩性等进行精准识别，而录井作业可以随钻进行地质跟踪勘察，针对性更强，时效性更高。将大数据分析技术纳入录井作业中，可以转变以往录井资料解释方法的弊端，更好适应了复杂储存勘探录井解释方法的应用需求，再加上人工智能算法，提高录井解释精准度，为完井解释奠定良好基础和保障。

1 大数据分析技术在油气录井作业中的应用适配性

在我国传统油气开发及地质勘探领域中，各个环节均需要应用勘测数据内容以供评价与决策，例如测量录井、数据解释、油藏管理优化等，大多需要工程师通过丰富经验累积或将相关数据带入物理模型中进行模拟，实现数据的全面化分析，虽然此种分析手段和录井数据解释收获良好成效，但无法真正适用于复杂程度更高的油气储存之中。在现如今的油气开发领域，生产环节复杂程度越来越高，外界影响因素与工作人员、工程师等主观影响因素众多，单一应用物理过程进行模型模拟难度更大，且针对某一单一问题可以进行多种解释，所参考的数据类型也存在差异，甚至出现过于依赖主观判断实施解释的情况，难以真正体现油气勘探开发的理性客观解释要求，无法真正保障前期数据信息的应用效率和精准度^[1]。

针对此，可以全面应用大数据分析技术，也就是通过海量信息数据的自动化分析得出客观结论，将大

数据技术与数据挖掘技术等取得联系，可以从海量数据中挖掘数据信息的潜在规律，进而实现辅助决策等目标。例如在油气勘探开发领域中，前期地质勘探和该井场其他井位生产数据是录井作业的基础，通过对生产数据和勘探测量数据进行挖掘把控，摸清数据潜在规律，往往能够达到更高层次的分析水平，而大数据分析技术更加倾向于高维数据空间形态和自主化学习，能够通过最优数值方程进行高维数据空间的可管理性描述，同时也可以实现交叉验证和新模型训练探索。所以将大数据分析技术应用于油气录井数据分析与软件开发应用中，能够及时筛选海量数据，既可以提高隐藏数据和内在逻辑规律的把握水平，同时也可以将海量数据中的伪数据进行去除，避免发生结果误导情况。与此同时，很多基础数据无法实施等量齐观，为避免出现过于主观或分析不到位等情况，可以建立客观度更高、操作效果更好的数学模型，进行计算机自动化分析与处理，也就是将录井作业过程中收集的地层岩性、气测数据等资料和信息，结合以往以投产验证的储存资料进行对照，从而建立数学模型，进一步合理评价与预测即将开采的储藏岩性及含油性，为油气解释和石油勘探工程预警奠定了完善的数据基础和保障^[2]。

2 基于大数据分析技术的油气录井软件开发实践

2.1 油气录井软件开发系统框架

以大数据分析技术为核心，健全完善油气录井软件开发系统需要从多个功能区间进行分析，主要涵盖了数据信息的导入及预处理、数据信息内容存储、模

型搭建与训练以及录井数据实时预测等等。首先在数据信息导入方面,分别体现为历史数据记录和实时数据接入,前者侧重于对专业库以及数据库中原有数据特征值和数据标签进行提取,而后者应用传感器数据、实时数据库中数据实时接入。随后需要进行数据预处理,对海量数据进行清洗、集成、规约和变换,从而建立样本数据和实时数据。样本数据需要与数据存储结构模块进行衔接,并纳入模型数据库和实时数据内容,由模型训练平台进行演算、迭代与评估预测,最终实现大数据实时分析,例如现场解释与完井解释、工程预警等等。此类录井软件结构框架最终输出为大数据模型,可以对新型录井数据和勘探事件等进行实时预测与分析,最终以数据接口形式传输、呈现给相关用户或平台。

