

# 基于大维数据驱动的油气管网泄漏监控模糊决策方法

马大中<sup>1</sup> 胡旭光<sup>1</sup> 孙秋野<sup>1</sup>

**摘要** 输油管网状态量多及工艺复杂,难以建立精确的管网数学模型,为了能够实时监控管网的安全运行情况,本文提出一种基于大维数据驱动的管网泄漏监控模糊决策方法.首先利用管网现有的数据信息,在不对数据进行降维处理的情况下,从信息物理系统的角度出发,将油气管网的拓扑结构、阀门开度等管道物理数据以及压力、流量等运行信息数据结合起来对复杂管网系统建立数据驱动模型.然后基于大维随机矩阵谱理论,将得到的信息物理数据协方差矩阵谱分布及圆环率作为模糊决策的条件对管网运行情况进行判断.当管网拓扑发生动态变化时,提出的方法可以有效地解决误报率高的问题.最后通过仿真及实例的分析,可以证明所提出方法的有效性.

**关键词** 大维数据驱动,信息物理系统,谱分析,模糊决策,油气管网

**引用格式** 马大中,胡旭光,孙秋野.基于大维数据驱动的油气管网泄漏监控模糊决策方法.自动化学报,2017,43(8):1370-1382

**DOI** 10.16383/j.aas.2017.c160435

## A Large Dimensional Data-driven Fuzzy Detection Method for Oil-gas Pipeline Network Leakage

MA Da-Zhong<sup>1</sup> HU Xu-Guang<sup>1</sup> SUN Qiu-Ye<sup>1</sup>

**Abstract** It is difficult to build an exact math model for the oil-gas pipeline network because there are multiple state variables and complex processing technology. A leakage detection method using fuzzy detection based on large dimensional data for oil pipeline network is proposed, which can monitor oil pipeline network operational status in real time. Firstly, from the view-point of cyber-physical system (CPS), existing oil pipeline network data containing network topology, physical data like valve opening, and operating data like pressure and flow are used to build a data-driven model without dimensionality reduction. Then, as for fuzzy decision conditions, the eigenvalue spectrum distribution and ring law based on spectral analysis of large dimensional random matrices are used to judge the network operating status. When the topological structure of oil pipeline network is dynamically altered, the fuzzy method can solve the problem of high rate false alarm. Finally, the effectiveness of the proposed method is demonstrated through simulation and case study.

**Key words** Large dimensional data-driven, cyber-physical system (CPS), spectral analysis, fuzzy detection, oil-gas pipeline network

**Citation** Ma Da-Zhong, Hu Xu-Guang, Sun Qiu-Ye. A large dimensional data-driven fuzzy detection method for oil-gas pipeline network leakage. *Acta Automatica Sinica*, 2017, 43(8): 1370-1382

管道运输是我国流体输送的主要运输方式之一.在流体运输的过程中,由于受到自然腐蚀、碰撞及人为破坏等因素的影响,管道破裂污染环境的事件时

有发生,因此管道的泄漏监控是科研工作者致力研究的热点之一<sup>[1-2]</sup>.针对单段的管道,Feng等利用管道两端压力、流量及输差数据对管道进行模糊决策,如发生泄漏采用压力梯度法和负压波法加权对泄漏点进行定位<sup>[3]</sup>.刘金海等对流体管道进行模糊建模,对流体管道压力、流量、温度等信号进行实时分析,同时结合管道自身固有的参数及专家经验实现管道泄漏实时检测<sup>[4]</sup>.但是随着管道管网的结构更加地庞大和复杂,原有方法已出现劣势,许多科学家开始从管网的角对管道安全监控运行进行研究.Huang等将瞬变流动模型和模拟退火法结合起来用于检测管网泄漏<sup>[5]</sup>.Ozevin等提出了利用声波法对管网进行检测,并且利用相关系数法和管网结构来确定泄漏的位置<sup>[6]</sup>.Kim等对管网压力数据进行累积积分,利用曲率函数和区间估计来确定管道的泄

收稿日期 2016-05-28 录用日期 2016-09-20  
Manuscript received May 28, 2016; accepted September 20, 2016

国家自然科学基金重大项目(61627809),国家自然科学基金重点项目(61433004),国家自然科学基金(61473069, 61573094),中央高校基本科研业务费专项基金(N160404005)资助

Supported by the Major Program of National Natural Science Foundation of China (61627809), the Key Program of National Natural Science Foundation of China (61433004), National Natural Science Foundation of China (61473069, 61573094) and Fundamental Research Funds for the Central Universities (N160404005)

本文责任编辑 文成林  
Recommended by Associate Editor WEN Cheng-Lin

1. 东北大学信息科学与工程学院 沈阳 110004  
1. College of Information Science and Engineering, Northeastern University, Shenyang 110004

漏点<sup>[7]</sup>. 在管网工况平稳运行时, 上述方法都可以取得比较理想的定位精度.

目前随着管网的规模不断扩大以及多种类型设备的投入运行, 使得管网的结构复杂性和运行不确定性大大增加, 信息错综复杂. 对于复杂的管网和多变的工况而言, 很难建立精准的数学模型去进行工况的识别. 尤其是由于泵的启停、阀门的开关等引起的管道拓扑结构发生变化时, 只考虑压力、流量、输差等表征管道运行信息建模去判断管网是否发生泄漏的方法就呈现出误判率高的特点. 针对实际系统来说, 数据驱动不需要建立精确的模型, 计算复杂度低, 能够充分利用工业生产中存储的大量与工艺相关的离线和在线数据, 实现对系统的监控、诊断和决策等功能<sup>[8-10]</sup>. 因此本文采用数据驱动的方法来监控管网的运行状态.

现有管网系统需要依靠采集到的压力、流量、设备等信息进行决策控制, 并且如果管网遭受到恶意数据攻击或者通信偶然错误会导致管网信息的改变和管网的非正常运行. 为了确保输油顺利进行, 实际中会经常对泵、阀门等设备进行操作, 改变压力、流量等变量数据; 反之, 根据管道压力、流量等信息, 不同的站场会按照时间要求进行启泵、增大阀门开度等操作以便于完成输油计划. 因此设备物理量与管道运行信息之间相互影响.

信息物理系统 (Cyber-physical system, CPS) 通过计算、网络及控制技术的相互协作与融合, 实现对实际多维复杂系统信息的及时检测、可靠传输和综合处理<sup>[11]</sup>. CPS 的概念最早由研究嵌入式系统的学者提出, 近年来已被用于能源、交通等领域<sup>[11-13]</sup>. 因此为了有效地建立输油管网的物理拓扑结构和表征管网运行状态的的压力和流量等信息之间的对应关系, 本文首次提出了复杂管网的基于数据驱动的信息物理系统建模方法.

