文章编号:1673-5005(2018)03-0154-08

doi:10.3969/j.issn.1673-5005.2018.03.019

# 基于 Hessian 正则化多视角学习的抽油机井 工况识别新方法

周 斌, 王延江, 刘伟锋, 刘宝弟

(中国石油大学(华东)信息与控制工程学院,山东青岛 266580)

摘要:针对大数据下抽油机井采油生产系统特点以及目前抽油机井工况识别研究中存在的问题,为进一步提高抽油 机井工况识别精准率和工程实用性,提出一种基于 Hessian 正则化多视角学习的抽油机井工况识别新方法。首先通 过先验知识和专家经验,选择实测地面示功图、电功率和井口温度信号3个视角并进行特征提取,然后利用 log 损失 函数建立 Hessian 正则化多视角 logistic 回归工况识别模型,最后采用交替优化算法求取最优解并进行分类识别。应 用该方法对胜利油田某区块11种抽油机井典型工况进行识别,其识别效果分别比传统的基于实测地面示功图、实测 电功率及特征连接多源识别方法提高了2.4%、11%和13.8%,而在少量标记训练样本下该方法识别效果更优,从而 验证了该方法的有效性。

关键词:抽油机井;工况识别;多视角学习; logistic 回归; Hessian 正则化

中图分类号: TE 938; TP 391.4 文献标志码: A

**引用格式**:周斌,王延江,刘伟锋,等.基于 Hessian 正则化多视角学习的抽油机井工况识别新方法[J].中国石油大学 学报(自然科学版),2018,42(3):154-161.

ZHOU Bin, WANG Yanjiang, LIU Weifeng, et al. A working condition recognition method of sucker-rod pumping wells based on multi-view learning and Hessian regularization [J]. Journal of China University of Petroleum (Edition of Natural Science), 2018,42(3):154-161.

# A working condition recognition method of sucker-rod pumping wells based on multi-view learning and Hessian regularization

ZHOU Bin, WANG Yanjiang, LIU Weifeng, LIU Baodi

(College of Information and Control Engineering in China University of Petroleum(East China), Qingdao 266580, China)

**Abstract**: To resolve the problems in working condition recognition of sucker-rod pumping wells and to further improve the accuracy and practicality, a novel method based on multi-view learning and Hessian regularization to identify the working condition was proposed. Firstly, the measured dynamometer cards, electrical power and wellhead temperature data were characterized based on the prior information and empirical knowledge. Then a multi-view logistic regression model with log loss function and Hessian regularization for working condition recognition was established. Finally, the working condition was classified and recognized by an alternating optimization algorithm. The proposed method was applied to eleven cases of typical working condition recognition in a block in Shengli Oilfield, and the results were compared with traditional recognition methods based on measured dynamometer cards, electrical power data and multi-sources of feature connection, respectively. The comparison shows that the recognition rates are improved by 2.4%, 11% and 13.8%, respectively. The performance is even much better with a small amount of marked training samples.

Keywords: sucker-rod pumping wells; working condition recognition; multi-view learning; logistic regression; Hessian regularization

收稿日期:2017-06-13

基金项目:国家自然科学基金项目(61671480)

作者简介:周斌(1970-),女,博士研究生,研究方向为故障诊断与智能信息处理。E-mail:freetzb@163.com。

通信作者:王延江(1966-),男,教授,博士,博士生导师,研究方向为模式识别与智能信息处理。E-mail:yjwang@upc.edu.en。

· 155 ·

由于抽油机井具有油田开采普及率高、工况复 杂多变、故障发生率高等特点,及时精准地识别油井 工况对油井极大实现降本增效具有重大意义。目前 抽油机井工况识别方法主要有以下几类,①基于示 功图识别方法:②基于电参数识别方法:③基于多源 数据识别方法。多数工况识别方法基于示功图识别 技术[16],主要是利用泵功图(实测地面示功图或实 测电参数通过相应模型计算得到)或实测地面示功 图结合人工智能方法(矩特征、小波包变换、神经网 络、支持向量机等)进行工况识别。基于电参数识 别方法主要是利用电功图[7](由实测电参数计算得 到)或实测电参数<sup>[8]</sup>进行工况识别。目前对基于多 源数据识别方法的研究很少,主要是利用泵功图结 合油井生产信息(产量、抽汲参数、井况数据等)进 行工况识别<sup>[9-10]</sup>。上述研究取得了较好成效,但仍 存在4个方面的问题:一是受阻尼系数、"除零"问 题影响,通过模型计算得到的泵功图和电功图存在 精度误差,影响了对特征参数值的精确计算:二是对 于机电液耦合复杂系统,同一现象有时可能由不同 原因造成,用单一信息源判断油井工况易产生误报 现象:三是受先期油井海量实时数据采集和存储技 术限制、井况复杂多变、生产统计数据不可靠等因素 影响,原有的基于多源信息工况识别的模型鲁棒性 较差。另外,传统的神经网络、支持向量机等方法主 要采用特征连接方式进行识别,对海量多源数据的 智能信息处理已难以再取得理想效果[11];四是多数 工况识别方法需要大量标记工况训练样本,而实际 工程中标记工况样本获取很难。无标记样本训练的 工况识别方法又极大浪费专家资源。此外,多数工 况特征提取方法计算量大且对特征数据之间的非相 关性要求较高。这些都影响了已有工况识别方法的 实际应用。大数据和油气生产物联网环境下,抽油 机井采油生产系统采集和存储了海量多源实时信 息,如实测的地面示功图、电参数、井口温度等,也获 取了海量未知工况样本。多视角学习方法能够有效 地解决海量而又复杂目标问题,突破传统单一特征 学习精确度低的局限性,充分利用目标本身的多特 征性质,各特征之间相互补充完善,提高学习算法性 能。研究表明,通过学习多视角特征和流形信息可 以进一步提高识别结果[12-13]。与传统信息处理方 法不同,半监督学习(semi-supervised learning,SSL) 方法大多基于流形假设,其中 Hessian 局部线性特 征映射算法具有一个明显的优势:当标记样本较少 时,其正则化可以极大改善半监督分类结果<sup>[14]</sup>。基

