

长输油气管道大数据挖掘与应用

于涛¹, 刘丽君², 陈泓君¹, 于瑶¹

(1. 中国石油天然气股份有限公司北京油气调控中心, 北京 100007; 2. 北京中油瑞飞信息技术有限责任公司, 北京 100007)

摘要: 针对未来长输油气管道智能化建设的需要, 结合油气管道数据采集与监视控制 (SCADA, supervisory control and data acquisition) 系统及运行参数, 对比传统理论方法和大数据挖掘方法的特点, 提出了大数据推动管道智能化的研究方向以及管道智能化研究的数字信息化、理论化和智能化 3 个步骤, 建立了管道智能化架构, 包括物理层、数据层、数据挖掘层、应用层和用户层共 5 个层次, 并确定以数据挖掘层为架构核心。统计分析、时序性预测和工况识别等应用案例表明, 利用大数据挖掘可有效解决实际生产的业务需求, 指导未来管道智能化的研究与建设。

关键词: 长输油气管道; 大数据; 智能化

中图分类号: TE89

文献标识码: A

doi: 10.11959/j.issn.2096-3750.2020.00174

Big data mining and application of long-distance oil and gas pipeline

YU Tao¹, LIU Lijun², CHEN Hongjun¹, YU Yao¹

1. PetroChina Co., Ltd., Beijing Oil & Gas Pipeline Control Center, Beijing 100007, China

2. Richfit Information Technology Co., Ltd., Beijing 100007, China

Abstract: In response to the need of the intelligent construction of the long-distance oil and gas pipeline in the future, combining with the supervisory control and data acquisition (SCADA) system and operating parameters of the oil and gas pipeline, comparing with the characteristics of traditional theoretical methods and big data mining methods, the direction of big data to promote intelligent pipeline and three steps of digital informatization, theorization, and intelligence of the pipeline intelligence research were proposed. The pipeline intelligent architecture was established, which included a physical layer, a data layer, a data mining layer, an application layer, and a user layer. The data mining layer was the core of the architecture. The statistical analysis, time series prediction and working condition identification and other application cases showed that the use of the big data mining could effectively solve the actual production business needs and guide the future pipeline intelligent research and construction.

Key words: long-distance oil and gas pipeline, big data, intelligent

1 引言

近年来, 大数据在工业数据领域的参数预测、设备监测和运行优化等方面发展迅速^[1], 有效提高了工业控制系统的运行管理水平^[2]。2020年初, 我国提出了“新基建”概念, 作为国家经济发展动脉的长输油气管道, 可借助“新基建”机遇, 通过管

道智能化的建设吸收数字化时代赋予的科技力量, 推动数字经济与管道上、下游产业的深度融合, 将海量数据资源转换为商业价值, 提升企业大数据创造价值的的能力。充分利用“新基建”中的 5G 通信网络及工业互联网、物联网等技术, 加速资源整合, 推动行业技术储备能力和应急管理体系的建设, 促进机器人检修、无人站场建设、无人机巡线等技术

收稿日期: 2020-05-06; 修回日期: 2020-06-30

通信作者: 于涛, 124114004@qq.com

基金项目: 国家自然科学基金资助项目 (No.51774311)

Foundation Item: The National Natural Science Foundation of China (No.51774311)

的应用,探索化石能源与新能源全面融合发展,实现管道数字孪生及远程分析监控,逐步建设低碳能源生态圈,有效提升能源利用率,从而提高油气管道的安全管控水平,实现提质增效的运行目标。

在长输油气管道运行中,调度员利用 SCADA 系统的压力、流量等参数,结合管道的设备状态变化和报警信息等实现管道的远程监控运行。随着管道自控通信、物联网技术的发展以及调度员远程调控运行经验和业务知识的积累,结合 SCADA 系统的实时数据及历史数据,探索将运行经验、业务知识结合系统报警信息等进行数据理论化和模型化,构建参数预测和工况智能识别模型,并使用实际生产数据驱动模型提高模型的预测精度和适应性,将是未来管道智能化发展的途径^[3-5],也是油气管道在“新基建”人工智能方面的研究前提。相比其他行业,油气管道具有点多、线长等特点,生产和运行维护环节多、难度大,实施大数据挖掘和智能化研究的困难多,因此,需要根据大数据挖掘特点,结合业务需求,构建管道智能化架构,从而指导未来管道智能化的建设与应用。

2 管道大数据挖掘与智能化

2.1 传统理论与大数据挖掘方法对比

长输油气管道的生产数据具有海量性 (volume)、多样性 (variety)、高速性 (velocity)、真实性 (veracity)、可见性 (visibility) 和价值 (value) “6V” 工业大数据特性^[6],同时具有更强的专业性、关联性、流程性、时序性和解析性等,由于同一工况受多个参数影响,因此具有模型维度高的特点。对长输油气管道生产数据的研究可采用传统理论方法和大数据挖掘方法^[7],传统理论与大数据挖掘方法流程对比如图 1 所示。