信息物理系统实现了数据资源与物理世界的紧密融合, 泵、阀门等设备状态量和压力、流量等管道运行信息分别对应 CPS 中的物理系统与信息系统. 实际运行时, 物理系统的状态为信息系统的输入, 经过传输、转换及处理后, 最终转化为决策指令, 反馈作用于物理系统. 每一单独的监控点所采集的数据点随着站场的规模大小可能达到几十个或几百个. 当组成管道监控管网的站点逐渐增多时, 所需要监控的数据的维数也会迅速地增长, 当数据量趋于无穷时数据的维数与数据量的比值不再趋于零, 而是趋于一个小于 1 的常数, 即呈现出大维数据的特点<sup>[14]</sup>. 管网系统已经成为大量物理数据和信息数据等不同数据类型交叉的大维数据系统. 由于数据的维数和数据的长度等比例地进行增长, 因此原有的针对于大数据系统的方法<sup>[15]</sup> 已经不再适用于分析

大维数据系统. Stubbs 等利用典型变量分析对工业过程进行监控<sup>[16]</sup>. 樊继聪等利用联合指标独立成分分析对故障进行检测和诊断<sup>[17]</sup>. 颜雪军等利用二维主成分分析方法解决局部特征描述向量的高维数问题<sup>[18]</sup>. 但是当数据维数大量增加即维数很大时, 采用降维处理的方式会删掉大量有效数据, 导致得出错误的结论. 因此, 为了对管网数据进行处理及运行状态分析, 在不降低数据维数的情况下, 本文采用大维随机矩阵谱理论来处理管网复杂的物理数据和信息数据, 从统计的角度研究管网的物理和信息对应关系以及数据变化的情况. 然后根据建立的模糊准则和阈值关系<sup>[19-22]</sup>, 判断油气管网的运行状态.

综上所述, 本文在建立复杂管网的信息物理系统的模型基础上, 提出基于大维数据驱动的管网泄漏监控模糊决策方法 (Fuzzy detection based on large dimensional data, FDLDD). 本文创新点在于:

1) 建立了复杂管网的信息物理系统模型, 该模型充分地考虑系统中所存在的物理数据和信息数据, 并将管网的物理拓扑结构的变化直接与信息建立对应的关系;

2) 针对复杂管网所存在的大维数据, 采用大维随机矩阵谱理论的方法进行分析, 该方法可以在不降低数据维数的前提下, 有效地发现数据中所存在的异常情况, 及时报警;

3) 提出了基于大维数据驱动的管道泄漏监控模糊决策方法. 所设计的方法不需要建立大量的模糊规则, 降低了建立模糊规则的难度. 根据所设计的模糊规则判断管道是否发生泄漏或者判断管道的信息是否受到干扰或攻击, 能够准确判断管网运行状态, 减少误报率.

本文从以下三个方面进行阐述: 第 1 节介绍管网具体情况, 将管网运行中涉及到的变量进行数据映射, 从而建立起管网信息物理模型; 第 2 节介绍判断管网运行状态的方法, 利用模糊决策对管网经过处理的数据进行状态分析, 最终得到管网的运行情况; 第 3 节从仿真及实例两方面说明本文提出的 FDLDD 方法针对管网的各种情况均能达到预期的效果.

## 1 基于数据驱动的管网信息物理系统建模

### 1.1 管道管网系统描述

管道管网是一种复杂并且具有扩展性的大型网络系统. 管道管网系统包含 3 个部分如图 1 所示:

第一个是分布式数据采集系统, 能够及时地获得管道运行状态和各种工艺参数数值, 并将这些信号转换为数字信号传输到监控室或者管网调控中心.

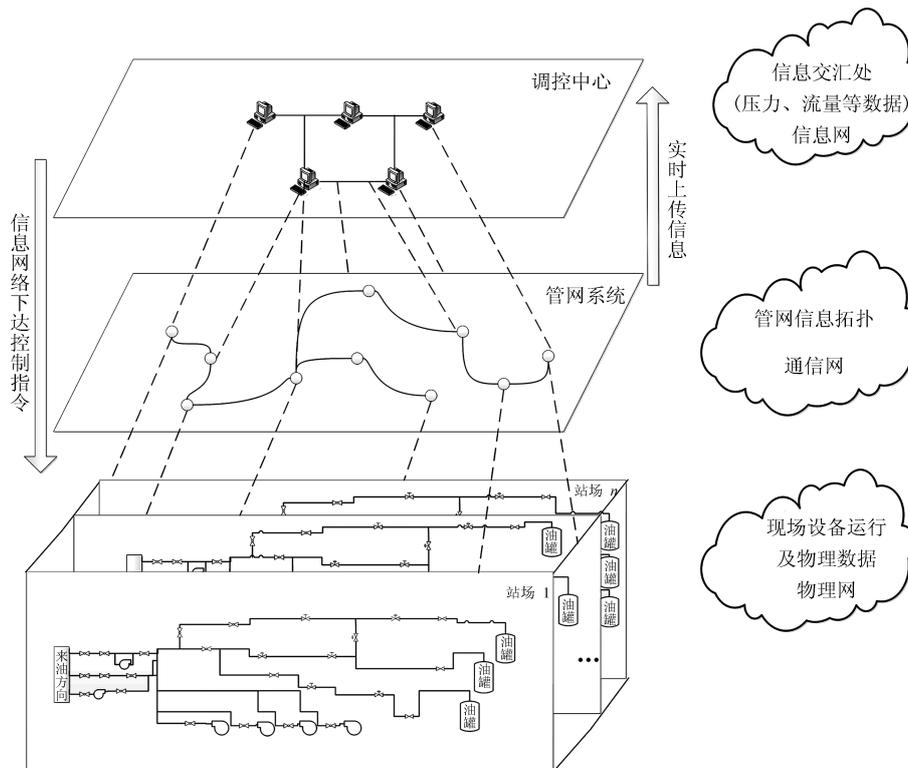


图 1 管网整体结构

Fig. 1 The whole structure of pipeline network

第二个是过程监控与管理系统, 它可以分成两个部分: 管网的调控中心和站场的监控室; 过程监控与管理系统将收集到的运行情况经过处理以图形、报表、声音等形式呈现给工作人员并且根据需求调节现场设备状态, 达到实时监控管网运行状态的目的。

第三个是数据通信网络。通信网络是管网系统中至关重要的环节, 它联系着分布式数据采集系统和过程监控与管理系统, 实现二者之间的数据通讯, 因此数据通信网络对于管网系统来说显得尤为重要。

管网系统这三个组成部分的功能不同, 但三者的有效集成构成了功能强大的管网系统, 完成对整个运输过程的有效监控。

## 1.2 管网信息物理系统建模

根据第 1.1 节的描述可知, 管网系统结构复杂, 涉及到很多变量。在保证逻辑完整性和功能完整性的前提下, 本文从数据流的角度对输油管网系统建立多数据映射的数据流模型。