于此,针对大数据下抽油机井采油生产系统特点及 其工况识别研究目前所存在的问题,笔者有效利用 大数据下抽油机井采油生产系统的海量多源实时信 息,突破单一信息源识别的局限性和传统特征连接 多源识别方法的技术瓶颈,提出一种基于 Hessian 正则化多视角学习的抽油机井工况识别新方法,并 建立抽油机井工况识别模型,以进一步提高抽油机 井工况识别精准率。

# 1 Hessian 正则化多视角学习方法

#### 1.1 Hessian 正则化思想

在半监督学习 SSL 中, Hessian 正则化<sup>[15-17]</sup>反映 了样本数据分布流形的高阶信息, 能够更准确地描 述数据的内在局部几何特征。Hessian 正则化算法 既可以很好地匹配训练样本区域内数据, 也能较好 地预测区域外数据。利用 Hessian 正则化可以在标 记样本较少下极大改善半监督分类结果。

在 SSL 中, 假设有 l 个标记样本  $L = \{(\mathbf{x}_{i}^{1}, \mathbf{x}_{i}^{2}, ..., \mathbf{x}_{i}^{V}\mathbf{y}_{i})\}_{i=1}^{l}, u$  个 未 标 记 样 本  $U = \{(\mathbf{x}_{i}^{1}, \mathbf{x}_{i}^{2}, ..., \mathbf{x}_{i}^{V})\}_{i=l+1}^{l+u}, V$  代表视角数,  $x_{i}$  代表第 i 个样本,  $x^{k}, k \in \{1, 2, ..., V\}$  代表第 k 个视角的特征,  $\mathbf{x}_{i}^{k}$  代表第 i 个样本的第 k 个视角的特征矢量,  $y_{i} \in \{-1, 1\}$  代表样本  $x_{i}$  的标记。标记样本概率记为 p, 无标记样本边缘概率记为 $p_{x}, p_{x}$  构成一个紧流形 M, 在 M 上条件分布概率 p(y|x) 光滑变化。由于相邻 两个样本  $x_{i}$  和  $x_{j}$  之间要有相似的条件分布  $p(y_{i}|x_{i})$  和  $p(y_{j}|x_{j})$ , 因此对 M 的局部几何特性的准确研 究非常重要。Hessian 不仅有丰富的零空间,还可以 驱动分类函数按流形线性变化, 因此 Hessian 正则 化能够更好地表征数据分布的内在几何特性。

为便于描述,将本文中重要的变量符号列表说明,如表1所示。

表1 重要的变量符号

Table 1 List of important notations

符号	描述	符号	描述
l	标记样本数	и	未标记样本数
L	标记样本	U	未标记样本
V	视角数	$x^k$	第 k 个视角的特征
$x_i$	第 i 个样本	$\boldsymbol{x}_i^k$	第 i 个样本的第 k 个视角的特征矢量
$y_i$	样本 $x_i$ 的标记	K	多视角核
$\ m{f}\ _I^2$	f的惩罚项	Kk	核的第 k 个视角
$\ m{f}\ _k^2$	分类复杂度惩罚项	$\theta^k$	第 k 个核的权重
$\gamma_I$	f  ; 的参数	H	多视角 Hessian
$\boldsymbol{\gamma}_k$	$  f  _k^2$ 的参数	H <sup>j</sup>	Hessian 的第 j 个视角
$\gamma_{eta}$	β的平衡参数	$\beta^{j}$	Hessian 的第 j 个视角的权重
$\gamma_{\theta}$	θ 的平衡参数	$H_k$	再生核希尔伯特空间(RKHS)

#### 1.2 Hessian 正则化多视角学习方法

对于半监督分类问题,其目标函数可表示为

$$\min_{f \in C^{\infty}(M)} \frac{1}{l} \sum_{i=1}^{l} L_{\text{loss}}(x_i, y_i, f) + G(\|f\|).$$
(1)

式中, $C^{\infty}(M)$ 为流形 M上的平滑函数集;G(||f||)为函数惩罚项与流形惩罚项之和; $L_{loss}(x_i, y_i, f)$ 为损失函数,是区分各种分类方法的关键因素,包含误差项和正则化项两部分,表示如下:

$$J(\boldsymbol{\omega}) = \sum_{i} L_{\text{loss}}(m_i(\boldsymbol{\omega})) + \gamma R(\boldsymbol{\omega}).$$

式中, $L_{\text{loss}}(m_i(\omega))$ 为误差项; $R(\omega)$ 为正则化项。

添加 Hessian 流形正则化项后,式(1) 可优化如下:

$$\min_{f \in H_k} \frac{1}{l} \sum_{i=1}^{l} \psi(f, x_i, y_i) + \gamma_k \|f\|_k^2 + \gamma_I \|f\|_I^2.$$
(2)

式中, $\psi$  为损失函数;在一个恰当的再生核希尔伯特 空间  $H_k$ 中,  $\|f\|_k^2$  为分类复杂度惩罚项,  $\|f\|_l^2$  为流形 惩罚项, $\gamma_k$  为  $\|f\|_k^2$  的调节参数, $\gamma_l$  为  $\|f\|_l^2$  的调节参数。