传统理论方法已经形成了系统的方法论,物理概念清晰,对于管道特定的参数预测有系统的理论

体系和公式。但在工业控制系统应用中,传统理论方法存在局限性,包括:1) 在模型构建过程中需要对参数进行理想假设和简化,影响误差;2) 自适应性差,在实际应用时,影响参数的变化导致模型适用性差;3) 分析较片面,难以反映宏观的时空关联特性。如长输热油管道使用苏霍夫公式计算管道沿线的油温^[8],近年来,其他学者也大多基于该公式进行研究,利用解析法或者数值分析法建立预测公式和模型^[9-11],但研究侧重于非稳态过程,且建立的预测公式和模型需要管道准确的设计参数以及沿线温度场、土壤导热系数等参数。由于长输热油管道站间距长,参数获取难度大,影响了研究成果的推广应用,实际生产适应性较差。

大数据挖掘方法不依赖于理论机理,可将历史数据和实时数据进行综合分析,得到多维度、宏观的时空关联特性,并根据预测结果进行调参优化,提升了预测结果的准确性和适应性。大数据挖掘方法与传统理论方法并不矛盾,在研究过程中,可利用理论方法的参数,确定模型的影响因素,提升研究效率,同时可进一步完善科学研究体系,从而推动研究方法的发展,大数据挖掘方法推动油气管道智能化的 4 个方向如下。

- 1) 根据业务需要对生产数据进行分析、建模,实现对管道摩阻、油温控制、设备运行管理等长周期数据进行有效监控,可有效优化运行过程,节能降耗。
- 2) 利用大数据先进算法将运行人员对工况的分析处置经验模型化,开发完成事件工况的捕捉、分析与处置分析系统,最终推动管道智能化控制,如对泄漏等工况实现自我判断处置等。
- 3) 从数据中挖掘运行过程中的隐性问题,对隐性问题进行趋势分析,在其成为风险隐患之前解决问题。
- 4) 根据以上 3 个内容的实现,利用反向分析法

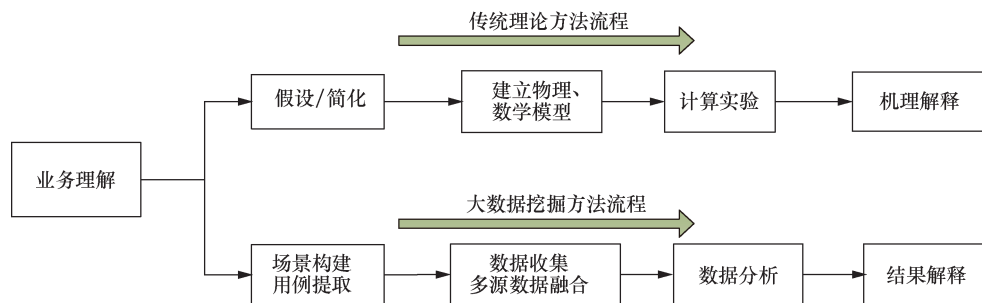


图 1 传统理论与大数据挖掘方法流程对比

对整个生产系统进行剖析建模，从设计和 SCADA 系统构建方面提升管道运行管理水平。

2.2 管道智能化

2.2.1 管道智能化研究流程

管道智能化是在信息论、人工智能及计算机科学发展基础上逐渐形成的一类高级信息技术^[12-13]。基于大数据研究的智能化突破了传统理论中必须基于数学模型的框架、利用先进的数学算法使其具备非线性的人类思维特征，未来可具备在线辨识、决策或总体自寻优能力以及分层信息处理、决策等功能^[14]。通过对参数及设备的自动监测、调节保护，可以提升管道安全优化运行水平，实现对异常工况的实时捕捉、判断，辅助运行人员的日常工作，提高工作及决策效率，逐步减少人员的工作量与工作压力。管道智能化控制需通过数字信息化、理论化、智能化 3 个步骤实现，利用生产系统实际的生产数据结合专家经验，建立基于实际生产数据驱动的算法模型，并应用于专家分析、智能处置及智能控制，管道智能化研究流程如图 2 所示^[10]。近年来，随着管道 SCADA 系统的推广应用，逐步建立了比较完善的数据采集、存储及使用系统，基本实现管道数字化。下一步需根据未来业务需求对数据进行挖掘分析，建立满足管道智能化控制需要的算法模型，并根据实际运行情况及时调整模型，提升模型的适