建立管网信息物理系统模型的目的在于对管网的运行情况进行检测, 其中物理系统为实际存在的输送设备或者测量油品性质的设备, 例如泵、阀门、密度计等设备。信息系统是泄漏检测中起关键作用的压力、流量等数据。通过管网变量间线性和非线性

性关系, 将原本独立的管道数据进行融合, 从而组成管网的 CPS 模型。通过这种方法构造的模型, 便于进一步对管网状态进行识别和分析。

将管道管网定义为二元组:  $PN = \langle AS, PS \rangle$ 。其中  $PN$  表示管道管网,  $AS$  表示管网设备运行状态集合即管网物理系统数据集合, 也就是指对周围环境进行感知和监测得到的设备状态量或者表征油品性质的物理数据, 比如泵状态、阀门开度以及油品密度等数据;  $PS$  表示管网信息系统数据集合, 也就是指管网中影响物理集合量和对管网运行状态判断起决定性作用的信息数据, 比如压力、流量等数据。

管网中的数据通过二元组分别定义为物理数据和信息数据, 可以将管道管网中有关于泄漏检测的异构数据进行有效地分类, 使影响管道的变量和泄漏常用的评价变量区分开来, 保持不同数据类型在分析时的优势。而且能够对管网数据流有更为清晰地认知, 便于对管网建立 CPS 模型。

与此同时  $AS$  和  $PS$  之间也具有耦合关系, 彼此间关联和影响。在管网运行的过程中,  $AS$  中泵的状态、阀门的开度会直接影响  $PS$  中流量及压力的值, 反过来  $PS$  的值会对  $AS$  中如泵、阀门开度等变量值的持续时间及大小产生影响。

在管网中  $AS$ 、 $PS$  的变量值需要互相传递并将变量值传输到过程监控与管理系统, 定义如下一对

一的数据映射关系为:

$$U_1 : x_i \rightarrow y_i = I_i \cdot x_i \quad (1)$$

其中  $I_i$  是与  $x_i$  具有相同维数的单位矩阵.

管网中泵和阀门的开度会影响储油罐的液位以及流体流量, 因此定义式 (2) 来说明变量间线性关系为:

$$U_{line} : (x_1^m, x_2^m, \dots, x_{var}^m) \rightarrow y_l^m = F_l^m \cdot \left( \bigcup_{i=1}^{var} x_i^m \right) \quad (2)$$

其中  $F_l^m$  表示线性映射函数,  $\bigcup_{i=1}^n x_i^m$  表示  $x_i^m$  间的关系,  $x_i^m$  表示站场  $m$  中的变量  $i$ ,  $i = 1, 2, \dots, v, v+1, \dots, var$ , 其中前  $v$  个为  $AS$  变量,  $i = v+1, \dots, var$  为  $PS$  变量,  $y_l^m$  表示站场  $m$  中的变量  $l$ ,  $l = 1, 2, \dots, r, r+1, \dots, varn$ , 其中前  $r$  个为  $AS$  变量,  $i = r+1, \dots, varn$  为  $PS$  变量.

在站场内  $AS$  的值会影响  $PS$  的值, 因为管道物理集合信息的改变对于管道信息集合改变是非线性的, 而且不同管道间压力传播衰减也是非线性的, 定义下式来表征数据传递过程中的非线性关系:

$$U_{nonline} : (x_1^q, x_2^q, \dots, x_{var}^q) \rightarrow y_l^s = \Psi_l^q \left( \bigcup_{i=1}^{var} x_i^q \right) \quad (3)$$

其中  $\Psi_l^q$  表示非线性映射函数,  $x_i^q$  表示站场  $q$  中变量  $i$ ,  $i = 1, 2, \dots, v, v+1, \dots, var$ , 其中前  $v$  个为  $AS$  变量,  $i = v+1, \dots, var$  为  $PS$  变量,  $y_l^s$  表示站场  $s$  中变量  $l$ ,  $l = 1, 2, \dots, r, r+1, \dots, varn$ , 其中前  $r$  个为  $AS$  变量,  $i = r+1, \dots, varn$  为  $PS$  变量.

通过  $U_{line}$  和  $U_{nonline}$  建立起站场内部数据信息集合及管道间数据信息交互关系, 得到整个管网的数据信息流, 即:

$$U_{multi} : Y^p = f \left[ F_j^p \left( \bigcup_{i=1}^{varl} x_j^p \right) + \Psi_j^p \left( \bigcup_{i=1}^{varl} x_j^p \right) \right] \quad (4)$$

其中  $f$  为管网数据流算子,  $p = 1, \dots, n$ ,  $j = 1, \dots, jl, jl+1, \dots, varl$ , 前  $jl$  个为  $AS$  变量,  $j = jl+1, \dots, varl$  为  $PS$  变量.

为了能够实时通过管网模型对管网是否发生泄漏进行判断, 利用数据库中  $AS$  和  $PS$  数据, 使它们通过线性差分及式 (1)~(4) 分别构建满足一阶自回归  $AR(1)$  模型条件的大维物理集合和信息集合矩阵. 管网大维物理集合或信息集合矩阵表示为:

$$\varphi(T) = G \begin{bmatrix} Y^1(1) & \dots & Y^1(T) \\ \vdots & \ddots & \vdots \\ Y^n(1) & \dots & Y^n(T) \end{bmatrix} \quad (5)$$

其中,  $G$  为  $AR(1)$  条件算子,  $i$  为管网站场编号即  $i = 1, \dots, n$ ,  $\varphi$  为大维物理集合矩阵或大维信息集合矩阵, 即  $\varphi = AS$  或  $PS$ .

经过数据分析, 管网物理集合  $AS$  定义为多组:  $AS = \langle VA, PU, TE, DE, ST \rangle$ ,  $\Delta AS$  表示在相邻数据采集间隔内管网物理集合  $AS$  数据的差值,  $VA$ 、 $PU$ 、 $TE$ 、 $DE$ 、 $ST$  分别代表阀门状态集合、泵状态集合、温度集合、油品密度集合和储油罐液位集合, 五种管道数据信息分别表示为:  $VA = \{va_1, va_2, \dots, va_m\}$ ,  $va_m$  表示第  $m$  个阀门状态;  $PU = \{pu_1, pu_2, \dots, pu_l\}$ ,  $pu_l$  表示第  $l$  个泵的状态;  $TE = \{te_1, te_2, \dots, te_j\}$ ,  $te_j$  表示第  $j$  个站的监测温度;  $DE = \{de_1, de_2, \dots, de_p\}$ ,  $de_p$  表示第  $p$  个站管道内流体密度;  $ST = \{st_1, st_2, \dots, st_k\}$ ,  $st_k$  表示第  $k$  个储油罐液位; 定义管网信息集合  $PS = \langle P, F \rangle$ ,  $\Delta PS$  表示在相邻数据采集间隔内管网信息集合  $PS$  数据的差值, 其中  $P$  和  $F$  分别表示压力集合和流量集合,  $P = \{p_1, p_2, \dots, p_n\}$ ,  $p_n$  表示第  $n$  个站场的进站/出站压力;  $F = \{f_1, f_2, \dots, f_n\}$ ,  $f_n$  表示第  $n$  个站场的进站/出站流量. 图 2 为基于数据驱动的管道管网信息物理系统模型结构图.