添加多视角核 **K** 和多视角 Hessian 矩阵 **H** 后, 目标函数式(2)可实现多视角特征表示。

多视角核 K:假设在第 k 视角 K<sup>\*</sup> 是有效的(对称、正定)内积核, k = 1,2,…, V,则

$$K = \sum_{k=1}^{V} \theta^{k} K^{k}, \qquad (3)$$
  
s. t. 
$$\sum_{k=1}^{V} \theta^{k} = 1, \theta^{k} \ge 0, k = 1, 2, \cdots, V.$$

多视角 Hessian 矩阵 **H**: 假设 **H**<sup>'</sup> 是第 *j* 视角的 Hessian 矩阵, 且 **H**<sup>'</sup> 是半正定的, *j* = 1, 2, …, *V*, 则

$$H = \sum_{j=1}^{V} \beta^{j} H^{j}, \qquad (4)$$

t. 
$$\sum_{j=1}^{n} \beta^{j} = 1, \beta^{j} \ge 0, j = 1, 2, \dots, V.$$

添加式(3)和式(4)到式(2)中,式(2)可表示如下:

$$\min_{f \in H_{k}, \theta \in R^{v}, \beta \in R^{v}} \frac{1}{l} \sum_{i=1}^{l} \psi(f, x_{i}, y_{i}) + \gamma_{k} \sum_{k=1}^{V} \theta^{k} \|f\|_{K(k)}^{2} + \gamma_{k} \sum_{k=1}^{V} \theta^{k} \|f\|_{K(k)}^{2} + \gamma_{k} \|\theta\|_{K(k)}^{2} +$$

s. t. 
$$\sum_{j=1}^{V} \theta^{k} = 1, \theta^{k} \ge 0, k = 1, 2, \cdots, V,$$
  
 $\sum_{j=1}^{V} \beta^{j} = 1, \beta^{j} \ge 0, j = 1, 2, \cdots, V.$ 

式中,  $\|\boldsymbol{\theta}\|_{2}^{2}$  和  $\|\boldsymbol{\beta}\|_{2}^{2}$  为正则项;  $\boldsymbol{\gamma}_{\theta}, \boldsymbol{\gamma}_{\beta} \in \boldsymbol{R}^{+}$ 。

根据表示定理<sup>[18]</sup>,给定一个凸的损失函数,式 (5)的最优化解可表示为

$$f^{*} = \sum_{i=1}^{l+u} \alpha_{i} K(x_{i}, x).$$
 (6)

对于多视角核 K 正则化项和多视角 Hessian 矩

阵H 正则化项有:

$$\begin{cases} \|\boldsymbol{f}\|_{K}^{2} = \boldsymbol{f}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{f} = \sum_{k=1}^{V} \boldsymbol{\theta}^{k} \|\boldsymbol{f}\|_{K(k)}^{2}, \\ \|\boldsymbol{f}\|_{I}^{2} = \boldsymbol{f}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{f} = \sum_{j=1}^{V} \boldsymbol{\beta}^{j} \|\boldsymbol{f}\|_{I(j)}^{2}. \end{cases}$$
(7)

将式(6)和式(7)代入式(5)中,则目标函数的 最优化可表示如下:

$$\min_{\substack{f \in H_k, \theta \in R^v, \beta \in R^v}} \frac{1}{l} \sum_{i=1}^{l} \psi(f, x_i, y_i) + \gamma_k \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{K} \boldsymbol{\alpha} + \\ \gamma_l \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{K} \boldsymbol{H} \boldsymbol{K} \boldsymbol{\alpha} + \gamma_{\theta} \| \boldsymbol{\theta} \|_2^2 + \gamma_{\beta} \| \boldsymbol{\beta} \|_2^2. \tag{8}$$
s. t.  $\sum_{k=1}^{V} \theta^k = 1, \theta^k \ge 0, k = 1, 2, \cdots, V,$ 
 $\sum_{j=1}^{V} \beta^j = 1, \beta^j \ge 0, j = 1, 2, \cdots, V.$ 

其中

$$\boldsymbol{K} = \sum_{k=1}^{V} \theta^{k} \boldsymbol{K}^{k}, \boldsymbol{H} = \sum_{j=1}^{V} \beta^{j} \boldsymbol{K}^{j}.$$

选择合适的损失函数并通过式(6)优化后,对 目标函数式(8)的求解就变为对系数求解,采用交 替最优化方法<sup>[19]</sup>即可求解。

# Hessian 正则化多视角 logistic 回 归抽油机井工况识别模型

#### 2.1 视角选取

目前抽油机井采油生产管理系统采集和存储了 海量的多源实时信息,如实测的地面示功图、电参 数、井口温度、井口压力等。这些多源实时信息可以 从地面、井筒、地层全方位地及时反映抽油机井工 况。实测地面示功图主要反映抽油机井泵及地下情 况<sup>[20]</sup>,实测电功率信号可以反映抽油机井泵及地下情 况<sup>[20]</sup>,实测电功率信号可以反映抽油机井和面及地 下情况<sup>[78]</sup>,二者基本上可以全面反映抽油机井地 面、井筒及地层状况,但还有少量抽油机井工况不但 示功图形状相似且电功图特点类似<sup>[7,20]</sup>,如连抽带 喷和油管底部严重漏失,可以借助井口温度等实时 信息进行准确识别。连抽带喷工况下,井温上升,井 压下降,所有漏失工况下,井温和井压都下降,通过 井温可以准确识别出这些工况。