用性和准确性，为管道智能化控制奠定理论基础。

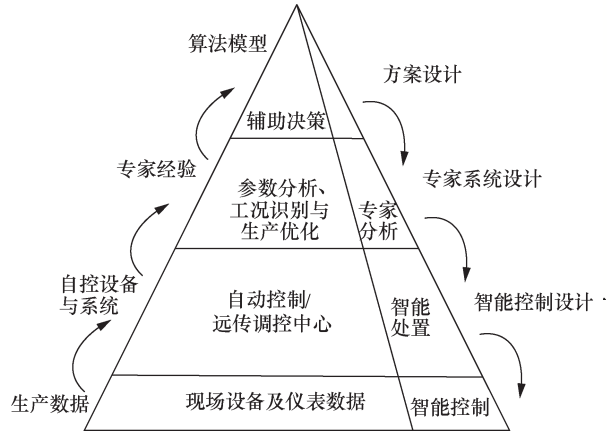


图 2 管道智能化研究流程

2.2.2 管道智能化架构

管道智能化的核心是基于生产数据构建相应的预测、识别模型，实现管道调控运行、管理业务的有效提升与智能展现。管道智能化系统架构如图 3 所示^[15]，包括物理层、数据层、数据挖掘层、应用层和用户层。在系统的总体架构中，需要充分结合国家、行业及企业标准规范，形成系统需要的标准规范体系。同时，考虑系统建设过程中的网络、数据安全因素，建立安全保障体系，确保系统的安全性。

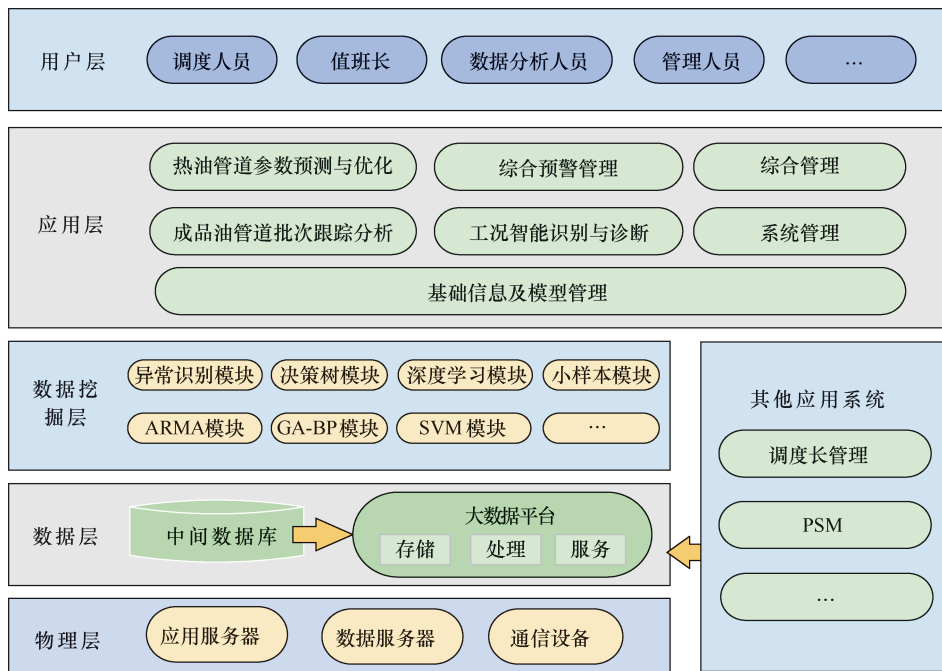


图 3 管道智能化系统架构

管道智能化系统架构的核心层是数据挖掘层；基础层为物理层、数据层，主要取决于管道的 SCADA 系统和信息化构建；应用层和用户层体现业务需求的展示与管理功能。在实际开发过程中，重点通过应用层和用户层提出业务需求，从数据层提取数据并在数据挖掘层利用理论公式、数据回归、数据挖掘等方式，将基础数据进行加工处理，使其成为可以为应用系统提供支撑的数据或预测模型，数据挖掘层的应用架构如图 4 所示。

数据驱动的模式是未来管道智能化的核心，数据的集成和清洗是大数据挖掘方法的基础，大数据的价值产生于数据分析挖掘。管道 SCADA 系统和相关生产系统产生了大量生产数据，利用大数据挖掘方法所研究建立的预测和识别模型，通过大数据平台实现管道工况识别、报警预警和预测优化等功能，指导调度员和管理人员的工况处置和管理决策，同时可实现管道生产数据的平台展示与报表分析，提升了管道的运行安全与管理效率。