## 2 基于大维数据驱动的管道管网泄漏监控

### 2.1 基于大维数据驱动的管网泄漏监控模糊决策方法

当管道管网信息物理系统模型建立后, 为了实时通过所建立的信息物理模型判断管道管网的运行状态, 根据数据间的关系将所采集的信息数据和物理数据分别构成信息集合矩阵和物理集合矩阵. 由于实时监控所采集的信息和物理数据的维数和数据量长度的比值是一个小于 1 的常数, 符合大维数据的定义<sup>[14]</sup>, 因此采用基于大维随机矩阵的谱理论对所建立的基于大维数据的管道管网的信息物理系统模型进行分析.

设管网物理集合矩阵为  $X$ , 矩阵  $X$  的任意列向量  $x_k$  满足  $AR(1)$  模型, 即  $x_t = \alpha x_{t-1} + \varepsilon_t$ , 其中  $t = 1, 2, \dots, M$ ,  $|\alpha| < 1$  是常数,  $\varepsilon_t \sim N(0, 1)$ . 它的功率谱密度函数为:

$$f_x(\lambda) = \frac{1}{2\pi} \left( \frac{1}{1 + \alpha^2 - 2\alpha \cos \lambda} \right), -\pi \leq \lambda \leq \pi \quad (6)$$

矩阵  $X$  的协方差矩阵  $S_M$  的经验谱分布为:

$$F_M(x) = \frac{1}{M} \sum_{k=1}^M I(\lambda_k \leq x) \quad (7)$$

其中  $I$  代表示性函数,  $\lambda_k, k = 1, 2, \dots, M$  为矩阵  $S_M$  的特征根.

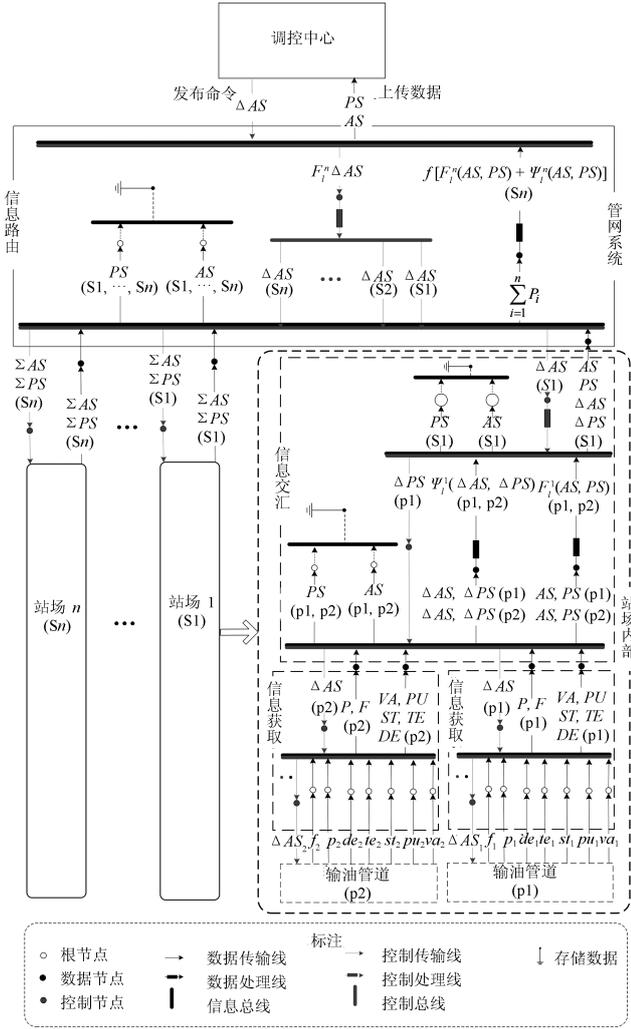


图2 基于数据驱动的管道管网信息物理系统模型

Fig. 2 The CPS model of pipeline network based on data-driven

当  $M/T \rightarrow c \in (0, 1)$  时, 协方差矩阵  $S_M$  的经验谱分布依概率 1 收敛到极限谱分布  $F(x)$ , 其密度函数  $f(x)$  为:

$$\frac{d}{dx}F(x) = \begin{cases} \frac{1}{2\pi x} \sqrt{R + 2p + 2\sqrt{\frac{q^2}{R}}}, & x \in \Omega \\ 0, & \text{其他} \end{cases} \quad (8)$$

当  $c > 1$  时,  $F(x)$  在原点有一个质量为  $1 - 1/c$  的测度. 其中:

$$\begin{aligned} u &= \max\{(1 + \alpha)^2, (1 - \alpha)^2\} \\ v &= \min\{(1 + \alpha)^2, (1 - \alpha)^2\} \\ p &= -1 - \\ &\frac{3(1 - ux)^2 + 3(1 - vx)^2 - 2(1 - ux)(1 - vx)}{8c^2} \end{aligned}$$

$$q = \frac{[2 - (u + v)x][4c^2 + (u - v)^2x^2]}{8c^3} - \frac{2}{c}$$

$$r = -\frac{(u + v)x}{(2c^2)} + \frac{-[2 - (u + v)x]^2}{256c^4} \times$$

$$[3(u - v)^2x^2 - 4(1 - ux)(1 - vx) + 16c^2]$$

$$w = 2p^3 + 27q^2 - 72pr -$$

$$\sqrt{4(-p^2 - 12r)^3 + (2p^3 + 27q^2 - 72pr)^2}$$

$$R = -\frac{2p}{3} + w + w^{1/3}/(3 \times 2^{1/3})$$

$$\Omega = \left\{ x : \frac{(1 - \sqrt{c})^2}{u} \leq x \leq \frac{(1 + \sqrt{c})^2}{v}, \right.$$

$$\left. (2p^3 + 27q^2 - 72pr)^2 > 4(p^2 + 12r)^3 \right\}$$

因为物理集合矩阵  $X$  的元素均为实数, 利用式 (9) 将矩阵  $X$  变换成标准非厄米特矩阵  $Z$ :

$$z_{i,j} = (x_{i,j} - \mu(\mathbf{x}_i)) \times \frac{\sigma(\mathbf{z}_i)}{\sigma(\mathbf{x}_i)} + \mu(\mathbf{z}_i) \quad (9)$$

其中  $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,T})$ ,  $\mu(\mathbf{z}_i) = 0$ ,  $\sigma(\mathbf{z}_i) = 1$ ,  $i = 1, 2, \dots, M$ ,  $j = 1, 2, \dots, T$ .