综上所述,选择实测地面示功图信号、实测电功 率信号和实测井口温度3个信息源作为特征视角。

#### 2.2 视角特征提取

对目前抽油机井工况识别研究中实测地面示功 图和实测电参数进行特征提取时所存在的问题进行 分析,为更好地提高工况识别精准度和适合工程实 际应用,充分利用先验知识和专家经验,结合机制分 析进行特征提取。

理论示功图如图1所示,其中横坐标代表位移, 用*S*表示,纵坐标代表载荷,用*P*表示。*S*。为活塞冲

· 157 ·

程(有效冲程), $P_1$ 为活塞上液柱重量,A(E)点为游动凡尔关闭点,也是下死点,B点为固定凡尔打开点,C(F)点为固定凡尔关闭点,也是上死点,D点为游动凡尔打开点。





#### Fig. 1 Theoretical dynamometer card

结合理论示功图对实测地面示功图、实测电功 率、实测井口温度的特征提取分析与计算如下:

实测地面示功图依据抽油泵一个工作周期内冲 程、冲次、功图面积、活塞上液柱重量、最大和最小载 荷、有效冲程及冲程损失的变化进行特征提取,其特 征参数值精准计算可以通过对实测示功图上泵凡尔 开闭点位置的精确提取实现。

实测地面示功图的特征数据有冲程、冲次、示功 图实际面积、最大和最小载荷、最大最小载荷比、活 塞上液注重量、有效冲程、加载和卸载冲程损失、提 前加载和卸载位置12个特征参数,具体计算如下:

冲程、冲次、最大和最小载荷可以从实时采集 的示功图数据中直接获得;功图实际面积等于示 功图采集点所围成的封闭曲线面积;最大最小载 荷比等于最大载荷与最小载荷的比值;活塞上液 柱重量等于最大载荷与最小载荷之差;有效冲程 等于游动凡尔打开点与游动凡尔关闭点之间的位 移差;加载冲程损失等于固定凡尔打开点与游动 凡尔关闭点之间的位移差;卸载冲程损失等于固 定凡尔关闭点与游动凡尔打开点之间的位移差; 提前加载位置等于从游动凡尔关闭点开始到游动 凡尔打开点之间斜率正负反向的第一个点的位 移;提前卸载位置等于从固定凡尔关闭点开始到 固定凡尔打开点之间斜率正负反向的第一个点的

实测电功率信号依据"功特征"和曲线下面积 "AUC(area under curve)特征"<sup>[78]</sup>进行特征提取, 其特征参数值精准计算可以通过在实测示功图上精 确提取上、下死点位置实现。实测电功率信号的特 征数据有上行功、下行功、周期功、上行面积、下行面 积、周期面积、平衡率(上、下行程做功比)7个特征 参数.具体计算如下: 从实时采集的示功图数据中可以直接得到上死 点和下死点,上死点为位移最大点,下死点为位移最 小点,即功图数据的起始点。上行功等于上冲程 (从下死点到上死点)时间段内所作的功;下行功等 于下冲程(从上死点到下死点)时间段内所作的功; 周期功等于上行功与下行功之和;上行面积等于上 冲程时间段内电功率信号曲线与时间水平轴所围曲 线面积;下行面积等于下冲程时间段内电功率信号 曲线与时间水平轴所围曲线面积;周期面积等于上 行面积与下行面积之和;平衡率等于上行功与下行 功的比值。

目前对实测井口温度的特征提取没有任何参考 文献。受地面、地层等环境变化因素及数据采集精 度影响,现场采集到的井口温度数据大多并不严格 遵循对应工况特征状态规律,但从本质上可以反映 出对应工况每个冲程的热能耗损。实测井口温度信 号特征参数值精准计算可以通过对实测示功图精确 提取上、下死点位置实现。实测井口温度信号的特 征数据有上行热能(温度)耗损、下行热能(温度) 耗损、周期热能(温度)耗损3个特征参数,具体计 算如下:

一个冲程内井口温度实时采集点个数通常少于 示功图实时采集点个数,可以采用插值拟合方法将 二者实时采集点同步,同时从实时采集的示功图数 据中得到上死点和下死点。上行热能(温度)耗损 等于上冲程时间段内所耗损的热能(温度);下行热 能(温度)耗损等于下冲程时间段内所耗损的热能 (温度);周期热能(温度)耗损等于上行热能(温 度)耗损与下行热能(温度)耗损之和。

### 2.3 工况识别模型建立

基于 Hessian 正则化多视角 logistic 回归抽油机 井工况识别模型建立流程如图 2 所示,主要包括损 失函数选择和算法求解两大部分。

#### 2.3.1 损失函数选择

损失函数是区别各个分类方法的核心因素,损失 函数越小,模型鲁棒性越好,而且损失函数尽量是一 个凸函数,便于收敛计算。常用的损失函数有 Gold Standard loss(理想状态)、Hinge loss(用于 SVM)、 Log loss(用于 logistic 回归)、Squared loss(用于线性 回归)、Exponential loss(用于推进法)。其中,Log 损 失函数采用对数损失函数,具有平滑性、对异常点敏 感、可预测概率和适用于大数据实验等优势<sup>[21]</sup>。