3 管道大数据应用

3.1 应用需求

长输油气管道的大数据应用需要根据不同的业务需求和数据特点选择合适的研究方法，以热油管道为例，油温预测是管道安全优化运行的前提，工况识别是未来管道智能化研究的基础。当前的油温预测计算与分析局限于理论公式，而在实际生产

中主要体现不同站场的横向油温对比，无法实现非稳态油温的准确预测与不同站场、不同时刻的纵向油温对比。研究工况智能识别的学者较少，且已有研究大多侧重于泄漏等工况，局限性较大，无法应用于未来管道智能化控制。针对以上热油管道的实际业务需求，选取统计分析、循环神经网络 (RNN, recurrent neural network)、决策树和深度学习等大数据挖掘方法进行研究应用。

3.2 应用与分析

3.2.1 统计分析

统计分析方法针对日常运行的历史数据，通过整理、分析、描述等方法对当前数据进行归纳，从而获得工况参数潜在的价值。某高含蜡原油管道 2013—2017 年地温数据趋势如图 5 所示，此管道 2015—2017 年全线温降对比趋势如图 6 所示。

在图 5 中，管道的两个站场地温数据趋势符合正弦分布，同时可获得每年的最高温度及最低温度时段。根据图 5 的数据，运行时可以建立不同流量台阶，以每年的 10 月为例，地温处于下降时段，建议安排上旬管道输量较小，随着地温降低，输量增大，通过调整输量来减少地温导致的油温变化，同时可建立周期性的油温预测函数，为管道油温预测提供基础数据。在图 6 中，对比每年不同时期的全线温降，可获得全线温降的横向和纵向对比，指导站场启炉和加热炉负荷调整，既保证了管道的安全运行，又可以根据以往经验结合管道实际参数进