矩阵  $Z$  通过式 (10) 得到具有相同奇异值的矩阵  $Z_u$ :

$$Z_u = \sqrt{ZZ^T}U \quad (10)$$

其中,  $U$  为 Haar 分布的酉矩阵.

进一步, 将矩阵  $Z_u$  通过下式变换成标准矩阵  $Z_l$ :

$$z_{li} = \frac{z_{ui}}{\sqrt{M}\sigma(\mathbf{z}_{ui})} \quad (11)$$

其中  $\mathbf{z}_{li} = (z_{li,1}, z_{li,2}, \dots, z_{li,M})$ ,  $\sigma^2(\sqrt{N}\mathbf{z}_{li}) = 1$ ,  $\mathbf{z}_{ui} = (z_{ui,1}, z_{ui,2}, \dots, z_{ui,M})$ ,  $li = 1, 2, \dots, M$ ,  $ui = 1, 2, \dots, M$ .

矩阵  $Z_l$  的特征值分布于区间  $[(1 - c)^{1/2}, 1]$  中<sup>[23-24]</sup>. 区间  $[(1 - c)^{1/2}, 1]$  的物理意义在于将  $Z_l$  的所有特征值在复平面上表示出来, 特征根近似分布为一个内径为  $(1 - c)^{1/2}$ , 外径为 1 的圆环.

当物理集合矩阵  $X$  发生变化时, 协方差矩阵  $S_M$  的极限谱分布密度函数会窄于正常的情况; 矩阵  $Z_l$  中第  $i$  行第  $l - k$  个元素也会发生趋势变化或者突变, 特征值分布会变散, 并且会有特征值不包含在圆环中. 据此可知可以通过分析矩阵异常特征值分布及数量来检测管网是否发生异常.

以同样的方法可以得到管网的信息集合矩阵的协方差矩阵特征值分布.

由于管网数据形成的物理集合矩阵和信息集合矩阵经过变换得到的特征值和特征值的分布能够及时反映管网的特征, 物理集合矩阵经过谱分析得出的是关于管道管网是否发生工况调整的结论, 信息

集合矩阵得出的结论则是关于管网的运行状态是否发生变化. 将它们得出的结论作为模糊决策条件进行分析, 特点在于将原有对于变量的分析转化为对特征值的分析, 模糊决策也就转变成了对特征值相关内容的判别, 降低了建立模糊规则的难度, 减少了模糊规则数目. 而且在兼顾物理数据和信息数据的同时, 也可以消除管网频繁调整带来的数据不确定性, 解决在工况调整时频繁误报的情况.

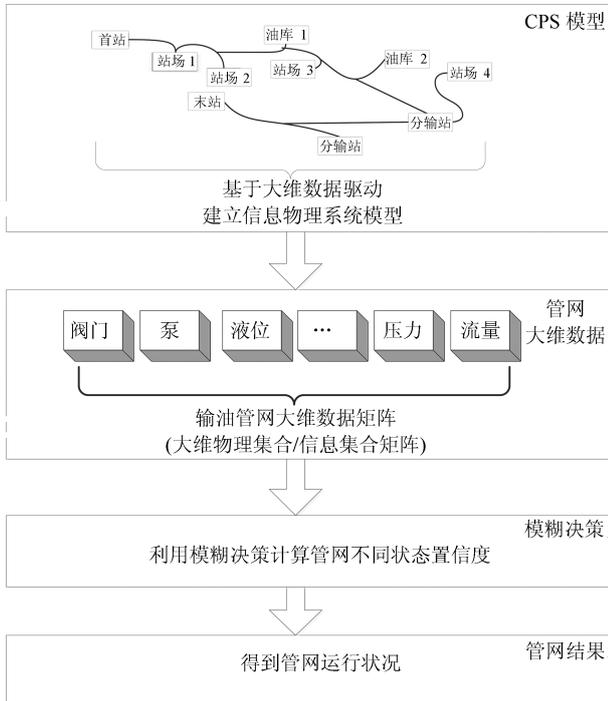


图 3 管网系统检测示意图

Fig. 3 Flowchart for leak detection of pipeline network

模糊决策条件空间为  $Cond = \{A_1, A_2, A_3, A_4\}$  表示管网异常状态表示量,  $A_1$  为物理集合矩阵异常点数量;  $A_2$  为物理集合矩阵异常特征值均值;  $A_3$  为信息集合矩阵异常点数量;  $A_4$  为信息集合矩阵异常特征值均值; 其中, 每个异常状态表示量又分成 3 个程度的模糊子集,  $A_{sub} = \{a_1, a_2, a_3\}$ .  $a_1$  表示无异常,  $a_2$  表示小异常,  $a_3$  表示大异常. 决策结果空间  $Res = \{B_1, B_2, B_3\}$ ,  $B_1$  表示正常输油,  $B_2$  表示调整工况,  $B_3$  表示产生泄漏; 其中, 每个管道运行状态量又分成 2 个程度的模糊子集,  $B_{sub} = \{b_1, b_2\}$ ,  $b_1$  表示正常/小幅调整工况/发生小泄漏量泄漏,  $b_2$  表示异常/大幅调整工况/发生大流量泄漏. 经过原理分析和大量的实际数据进行学习后, 选择三角函数和梯形函数作为物理集合矩阵和信息集合矩阵的隶属度函数.

根据现场及实际管网的情况得到一组模糊决策规则, 形成一个决策规则库. 第  $i$  条模糊决策规则  $R_i$  为

$R_i$ : if  $A_1$  is  $a_i$  and,  $\dots$ , and  $A_4$  is  $a_j$ ,  
then  $B_1$  is  $b_k$  and,  $\dots$ , and  $B_3$  is  $b_l$ .

其中  $i \in (1, 2, 3)$ ,  $j \in (1, 2, 3)$ ,  $k, l \in (1, 2)$ , 分别表示管网中物理集合、信息集合的状态及管网的运行状态.

模糊决策规则前提中的变量  $A_i (i = 1, 2, 3, 4)$  均会影响决策空间的结果  $B_j (j = 1, 2, 3)$ , 即改变  $A_i$  会使  $B_j$  发生改变, 这种影响用测度矩阵  $\delta_j$  来表示:

$$\delta_j = \begin{bmatrix} \delta_{j11} & \delta_{j12} \\ \vdots & \vdots \\ \delta_{ji1} & \delta_{ji2} \end{bmatrix} \quad (12)$$

因为每个前提变量对于结果变量的影响不同, 因此定义权重向量  $\gamma_j$  来表示  $A_i$  对  $B_j$  的影响程度:

$$\gamma_j = (\gamma_{j1}, \dots, \gamma_{ji}) \quad (13)$$

对于最后决策空间的结果, 定义置信度矩阵来决定具体结果:

$$C_j = \gamma_j \times \delta_j \quad (14)$$

当置信度满足下式时, 则认为管道运行的情况为  $B_j$ :

$$C_{jk} > \lambda_{jk} \quad (15)$$

其中  $\lambda_{jk}$  为给定的置信度阈值, 可以根据具体情况进行调整, 本文取  $\lambda_{jk} = 0.9$ ,  $j$  为模糊决策结果,  $k$  为模糊决策结果的子集元素.