本文中采用 logistic 损失 log( $1 + e^{-f}$ ) 作为损失 函数,等同于交叉熵损失函数,用于 logistic 回归。



#### 图 2 Hessian 正则化多视角 logistic 回归抽油机井工况识别模型建立流程

# Fig. 2 Modeling flow chart of working condition recognition of sucker-rod pumping wells based on multi-view

and Hessian regularization of logistic regression

#### 2.3.2 算法求解

结合 logistic 的损失函数和上述所提方法,本文 中采用 Hessian 正则化多视角 logistic 回归算法建 模。

将式(6)代入 logistic 的损失函数中,损失函数 的最优化可表示为

 $log(1 + e^{-y_i K(x_i,x)\alpha}).$  (9) 将式(9) 代入式(8) 中,则目标函数的最优化可表 示为

$$\min_{\substack{f \in H_k, \theta \in \mathbf{R}^v, \beta \in \mathbf{R}^v}} \frac{1}{l} \sum_{i=1}^{l} \left( \log(1 + e^{-y_i K(x_i, x)\alpha}) \right) + \gamma_k \alpha^T K \alpha + \gamma_l \alpha^T K H K \alpha + \gamma_\theta \| \boldsymbol{\theta} \|_2^2 + \gamma_\beta \| \boldsymbol{\beta} \|_2^2. \tag{10}$$
s. t.  $\sum_{i=1}^{V} \theta^k = 1, \theta^k \ge 0, k = 1, 2, \cdots, V,$ 

$$\sum_{j=1}^{V} \beta^{j} = 1, \beta^{j} \ge 0, j = 1, 2, \cdots, V,$$
  

$$\ddagger \Psi \mathbf{K} = \sum_{k=1}^{V} \theta^{k} \mathbf{K}^{k}, \mathbf{H} = \sum_{j=1}^{V} \beta^{j} \mathbf{H}^{j}_{\circ}$$

Hessian 正则化多视角 logistic 回归算法(式 (10)) 求解步骤如下:

(1) 固定 θ 和 β, 求解 α。固定 θ 和 β 后, 式
 (10) 可表示为

$$\min_{\boldsymbol{\alpha} \in \mathbf{R}^{l+u}} \frac{1}{l} \sum_{i=1}^{l} \left( \log(1 + e^{-y_i \boldsymbol{K}(x_i, x)\boldsymbol{\alpha}}) \right) + \gamma_k \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{K} \boldsymbol{\alpha} + \gamma_l \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{K} \boldsymbol{H} \boldsymbol{K} \boldsymbol{\alpha}.$$
(11)

其中

$$\boldsymbol{K} = \sum_{k=1}^{V} \theta^{k} \boldsymbol{K}^{k}, \boldsymbol{H} = \sum_{j=1}^{V} \beta^{j} \boldsymbol{H}^{j}.$$

logistic 的损失函数可微,本文中采用梯度下降 法求解式(11)。式(11)的一阶导数可表示为

$$\nabla f(\boldsymbol{\alpha}) = -\frac{\log e}{l} \sum_{i=1}^{l} \left( \frac{y_i}{1 + e^{y_i \boldsymbol{K}(x_i, x)\boldsymbol{\alpha}}} \boldsymbol{K}^{\mathrm{T}}(x_i, x) \right) + \gamma_k (\boldsymbol{K} + \boldsymbol{\kappa})$$

 $K^{\mathrm{T}}$ )**α** +  $\gamma_{I}(KHK + KHK^{\mathrm{T}})$ **α**. 梯度下降算法求解过程如下:

① 初始化 $\boldsymbol{\alpha}^{0} \in \mathbf{R}^{l+u}, \delta, d^{0} = -\nabla f(\boldsymbol{\alpha}^{0}), 0 < \varepsilon \ll$ 1, m = 0;

$$\begin{array}{l} (\underline{2}) \stackrel{\text{d}}{\rightrightarrows} \left| f(\boldsymbol{\alpha}^{m+1}) - f(\boldsymbol{\alpha}^{m}) \right| > \varepsilon \text{ fr}, \\ \boldsymbol{\alpha}^{m+1} = \boldsymbol{\alpha}^{m} + \delta d^{m}, \\ d^{m+1} = -\nabla f(\boldsymbol{\alpha}^{m+1}) + \frac{\left\| \nabla f(\boldsymbol{\alpha}^{m+1}) \right\|^{2}}{\left\| \nabla f(\boldsymbol{\alpha}^{m}) \right\|^{2}} d^{m}, \\ m = m + 1; \end{array}$$

(3)
$$\boldsymbol{\alpha}^{*} = \boldsymbol{\alpha}^{m+1}$$
。  
(2) 固定  $\boldsymbol{\alpha} \ \pi \boldsymbol{\beta}, \bar{\boldsymbol{x}}$ 解  $\boldsymbol{\theta}_{\circ}$  固定  $\boldsymbol{\alpha} \ \pi \boldsymbol{\beta} \ \boldsymbol{\beta}, \bar{\boldsymbol{x}}$   
(10) 可表示为  

$$\min_{\boldsymbol{\alpha} \in \mathbf{R}^{l+u}} \frac{1}{l} \sum_{i=1}^{l} (\log(1 + e^{-y_{i}(\sum_{k=1}^{V} \theta^{k} K^{k}(x_{i}, x))\boldsymbol{\alpha})) + \gamma_{k} \boldsymbol{\alpha}^{\mathrm{T}} (\sum_{k=1}^{V} \theta^{k} K^{k}(x_{i}, x))\boldsymbol{\alpha} + \gamma_{l} \boldsymbol{\alpha}^{\mathrm{T}} (\sum_{k=1}^{V} \theta^{k} K^{k}(x_{i}, x))\boldsymbol{H} (\sum_{k=1}^{V} \theta^{k} K^{k}(x_{i}, x))\boldsymbol{\alpha} + \gamma_{\theta} \|\boldsymbol{\theta}\|_{2}^{2}.$$
(12)  
s. t.  $\sum_{k=1}^{V} \theta^{k} = 1, \theta^{k} \ge 0, k = 1, 2, \cdots, V.$   
其中