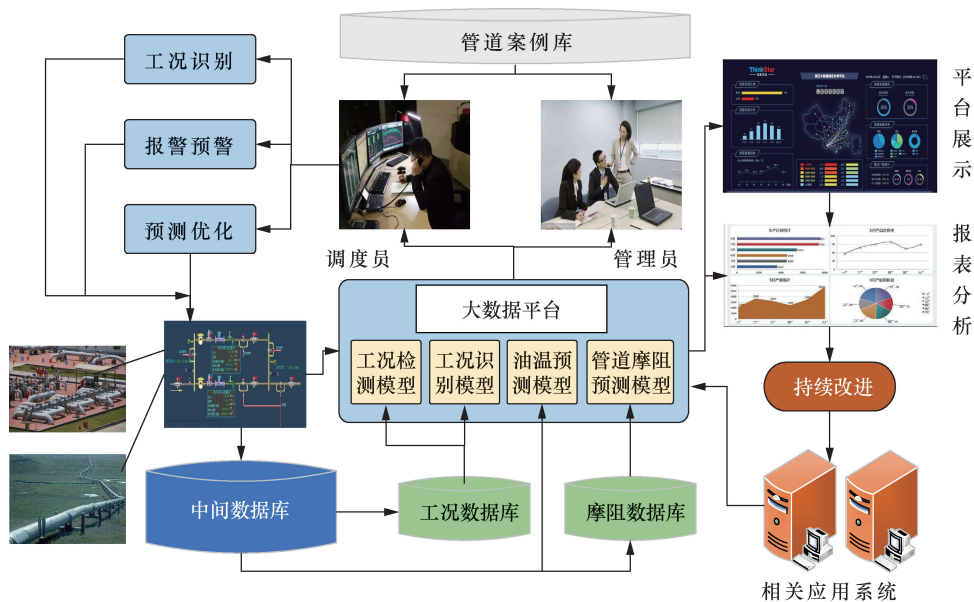


图 4 数据挖掘层的应用架构

行工况调整,从而实现节能降耗。同时可以判断得出管道加热炉调整数据中的温降突变点,获得管道清管启炉或热洗的时段和次数,以便于制定未来管道清管方案。

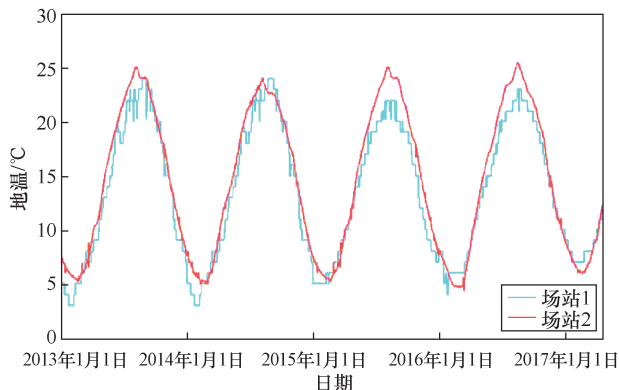


图 5 某高含蜡原油管道 2013—2017 年地温数据趋势

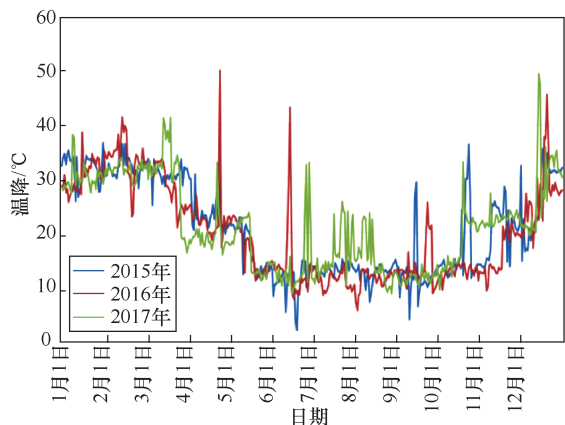


图 6 某高含蜡原油管道 2015—2017 年全线温降对比趋势

3.2.2 时序性预测

热油管道非稳态油温的研究主要有解析法和数值算法^[16-17],但因为研究过程需要准确的油品物理性质、沿线温度场、管道结蜡层及土壤导热系数等参数,所以模型边界条件和初始条件设定难度大,且现有的非稳态油温预测模型很难融入大数据平台,研究成果应用具有一定的局限性,实际生产的适应性差,不利于实时参数的预测及管道智能化控制平台的应用。热油管道的非稳态油温具有时序性特点,根据上、下游站场油温的变化趋势建立非稳态油温的时序性预测模型。通过分析 RNN 相关算法在自然语言处理和时间序列回归预测的成果^[18-20],将非稳态油温预测看作两个非等比数列对应关系,选取 RNN-seq2seq (sequence 2 sequence) 算法构建非稳态油温预测模型, RNN-seq2seq 模型结构示意图如图 7 所示,并利用实际生产数据训练测试模型^[21]。

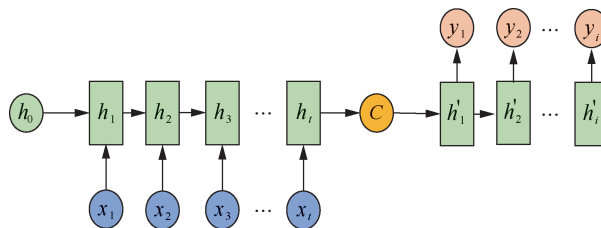
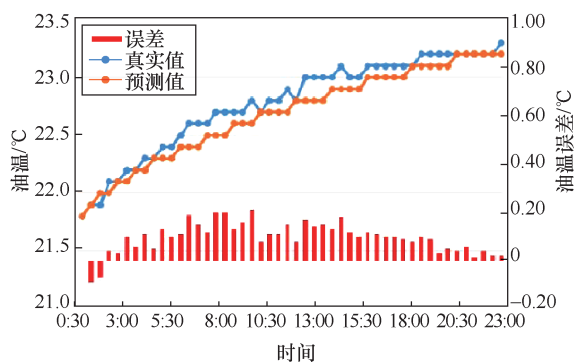
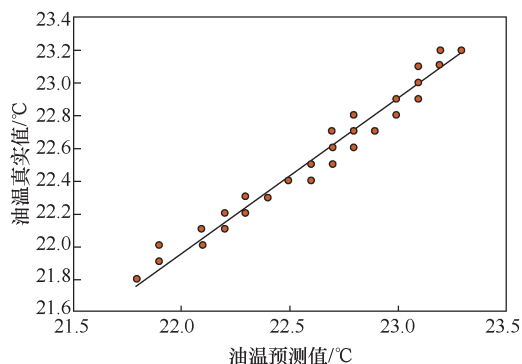


图 7 RNN-seq2seq 模型结构示意图

利用完成训练的模型预测上游启炉到达下游站场的非稳态变化过程,非稳态油温预测结果对比如图 8 所示,油温的预测值与真实值误差在 $\pm 0.2^{\circ}\text{C}$ 以内,整体变化趋势相同,相关系数平方为 0.97,且预测值的趋势变化更平滑,有利于分析油温的发展趋势。综上所述,通过 RNN-seq2seq 算法建立的非稳态油温预测模型可实现热油管道下游站场进站油温的非稳态预测,且模型具有预测准确性高、适合推广应用等特点。



(a) 油温趋势预测及误差对比



(b) 油温预测值与真实值相关性对比

图 8 非稳态油温预测结果对比

3.2.3 工况识别

及时发现长输油管道的非稳态工况有利于开展后续应急处置工作,降低管道运行风险。但工况发生后,参数变化较大,并伴有大量报警信息,尤其在工况频繁调节时,如何快速、准确地识别工况及减少漏报、误报是管道智能化的研究重点。根