## 2.2 管网运行情况检测步骤

管网异常情况发生时, 数据的时间序列会发生改变, 导致原来的模型参数不再适合于异常发生后的模型, 其协方差矩阵的极限谱分布密度函数的曲线形状及根据式 (9)~(11) 变换得到的标准矩阵在特征值复平面上的分布将会发生改变. 因此, 基于数据驱动的管道管网泄漏监控模糊决策方法的步骤如下, 具体检测示意图如图 3 所示:

- 1) 基于管网运行状况, 建立 CPS 模型;
- 2) 基于管网实时数据, 建立管网大维物理集合矩阵和大维信息集合矩阵;
- 3) 根据得到的矩阵计算管网状态置信度;
- 4) 根据置信度, 得到当前管网的运行情况.

## 3 仿真及模拟分析

### 3.1 仿真实验

为了说明 FDLDD 方法中关键参数的选择及验证该方法的有效性, 本文采用五个算例来进行说明, 算例 1 来阐述维度比的选择, 算例 2~4 利用某管网的运行数据及测试数据对不同的情况进行分析, 并

将结果与实际运行和测试情况进行比较, 算例 5 进行与模糊分类方法的对比分析, 通过以上算例验证提出方法的有效性. 由于信息集合矩阵的变化不明显, 因此在 FDLDD 方法中采用自适应放大来突出数据的变化.

**算例 1. 选择维度比参数**

选取某管网泄漏测试的数据, 并且以信息集合矩阵为例说明维度比对检测结果的影响, 具体结果如表 1 所示. 表中, 均值差比值为  $(MSR_N - MSR_A)/MSR_N$ .

从表 1 可以看出矩阵维度比在 0.5 附近时, 管网均值差比值是最明显的, 因此在选择维度比时应当选择在 0.5 附近比较好.

算例 2~4 均采用下述的信息物理系统模型: 输油管网总共有 10 个站场, 按照管网信息物理系统进行建模, 根据算例 1 中的维度比选取准则选取 10 个站场内与管道管网运行状态相关的泵、下载阀、调节阀、温度、密度等物理数据 100 个, 数据队列长度为 250, 即维度比为  $c_1 = 0.4$  的物理集合矩阵, 选取

管网内 10 个站场的进/出站的压力和流量等信息数据 24 个, 数据队列长度为 50, 即维度比为  $c_2 = 0.48$  的信息集合矩阵.

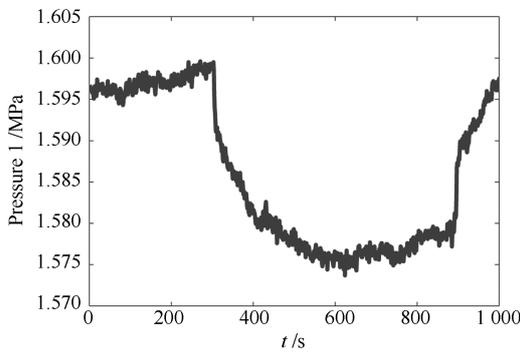
表 1 不同维度对比结果

Table 1 The results of different dimension ratio

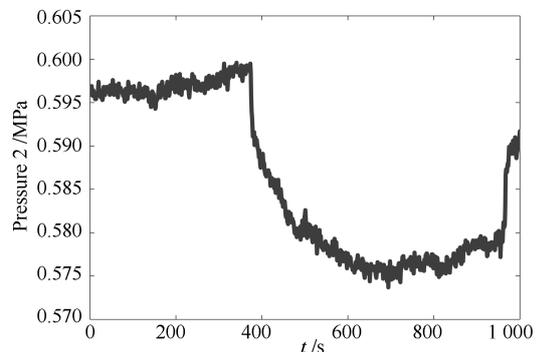
维度比	正常情况	异常情况	均值差比值
	特征值均值 $MSR_N$	特征值均值 $MSR_A$	
0.1	0.97	0.83	0.144
0.3	0.92	0.78	0.152
0.5	0.88	0.71	0.193
0.7	0.84	0.69	0.179
0.9	0.77	0.64	0.169

**算例 2. 正常调整工况**

选取该管网的正常输油数据, 图 4 为管网中某管道进行正常工况调整时产生的压力曲线, 该管道流量、泵、阀门等曲线和本管网其余站场的相关曲线未在本文中给出. 图 5 为管网 CPS 模型矩阵的特