$$\begin{aligned} \boldsymbol{H} &= \sum_{j=1}^{V} \boldsymbol{\beta}^{j} \boldsymbol{H}^{j}. \\ \vec{\mathbf{x}}(12) \quad \dot{\mathbf{b}} & \longrightarrow \boldsymbol{\beta} \boldsymbol{\beta} \boldsymbol{\mathfrak{W}} \boldsymbol{\Pi} \boldsymbol{\mathfrak{F}} \boldsymbol{\mathfrak{T}} \boldsymbol{\mathfrak{S}} \boldsymbol{\mathfrak{T}} \boldsymbol{\mathfrak{H}} \\ \nabla f(\boldsymbol{\theta}^{k}) &= -\frac{\log e}{l} \sum_{i=1}^{l} \left( \boldsymbol{K}^{k}(x_{i}, x) \boldsymbol{\alpha} \frac{\boldsymbol{y}_{i}}{1 + e^{\boldsymbol{y}_{i}(\sum_{k=1}^{V} \boldsymbol{\theta}^{k} \boldsymbol{K}^{k}(x_{i}, x)) \boldsymbol{\alpha}} \right) + \\ \boldsymbol{\gamma}_{i} \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{K}^{k} \boldsymbol{\alpha} + \boldsymbol{\gamma}_{i} \left( 2\boldsymbol{H}(\sum_{k=1}^{V} \boldsymbol{\theta}^{k} \boldsymbol{K}^{k}(x_{i}, x)) \boldsymbol{\alpha} \right)^{\mathrm{T}} \boldsymbol{K}^{k} \boldsymbol{\alpha} \end{aligned}$$

 $\gamma_{k}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{K}^{k}\boldsymbol{\alpha} + \gamma_{I} \left(2\boldsymbol{H}\left(\sum_{k=1}^{V}\theta^{k}\boldsymbol{K}^{k}(x_{i},x)\right)\boldsymbol{\alpha}\right)^{\mathrm{T}}\boldsymbol{K}^{k}\boldsymbol{\alpha} + \gamma_{\theta}\theta^{k}.$ 

梯度下降算法求解过程如下:

① 初始化 
$$\theta^{0} \in \mathbf{R}^{l+u}, \delta, 0 < \varepsilon \ll 1, m = 0, d^{k^{(0)}}$$
  
=  $-\nabla f(d^{k^{(0)}}), k = 1, 2, \cdots, V;$   
② 当  $|f(\theta^{m+1}) - f(\theta^{m})| > \varepsilon$  时,  
 $\theta^{k^{m+1}} = \theta^{k^{m}} + \delta d^{k^{m}},$   
 $d^{k^{m+1}} = -\nabla f(\theta^{k^{m+1}}) + \frac{\|\nabla f(\theta^{k^{m+1}})\|^{2}}{\|\nabla f(\theta^{k^{m}})\|^{2}} d^{k^{m}},$   
 $m = m + 1;$   
③  $\theta^{*} = \theta^{m+1}$ 。  
(3) 固定  $\alpha$  和  $\theta,$ 求解  $\beta_{\circ}$  固定  $\alpha$  和  $\theta$  后, 式  
(10) 可表示为

$$\min_{\boldsymbol{\beta} \in \mathbf{R}^{l+u}} \gamma_{l} \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{K} \left( \sum_{j=1}^{V} \boldsymbol{\beta}^{j} \boldsymbol{H}^{j} \right) \boldsymbol{K} \boldsymbol{\alpha} + \gamma_{\boldsymbol{\beta}} \| \boldsymbol{\beta} \|_{2}^{2}, \quad (13)$$
s. t. 
$$\sum_{j=1}^{V} \boldsymbol{\beta}^{j} = 1, \boldsymbol{\beta}^{j} \ge 0, j = 1, 2, \cdots, V,$$

$$\pm \boldsymbol{\mu}$$

$$\boldsymbol{K} = \sum_{k=1}^{V} \theta^{k} \boldsymbol{K}^{k}.$$

式(13)可看作是多重 Hessian 最优线性组合的 学习过程。

通过上述交替优化算法就可求得式(10)的局 部最优解。

• 159 •

具体工况识别过程:对实测地面示功图、实测电 功率和实测井口温度信号3个视角原始工况数据进 行特征提取得到3个特征集后建立工况样本库并生 成训练样本集和测试样本集(训练和测试集中每个 工况样本都包含3个特征集)。训练样本分别建立 各自的核矩阵和 Hessian 矩阵,并通过各自的权重 系数( $\theta^1 + \theta^2 + \theta^3 = 1$ , $\beta^1 + \beta^2 + \beta^3 = 1$ )生成各自的多视角 核 K 矩阵和多视角 H 矩阵,选择合适的模参( $\gamma_k$ 、  $\gamma_I$ , $\gamma_\theta$ , $\gamma_\beta$ )后通过上述方法训练样本,经过多次迭代 收敛后得到最优解集,将这个解集和测试集中测试 样本各自的多视角核 K 矩阵结合就得到测试样本 的预测值集,对预测值和实际值进行比较来分类识 别抽油机井工况。

## 3 实验结果分析

为验证本文中所提方法及模型的有效性和实用 性,利用胜利油田某区块中近60口抽油机井进行工 况识别实验,区块油藏类型是典型的高压低渗稀油 油藏。实验所用抽油机井工况样本库根据油井作业 记录建立,包含11种工况,每种工况有150个样本, 共1650个样本。11种典型工况有正常、供液不足、 抽油杆断脱、连抽带喷、泵卡、泵游动凡尔失灵、结 蜡、油管漏失、泵漏、游动凡尔漏失、固定凡尔漏失。