据油气管道 SCADA 系统参数及报警信息特点，提出了基于报警及设备状态量变化的决策树模型和基于压力、流量等参数的深度学习工况智能识别模型。

1) 决策树

决策树是一类简单、应用广泛的回归与分类技术，其本质是一种在输入空间上不断划分特征空间并进行分类或者回归的算法。根据原油管道 SCADA 系统的泵和阀门等主要设备的状态量和报警信息，结合业务专家给出的工况判断流程，建立基于决策树算法的工况智能识别模型，典型站场工况识别决策树模型如图 9 所示。

2) 深度学习

深度学习是一类模式分析方法的统称，其中卷积神经网络 (CNN, convolutional neural network) 是深度学习典型的网络结构，主要应用于图像识别。CNN 能够学习大量的输入与输出之间的映射关系，

且只需用已知的模式对 CNN 进行训练。CNN 结构一般包括特征提取层和特征映射层，层级结构一般包括数据输入层、卷积层、池化层与全连接层等^[22-23]，并可以按照此结构不断扩充，CNN 结构如图 10 所示。

研究利用 CNN 的典型网络结构 VGG-16 (visual geometry group) 建立管道工况智能识别模型，即将流量、压力等工况数据转化为图像矩阵，将工况识别的分类问题转化为图像识别问题，利用图像识别领域的深度学习建立工况智能识别模型。

将每个工况数据归一化处理后，转化成灰度图，数据样本归一化如图 11 所示，对样本数据滑窗采样生成训练样本图像，数据的滑窗采样如图 12 所示。通过对样本集的训练优化，研究构建基于深度学习的工况智能识别模型，经验证最终识别准确率可达 91.9%。