2.2 油气录井软件功能模块

2.2.1 数据接入模块

数据接入模块是整个油气录井软件运行的基础要素,同样也是关系最终解释成果和预警成果的重中之重,这一部分主要划分为历史数据和实时数据,其中历史数据主要体现为样本数据,如特征、标签等等,需要将样本数据作为训练大数据解释模型的基础内容进行运转和探索,而实时数据为油气井场当前状态下以传感器形式传输的数据类型。历史数据接入应用于模型训练阶段,大多以线下形式进行实现,随后最终呈现为线上软件的模型文件,对此,此类样本数据可以单独存储,如选择公开存储形式,可以对其建立专业数据库或数据库导入应用系统,搭建不同样本数据存储库,针对某一模型对应一张或多张表格或数据库,实现样本数据的共享。在历史数据接入范围可以从录井作业现场解释、完井综合解释和工程预警三个维度进行推进,其中录井现场解释涵盖了钻井数据、核磁共振数据、随钻测井、地化录井、岩屑录井以及录井油气解释相关内容,完井综合解释涵盖了地层测试求产、测井感应数据和电缆测试数据等报表内容,工程预警重点围绕模拟井漏、钻时气测数据、地层压力、钻井参数等进行历史数据的接入与共享,打造更加庞大的大数据信息资源^[3]。

2.2.2 数据预处理

在数据预处理功能模块中需要根据油气勘探开发中录井技术的应用、业务需求进行数据抓取,从而将有意义的信息通过接口送入预处理模块,该模块由不同处理算法构成,是正式建模之前的核心模块。在预处理模块中,需要对数据信息进行清洗和标准化转换,

做好变换计算,对存在缺失的数据进行补充,实现噪声识别,并在此基础之上对数据进行特征分析和要素抓取、事件选取,进行数据离散程度。例如在录井数据中需要结合勘探实际情况进行数据去除噪声和补充,并进行数据归一化计算、整合,随后才可以将其导入数据存储和模型训练平台之中。

当完成预处理之后,需要保障该环节输出数据包含多维特征,为有属性向量的数据表,例如需要针对录井作业生产环节中各类数据进行全面收集,通过架设传感器收集实时数据,从数据库中调取历史数据,并结合工程师模型和经验解释数据等构成数据体系。由于不同测量环境中所得到的数据类型复杂程度均存在差异,仪器工作状态和测量标准也不一致,因此很多数据均会存在大量噪声,且各自标准不一致,甚至出现过度信息冗余的数据类型;如果此类数据直接进行数据挖掘,则会导致数据分析失败和结果误差,虽然在预处理环节进行数据筛选和归一化整合,会耗费大量作业时间和软件运转时间,但总体而言,数据的质量水平有所保障。

例如进行数据缺失值处理时,常规类型的处理手段可以直接删除缺省值,以进一步保障数据的可靠性和可应用性,而针对其他类型并不是必须应用的数据特征值,可以直接以模型预测手段预测该缺失值的特征要素;而针对异常值则需要根据工程师经验和常识进行判断,如随钻测井数据为负值时大多为无效数据,密度测井值一般大于 1g/cm^3 、中子密度需大于100%等等。

2.2.3 模型训练与调整模块

模型训练与调整模块需基于预处理完成后的数据,结合大数据分析技术进行录井解释。与其他模块相比,模型训练平台以大数据分析技术为核心,可以进行离线操作,只需要将大量数据录入其中,提供样本数据,加入机器学习算法,即可得到与之相关的大数据解释模型,可以根据不同应用需求对大数据模型进行训练。例如,数据录入数据库中需结合训练和测试结果进行分离训练和验证,数据集采取五重折叠交叉验证手段。当应用大数据分析技术进行模型构建、分析测井数据、解释问题时,那么所得出的测井数据类型为数据集,而其中油层和水层的相关解释数据则可以体现为数据点标识标签,以大数据算法为核心构建训练模型即可实现高维边界,将不同类型的数据点进行优化分解,从而实现不同区间的划分;当录入新型测井数据时,可以通过大数据模型自行判断该数据

的具体落点,对应与之相关解释。

在常规类型的录井技术采集参数中,需要针对岩屑、气体等常规参数进行收集,并在此基础之上应用气体比率法、皮克斯勒法、3H 比值法等多种手段进行录井解释模板的构建。因为原始数据有多元性和异构性,因此在数据表整理和分析时,避免对数据进行单一并集,需要进行重复抽样和差值操作,依照数据原有性质进行协同索引。由考察结果可知在大数据模型准确度影响要素中,电阻率的影响度最高,声波测井次之,气体特征与其他要素生成互补作用,不同要素在油气解释敏感度方面各具差异,但如果在数据库中加入密度测井和声波测井以及三维荧光数据,也可以在现有数据模型的准确度基础之上提高运行精准度水平。