#### 3.1 全标记训练样本下工况识别结果对比

实验时将样本库中的1650个工况样本平均分成两部分:训练集和测试集各包含825个样本,其中每种工况各包含75个样本。

采用 logistic 回归一对多两分类器,迭代次数为 1200 次。测试集中样本总数为 825,其中正例样本 数为 75,反例样本数为 750。识别准确率的计算方 法为测试集中分类正确的样本数占样本总数的比 例,测试样本是按其产生的预测值排序后正例样本 排在最前面为截断进行分类<sup>[22]</sup>。重复做 5 次实验, 每次实验中训练集和测试集互斥且不一样。

最后,将本文中所提识别方法(Hessian 正则化 多视角 logistic 回归,mHLR)分别与单视角识别方法 (logistic 回归,LR)、Hessian 正则化单视角识别方法 (Hessian 正则化单视角 logistic 回归,HLR)、传统特 征连接多源识别方法(多特征连接 logistic 回归, mCLR)对抽油机井工况的识别结果进行对比。其 中,取5次实验结果的平均值作为最终实验结果,具 体如表2所示。

实测地面示功图、实测电功率和实测井口温度 3 个视角都采用本文中提取的特征数据。

Table 2 Comparison results of working condition recognition of sucker-rod pumping wells by different methods based on all marked training samples % 视角个数 视角名称 LR HLR mCLB mHLR 1 实测地面示功图 94.84 94.70 86.20 1 实测电功率 86.20 实测地面示功图、实测电功率 2 84 98 96 77 3 实测地面示功图、实测电功率、实测井口温度 83.47 97.23

表 2 全标记训练样本下不同方法抽油机井工况识别结果对比

#### 实验结果分析如下:

(1)与实测地面示功图识别相比,本文中所提 方法(mHLR)与单视角识别方法(LR、HLR)对抽油 机井工况的识别率分别提高约 2.4%、2.5%;与实 测电功率识别相比,两种方法的识别率均提高约 11%。

(2)与实测地面示功图和实测电功率两个视角 识别相比,本文中所提方法(mHLR)与传统特征连 接多源识别方法(mCLR)对抽油机井工况的识别率 提高约 11.8%;与实测地面示功图、实测电功率和 实测井口温度 3 个视角识别相比,两种方法的识别 率提高约 13.8%。本文中所提方法的识别结果明 显优于传统特征连接多源方法的识别结果,并且在 选择合适的特征数据下,本文中所提方法会随着特 征数据的增加识别精准率提高,而传统特征连接多 源识别方法会随着特征数据的增加识别精准率下 降。

(3)分别利用实测地面示功图信号、实测地面示功图和实测电功率信号、实测地面示功图、实测电 功率和实测井口温度信号,采用不同方法对抽油机 井工况进行识别,都可以取得较好的识别结果。

(4)本文实验中 Hessian 正则化和单视角结合 对识别率影响不大,但和多视角结合可以提高识别 率。

#### 3.2 不同标记训练样本下工况识别结果对比

在第一部分实验内容基础上,分别从单视角和 多视角中选取识别率高的识别方法进行实验。将本 文中所提识别方法(mHLR)和单视角识别方法 (HLR 和 LR)按5组不同比例标记训练样本对抽油 机井工况的识别结果进行对比。每组各重复做5次 实验,分别取其5次实验结果的平均值作为最终实验结果,具体如表3所示。

表 3 不同标记训练样本下不同方法抽油机井工况识别结果对比

Table 3 Comparison results of working condition recognition of sucker-rod pumping wells

by different methods based on different marked training samples

视角个数	视角名称	方法	10%标 记/%	30%标 记/%	50%标 记/%	70%标 记/%	100%标 记/%
1	实测地面示功图	LR	84.63	87.73	90.40	93.17	94.84
1	实测地面示功图	HLR	84.57	87.53	89.98	93.50	94.70
3	实测地面示功图、实测电功率、实测井口温度	mHLR	87.55	90.86	92.54	94.97	97.23

#### 实验结果分析如下:

(1)受地质、设备等因素影响,抽油机井工况复 杂多变,工况样本不易获取,但在油气生产物联网环 境下抽油机井可以获取大量未知工况样本。与其他 方法相比,本文中所提方法(mHLR)可以实现在少 量已知工况样本下,充分利用大量未知工况样本来 实现对抽油机井工况更精准的识别,该方法更符合 实际工程应用。

(2)在不同比例标记训练样本下,本文中所提 识别方法(mHLR)对抽油机井工况的识别结果都比 单视角识别方法(HLR 和 LR)对抽油机井工况的识 别结果准确率高,尤其是在少量(低于 30%)标记训 练样本情况下效果更明显,识别率平均提高约 3%。

(3)在低于 30% 少量标记训练样本下,本文中 所提识别方法(mHLR)在识别率的提高幅度上高于 单视角识别方法(HLR 和 LR)。

## 4 结 论

(1)提出一种基于 Hessian 正则化多视角学习的抽油机井工况识别新方法,该方法通过集成多视 角多核学习算法和多视角图集成 Hessian 学习算 法,大大提高了算法的泛化能力。实验结果表明该 方法比单一信息源识别方法、传统特征连接多源识 别方法识别精度高。

(2)利用实测地面示功图、实测电功率和实测 井口温度信号,结合先验知识、专家经验和机制分析 进行特征提取,并将 log 损失函数引入 Hessian 正则 化多视角学习方法中,建立 Hessian 正则化多视角 logistic 回归抽油机井工况识别模型,提高了模型的 鲁棒性,且在全标记和少量标记训练样本下都取得 了较好的识别效果,工程实用性强。

(3) Hessian 正则化多视角学习方法还可以推广 到螺杆泵、无杆泵采油井以及其他领域的故障诊断 与识别方面,具有较好的推广价值。

### 参考文献:

 [1] 韩国庆,吴晓东,毛凤英,等.示功图识别技术在有杆 泵工况诊断中的应用[J].石油钻采工艺,2003,25
 (5):70-74.