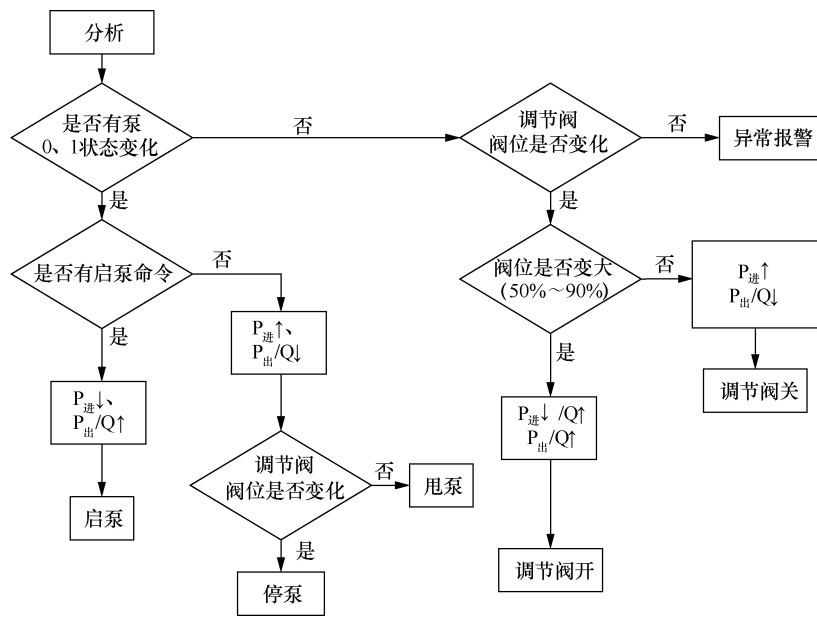


图 9 典型站场工况识别决策树模型

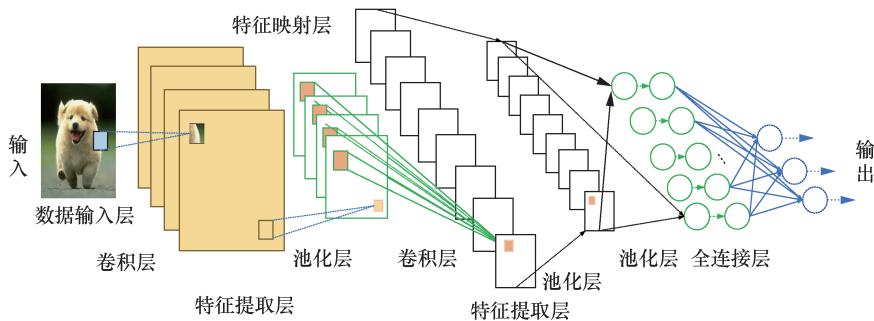


图 10 CNN 结构

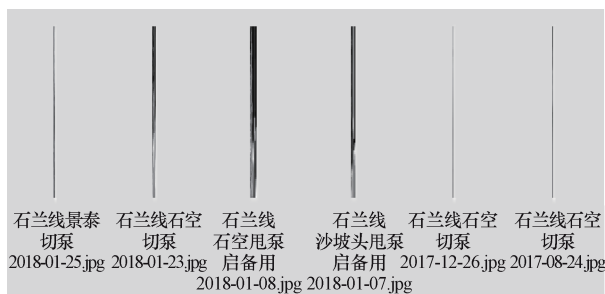


图 11 数据样本归一化

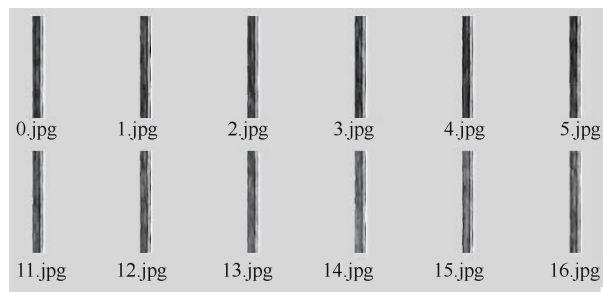


图 12 数据的滑窗采样

工况的智能识别包括决策树和深度学习工况智能识别两个模型，其中决策树模型侧重于管道 SCADA 系统的状态量，深度学习工况智能识别模型侧重于模拟量，将模拟量生成图像，并进行识别研究。在实际应用过程中，两者通过对模型的预测结果进行相互验证，实现识别报警，若只有一个模型识别出工况，则单独报警，经技术人员分析后进行处理，进而提升模型的预测准确性，防止工况的误报和漏报。

4 结束语

通过以上研究，可以获得如下结论。

1) 对比传统理论方法和大数据挖掘方法的研究特点，获得研究过程中可利用理论方法确定模型的影响因素，提升研究效率，同时得出了大数据挖掘方法推动油气管道智能化的 4 个方向。

2) 管道智能化控制需要经过数字信息化、理论化、智能化 3 个步骤，提出了管道智能化研究架构，指导后续应用研究工作的开展。

3) 研究确定以数据挖掘层为核心的管道智能化建设架构，为数据驱动模型构建智能化管道奠定了基础。

4) 根据油气管道大数据业务需求，给出统计分析、时序性预测和工况识别等应用案例，可见利用大数据挖掘方法可有效解决实际生产业务需求。

5) 目前，油气管道大数据研究应用还处于起步

阶段，管道智能化概念性较强，大多侧重于单业务数据孪生与应用，业务之间存在壁垒和信息“孤岛”，后续需要建立一套完整的理论研究体系，构建高质量的样本数据库，培养跨学科的研究人才，为管道智能化研究应用奠定理论基础和人才基础。

参考文献:

- [1] 吴军. 智能时代[M]. 北京: 中信出版社, 2016.
WU J. The age of intelligence[M]. Beijing: CITIC Press Group, 2016.
- [2] 李姐, 倪军, 王安正. 从大数据到智能制造[M]. 上海: 上海交通大学出版社, 2016.
LI J, NI J, WANG A Z. From big data to intelligent manufacturing[M]. Shanghai: Shanghai Jiao Tong University Press, 2016.
- [3] 唐大为, 朱静. 智能化管道数据采集系统设计思路初探[J]. 价值工程, 2015(5): 220-222.
TANG D W, ZHU J. Design thought of intelligence pipeline data acquisition system[J]. Value Engineering, 2015(5): 220-222.
- [4] 徐善丹. 关于智能化与数字化长输管道的探讨[J]. 广东化工, 2015, 41(11): 121-122.
XU S D. A discussion about digital and intelligent pipeline[J]. Guangdong Chemical Industry, 2015, 41(11): 121-122.
- [5] 于达, 熊毅, 朱婷婷, 等. 输油气管道智能化事故应急体系建设[J]. 油气储运, 2015, 34(10): 1038-1042.
YU D, XIONG Y, ZHU T T, et al. An intelligent emergency response system for failure of oil and gas pipeline[J]. Oil & Gas Storage and Transportation, 2015, 34(10): 1038-1042.
- [6] 李杰. 工业大数据[M]. 北京: 机械工业出版社, 2017.
LI J. Industrial big data[M]. Beijing: China Machine Press, 2017.
- [7] 谭伟业. 浅谈管道数字化与智能化[J]. 中国管理信息化, 2014, 17(18): 45-48.
TAN W Y. Discussion on pipeline digitization and intellectualization[J]. China Management Informationization, 2014, 17(18): 45-48.
- [8] 杨筱衡. 输油管道设计与管理[M]. 东营: 中国石油大学出版社, 2006.
YANG X H. Design and management of oil pipeline[M]. Dongying: China University of Petroleum Press, 2006.
- [9] 王海琴. 含蜡原油热输管道沿程温降计算[J]. 油气储运, 2002, 21(7): 9-12.
WANG H Q. Calculation and analysis of temperature drop along buried hot waxy oil pipeline[J]. Oil & Gas Storage and Transportation, 2002, 21(7): 9-12.
- [10] 张争伟, 凌霄, 宇波, 等. 埋地热油管道正常运行的数值模拟研究[J]. 工程热物理学报, 2008, 29(8): 1389-1392.
ZHANG Z W, LING X, YU B, et al. Numerical simulation study on normal operation of buried hot oil pipeline[J]. Journal of Engineering Thermophysics, 2008, 29(8): 1389-1392.
- [11] YU B, LI C, ZHANG Z, et al. Numerical simulation of a buried hot crude oil pipeline under normal operation[J]. Applied Thermal Engineering, 2008, 30(17): 2670-2679.
- [12] 张钟俊, 蔡自兴. 智能控制与智能控制系统[J]. 信息与控制, 1989(5): 30-39.
ZHANG Z J, CAI Z X. Intelligent control and intelligent control system[J]. Information and Control, 1989(5): 30-39.