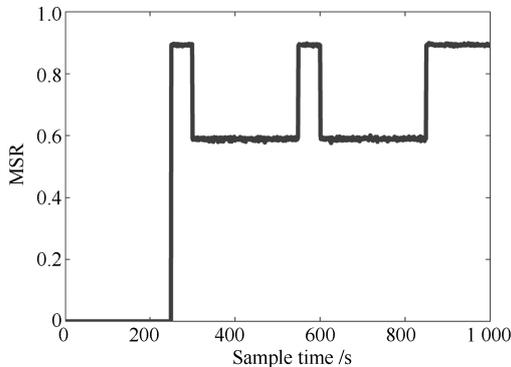
(a) 上游压力图  
(a) Inlet pressure



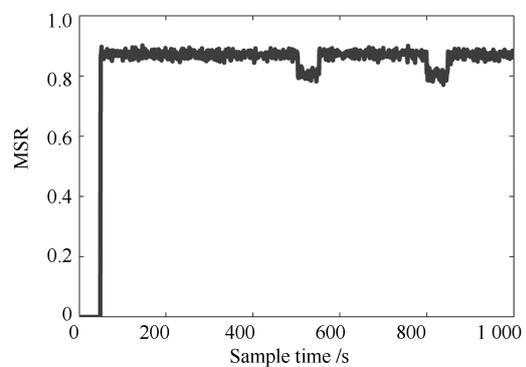
(b) 下游压力图  
(b) Outlet pressure

图 4 上下游压力曲线图

Fig. 4 The curve of pipeline pressure



(a) 物理集合矩阵 MSR 序列  
(a) MSR series of physical



(b) 信息集合矩阵 MSR 序列  
(b) MSR series of cyber

图 5 管网 MSR 时间序列

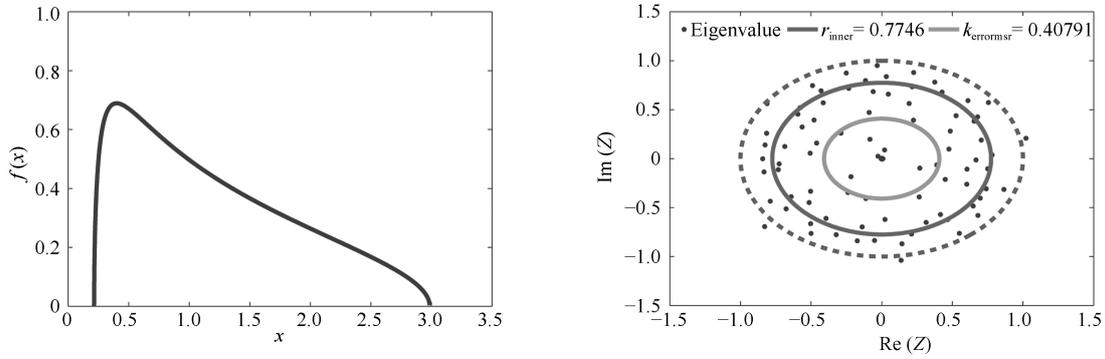
Fig. 5 MSR series of pipeline network

征值均值 (Mean spectral radius, MSR) 图, 从图 5 (a) 可以看到  $T = 304$  时物理集合矩阵的 MSR 开始下降, 说明此刻泵、阀门等状态发生改变;  $T = 503$  时由于工况调整引起的负压波已经影响到了其他站场, 使得管网的信息集合矩阵 MSR 开始下降, 如图 5 (b) 所示.

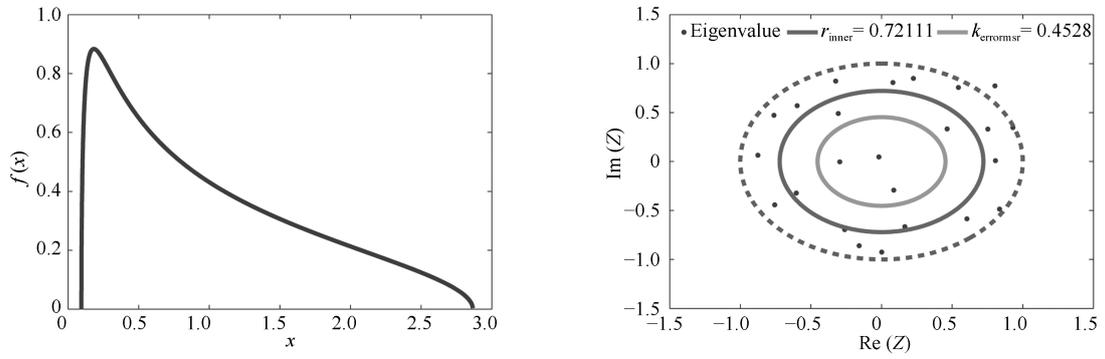
利用管网数据分别对两个矩阵进行预处理后画出其在  $T = 503$  时的具体谱分布曲线及复平面上的圆环曲线, 如图 6 所示. 图中  $k_{\text{errormsr}}$  表示异常

特征值均值, 即小于圆环内径  $r_{\text{inner}}$  的矩阵特征值均值.

从图 6 可以看到在  $T = 503$  时, 两个矩阵均出现异常特征值, 即有特征值不包含在极限谱分布函数的包络中. 物理集合矩阵的异常特征值均值为 0.40791, 数量为 47; 信息集合矩阵的异常特征值均值为 0.4528, 数量为 7, 经过模糊规则后,  $B_1$  中  $b_1$  的隶属度为 0.92,  $B_2$  中  $b_1$  的隶属度为 0.93, 因此认为管网运行正常, 处于小幅工况调整的状态,



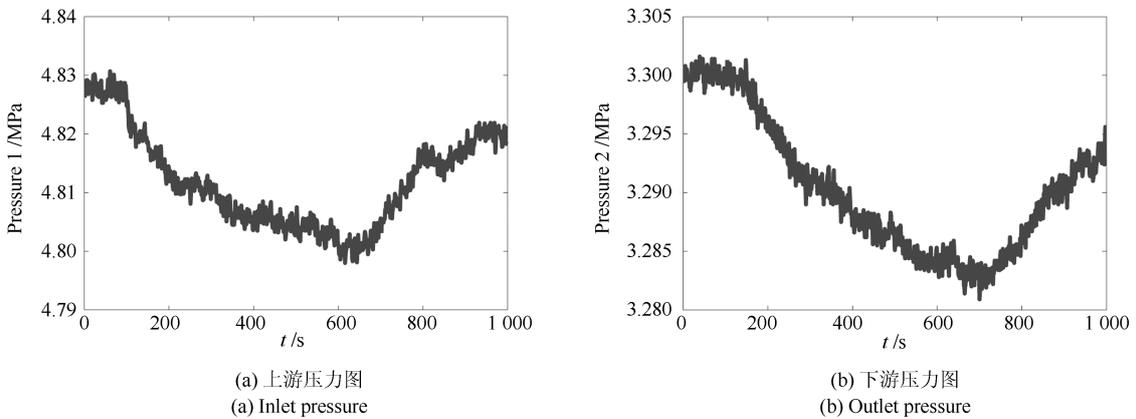
(a) 物理集合矩阵谱分布及圆环  
(a) Spectral distribution and ring law of physical matrix



(b) 信息集合矩阵谱分布及圆环  
(b) Spectral distribution and ring law of cyber matrix

图 6 管网矩阵谱分布及圆环

Fig. 6 Spectral distribution and ring law of pipeline network matrix



(a) 上游压力图  
(a) Inlet pressure

(b) 下游压力图  
(b) Outlet pressure

图 7 上下游压力曲线图

Fig. 7 The curve of pipeline pressure

系统会显示此时状态为小幅工况调整. 分析结果与实际中管网有工况调整的情况相符.

**算例 3. 测试大流量泄漏**

选取该管网的测试数据, 图 7 为测试阀室的上下游站场压力曲线, 该管道流量、泵、阀门等曲线和本管网其余站场的相关曲线未在本文中给出. 测试期间没有改变任何设备状态, 但是在  $T = 300$  时信

息集合矩阵 MSR 变小, 说明有异常情况出现, 如图 8 所示.