HAN Guoqing, WU Xiaodong, MAO Fengying, et al. Application of dynamometer card identification in diagnosis of working condition for suck rod pump[J]. Oil Drilling & Production Technology, 2003,25(5):70-74.

- [2] 陈培毅.基于抽油机实测电功率的悬点示功图仿真与 工况诊断[D].秦皇岛:燕山大学,2013.
   CHEN Peiyi. Simulation of dynamometer card and working condition diagnosis based on pumping unit's measured electric power[D]. Qinhuangdao: Yanshan University, 2013.
- [3] LI K, GAO X W, TIAN Z D, et al. Using the curve moment and the PSO-SVM method to diagnose downhole conditions of a sucker rod pumping unit [J]. Petroleum Science, 2013,10(1):73-80.
- [4] WU W, SUN W L, WEI H X. A fault diagnosis of suck rod pumping system based on wavelet packet and RBF network[J]. Advanced Materials Research, 2011 (189/ 190/191/192/193):2665-2669.
- [5] XU P, XU S J, YIN H W. Application of self-organizing competitive neural network in fault diagnosis of suck rod pumping system[J]. Journal of Petroleum Science & Engineering, 2007,58(1/2):43-48.
- [6] 梁华. 有杆抽油系统故障递阶诊断的故障识别研究
   [J]. 西南石油大学学报(自然科学版),2015,37(1):
   165-171.

LIANG Hua. Hierarchical fault diagnosis of rod pumping system based on fault distinguishing [J]. Journal of Southwest Petroleum University(Science & Technology Edition), 2015,37(1):165-171.

[7] 孙振华. 游梁式抽油机采油系统实时评价方法研究
 [D]. 青岛:中国石油大学(华东),2011.
 SUN Zhenhua. The real-time evaluation method of the beam-pumping recovery system[D]. Qingdao: China U-

niversity of Petroleum(East China), 2011.

- [8] 陈黎兴. 基于电功率数据的抽油机故障诊断方法研究
  [D]. 青岛:中国石油大学(华东),2016.
  CHEN Lixing. Electrical power data based fault diagnosis of oil pumping units[D]. Qingdao: China University of Petroleum(East China), 2016.
- [9] 王凯. 基于产生式规则系统的抽油泵故障诊断[J]. 石油勘探与开发,2010,37(1):116-120.
  WANG Kai. Fault diagnosis of rod-pumping unit based on production rules system [J]. Petroleum Exploration and Development, 2010,37(1):116-120.
- [10] LIU S, RAGHAVENDR C S, LIU Y, et al. Automatic early fault detection for rod pump systems [R]. SPE 146038, 2011.
- [11] 王凯,刘宏昭,熊俊,等.基于改进的超球支持向量机 的有杆抽油泵故障诊断研究[J].机械科学与技术, 2011,30(1):133-141.

WANG Kai, LIU Hongzhao, XIONG Jun, et al. Fault diagnosis of a rod-pumping unit based on improved hyper sphere vector machines [J]. Mechanical Science and Technology for Aerospace Engineering, 2011, 30(1): 133-141.

- [12] VEERARAGHAVAN A, CHELLAPPA R, ROY-CHO-WDHURY A K. The function space of an activity [C/OL]// 2006 IEEE Computer Society Conference on. IEEE, 2006: computer vision and pattern recognition [2017-03-20]. http://ieeexplore. ieee. org/abstract/document/1640855.
- [13] IOSIFIDIS A, TEFAS A, PITAS I. Multi-view action recognition based on action volumes, fuzzy distances and cluster discriminant analysis [J]. Signal Processing, 2013,93(6):1445-1457.

- [14] LIU W F, LIU H L, TAO D C, et al. Multiview Hessian regularized logistic regression for action recognition
   [J]. Signal Processing, 2015,110:101-107.
- [15] LIU W F, TAO D C. Multiview Hessian regularization for image annotation [J]. IEEE Transactions on Image Processing, 2013,22(7):2676-2687.
- [16] DONOHO D L, GRIMES C. Hessian eigenmaps: locally linear embedding techniques for high-dimensional data
   [J]. Proceedings of the National Academy of Sciences, 2003,100(10):5591-5596.
- [17] TAO D C, JIN L, LIU W F, et al. Hessian regularized support vector machines for mobile image annotation on the cloud[J]. IEEE Transactions on Multimedia, 2013, 15(4):833-844.
- [18] SCHÖLKOPF B, HERBRICH R, SMOLA A J. A generalized representer theorem [C/OL]//Computational Learning Theory. Berlin/Heidelberg: Springer, 2001: 416-426[2017-03-20]. http://www.umiacs.umd.edu/~hal/tmp/P139.pdf.
- [19] BEZDEK J C, HATHAWAY R J. Convergence of alternating optimization [ J ]. Neural Parallel & Scientific Computations, 2003,11(4):351-368.
- [20] 胡广杰,易斌,田宝库,等. 抽油机井实测示功图泵况 诊断分析[M]. 北京:石油工业出版社,2008:1-45.
- [21] 刘红丽. 多视角学习在智能信息处理中的若干应用研究[D]. 青岛:中国石油大学(华东),2015.
  LIU Hongli. Study on multi-view learning in intelligent information processing[D]. Qingdao: China University of Petroleum(East China), 2015.
- [22] 周志华.机器学习[M].北京:清华大学出版社, 2016:29-35.

(编辑 修荣荣)