- [13] HAN J W, MICHELINE K, JIAN P. 数据挖掘概念与技术[M]. 北京: 机械工业出版社, 2015.
HAN J W, MICHELINE K, JIAN P. Concept and technology of data mining[M]. Beijing: China Machine Press, 2015.
- [14] 杨静. 国内外智能化控制系统发展态势的研究[J]. 装备机械, 2016(1): 59-64.
YANG J. Research on development trend of intelligent control system at home and abroad[J]. Equipment Machinery, 2016(1): 59-64.
- [15] 吴潜. 大数据背景下石化工业实现智能制造的思考[J]. 化学工业, 2016, 34(2): 15-18.
WU Q. Big data utilization aiming at smarter manufacturing of petrochemical industry[J]. Chemical Industry, 2016, 34(2): 15-18.
- [16] 贾永英, 刘扬, 王玉洁, 等. 非牛顿原油层埋地热油管道沿线温度的确定[J]. 大庆石油学院学报, 2006, 30(6): 51-52.
JIA Y Y, LIU Y, WANG Y J, et al. Temperature determination of laminar of non-Newtonian crude oil in buried pipelines[J]. Journal of Daqing Petroleum Institute, 2006, 30(6): 51-52.
- [17] YU B, LI C, ZHANG Z, et al. Numerical simulation of a buried hot crude oil pipeline under normal operation[J]. Applied Thermal Engineering, 2008, 30(17): 2670-2679.
- [18] 胡新辰. 基于 LSTM 的语义关系分类研究[D]. 哈尔滨: 哈尔滨工业大学, 2015.
HU X C. Research on semantic relation classification based on LSTM[D]. Harbin: Harbin Institute of Technology, 2015.
- [19] ZHAO F, FENG J S, ZHAO J, et al. Robust LSTM-autoencoders for face de-occlusion in the wild[J]. IEEE Transactions on Image Processing, 2018, 27(2): 778-790.
- [20] GERS F A. Learning to forget: continual prediction with LSTM[J]. Neural Computation, 2000, 12(10): 2451-2471.
- [21] CHUNG E, PARK J G. Sentence-chain based seq2seq model for corpus expansion[J]. ETRI Journal, 2017, 39(4): 455-466.
- [22] 乔丽. 基于 CNN 的工作缺陷检测方法研究及系统设计[D]. 武汉: 华中师范大学, 2016.
QIAO L. Research and system design of workpiece defect detection method based on CNN[D]. Wuhan: Central China Normal University, 2016.
- [23] 李鹏辉. 基于深度学习的油井功图智能识别[D]. 洛阳: 河南科技大学, 2015.
LI P H. Intelligent recognition of oil well power diagram based on

deep learning[D]. Luoyang: Henan University of Science and Technology, 2015.

[作者简介]



于涛(1982-), 男, 山东青岛人, 博士, 中国石油天然气股份有限公司北京油气调控中心高级工程师, 主要研究方向为长输油气管道调控运行管理、数据挖掘分析与应用等。



刘丽君(1982-), 女, 山东烟台人, 北京中油瑞飞信息技术有限责任公司工程师, 主要研究方向为长输油气管道信息化、智能化建设。



陈泓君(1989-), 女, 山东菏泽人, 中国石油天然气股份有限公司北京油气调控中心工程师, 主要研究方向为长输油气管道调控运行管理、数据分析。



于瑶(1989-), 女, 吉林松原人, 中国石油天然气股份有限公司北京油气调控中心工程师, 主要研究方向为长输油气管道安全监控。