2.2.4 实时预测服务功能

基于前者数据库接入、数据预处理和模型训练与调整之后,可以完成模型的测试验证和优化选择,进而得到实时预测服务模块。这一模块恰恰是以大数据分析技术为核心得到的油气录井解释模型体系,如现场解释、完井综合解释和工程预警等,体系多样,输出精准度更高。

3 油气录井软件应用实例及价值

辽河凹陷位于我国渤海湾盆地东北部区域,辽河凹陷原有结构呈现多期化发展态势,经历了三大演化阶段,进而引申为由北东向主干断裂和近东西向次级断裂组成联合,凹陷主要划分为6个构造单元,其中尤以西部凹陷含油气水平最好,面积约为2500km²。该区域主要为油页岩结构,同时发育灰岩、粉砂岩等多种岩石特性。该区域油气富集,油田储量稳定增长,但受到地层岩性和流体性质等因素影响,钻井勘探难度大,通过录井手段可以直接获取钻井现场的数据资料等信息内容,从而得出分析结果。该区域油气勘探难度相对较高,作业环境存在诸多工程难题,录井技术应用区间需满足尤为复杂的采集环境,同时保障资料解释的精准度和后期预警要求。

针对于此,可以应用基于大数据分析技术的油气录井软件,对油气储藏区域进行深层复杂岩性识别和储存评价。在应用录井系统进行资料采集与处理时,要对照该区域经常以往录井数据要素进行岩性分析,应用气测录井技术和地球化学录井技术等,实施同位素录井技术等,实现井场数据采集与存储。该井场数据类型主要涵盖了地质、测井、钻井和测试共56项结构化数据,以针对不同应用区间和用户需求建立数

据库和服务终端。可以将录井作业过程中收集的地层岩性、气测数据等资料和信息,结合以往已投产验证的储存资料进行对照。结果可知,A井分别开展10层录井作业,其中人工解释与大数据解释基本一致,第一层井顶深3436.8,底深3440,人工解释结论为干层,而大数据解释中顶深为3436,底深为3440,结论为非产层;第四层人工解释顶深3630、底深3635,结论为水层,而大数据解释该层为气层,只有这一层的解释结论不一致。另外,在其他井位软件运行中,大数据解释的层次和储存划分更为精细、准确,通过实际验证和试油也证明了大数据解释内容的正确性。

针对此类油气勘探作业而言,高风险、高科技、高成本投入的特点异常明显,如果一味应用传统录井技术和手段,则难以真正满足我国油气勘探作业需求,对此需全面推进录井技术的智能化发展和数字化转型,并实现国产化设备革新进程,积极推进具有我国自主知识产权的录井技术体系应用。此次研究中以大数据分析技术为核心,研究提出了油气录井软件开发体系,通过实际验证不难发现,此类录井软件系统在数据源头处理和智能化分析与运用方面自主化程度更高,协同管理效能更为明显,能够有效适应环境复杂程度高、作业难度大的油气勘探开发模式。

4 结论

总而言之,现如今我国录井技术飞速发展,与之配套的软件系统越发优化,虽然相比于西方国家,我国相关技术体系起步较晚,但是随着国内油气勘探的全面深入,录井技术以及软件开发系统越发飞速发展,并逐渐实现了自主研发和对外出口。本文基于大数据分析技术,探讨了油气录井软件开发要素,首先分析了大数据分析技术在油气录井作业中的应用适配性,其次,分别围绕油气录井软件开发系统框架和功能模块实现进行重点分析,最后结合具体应用实例,探讨了此类软件系统的应用效果,希望能够对我国油气勘探开发领域录井技术的优化发展提供推进力量。

参考文献:

- [1] 王湘蜀,黄凌宇,等.基于大数据分析技术的油气录井软件开发[J].信息系统工程,2022,35(4):51-54.
- [2] 苏朝博,胡云,苑仁国,等.RISExpress V2.0软件在井场随钻地质跟踪中的应用[J].录井工程,2022,33(1):94-99.
- [3] 桑月浦,刘晓亮,王道伟.RISExpress 油气识别与评价软件的开发与应用[J].石化技术,2018,25(5):104-104.