按照第 2.3 节中的检测步骤, 分别对两个矩阵进行预处理后画出  $T = 300$  时谱分布曲线及复平面上的圆环曲线, 如图 9 所示. 图 9(a) 中物理集合矩阵谱分布及圆环未发生改变, 即管网中不存在工况调整, 而此时信息集合矩阵特征值却发生变化, 物理

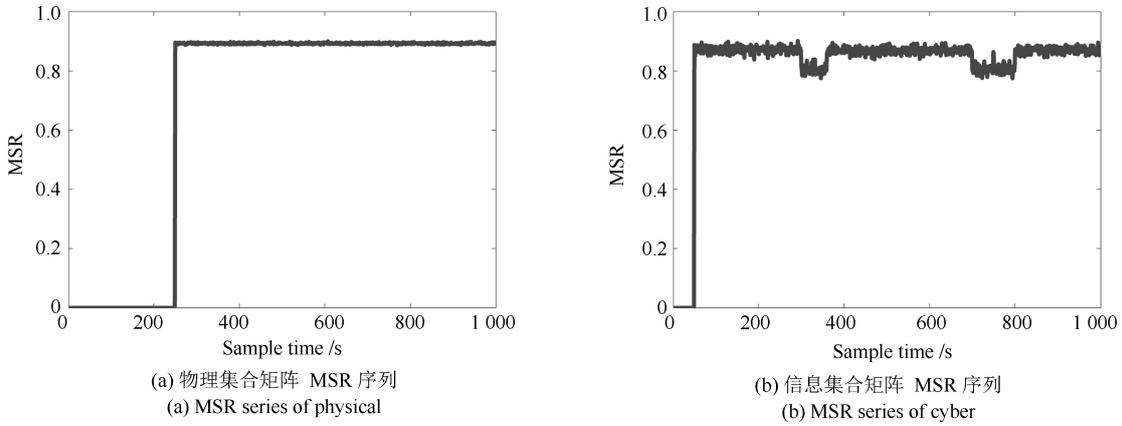


图 8 管网 MSR 时间序列  
Fig. 8 MSR series of pipeline network

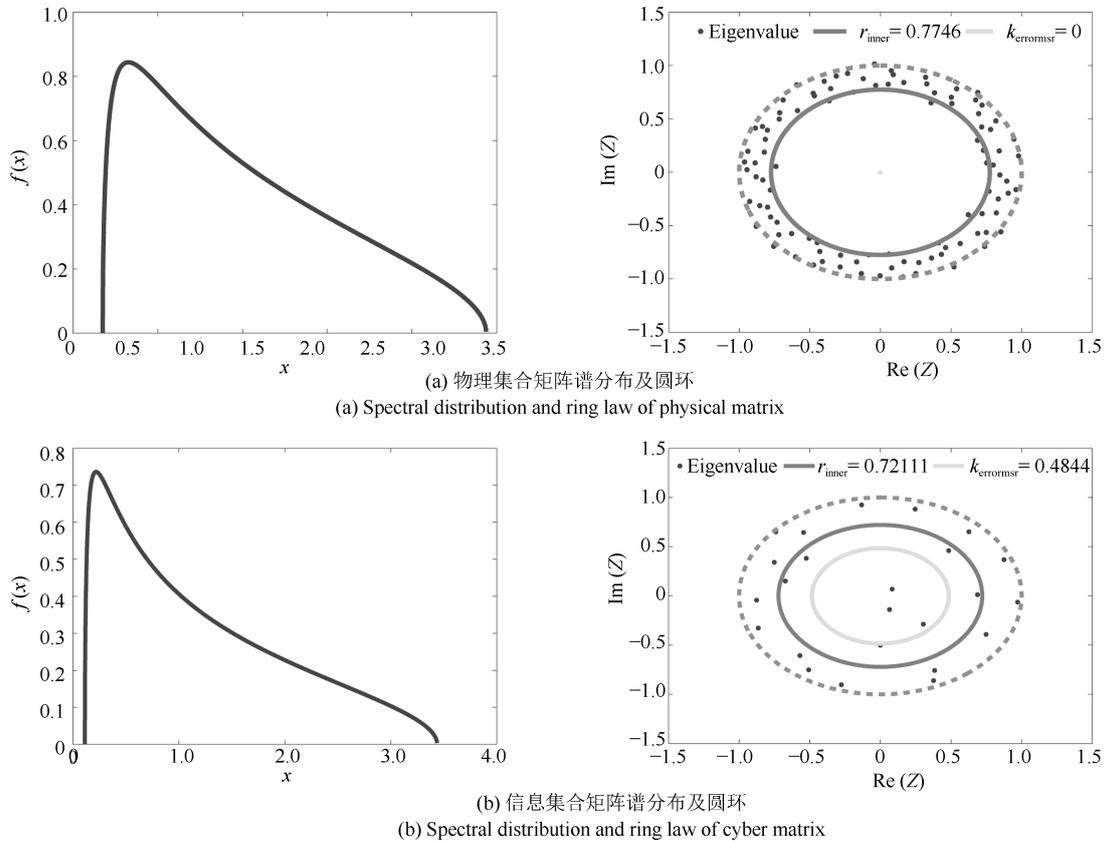


图 9 管网矩阵谱分布及圆环

Fig. 9 Spectral distribution and ring law of pipeline network matrix

集合矩阵的异常特征值均值为 0, 数量为 0; 信息集合矩阵的异常特征值均值为 0.4844, 数量为 8, 经过模糊规则后,  $B_1$  中  $b_2$  的隶属度为 0.94,  $B_3$  中  $b_2$  的隶属度为 0.93, 因此经过模糊决策认为管网存在异常, 发生大范围泄漏。

**算例 4. 测试小泄漏量泄漏**

算例 4 测试了 FDLDD 方法对管道缓慢泄漏的效果, 数据来源于该管网的测试数据, 和算例 3 属于同一次测试。图 10 为测试阀室的上下游站场压力曲线, 该管道流量、泵、阀门等曲线和本管网其余站场的相关曲线未在本文中给出。测试期间没有改变任何设备状态, 但是在  $T = 511$  时信息集合矩阵 MSR 变小, 说明有异常情况出现, 如图 11 所示。

按照第 2.3 节中的检测步骤, 分别对两个矩阵进行预处理后画出  $T = 511$  时谱分布曲线及复平面上的圆环曲线, 如图 12 所示。

图 12 (a) 中物理集合矩阵的异常特征值均值为 0, 数量为 0; 信息集合矩阵的异常特征值均值为 0.54729, 数量为 6, 经过模糊规则及隶属度函数判

断,  $B_1$  中  $b_2$  的隶属度为 0.91,  $B_3$  中  $b_1$  的隶属度为 0.92, 因此经过模糊决策认为管网存在异常, 发生小泄漏量泄漏。

**算例 5. 与模糊分类的方法对比**

文中提出的方法将物理数据和信息数据结合进行分析判断, 并与文献 [4] 所提出的基于模糊分类的管道检测方法进行对比。考虑到不同管道管网的情况, 分别选取了具有不同站场数量的管道管网系统的历史数据进行仿真, 具体对比结果如图 13 所示。

图 13 (a) 的纵坐标为检测方法的实时计算响应时间。在管网站点数比较少的时候, 两种方法的响应时间相差不多。但是随着站点数的增加, 管网中每个管道都需要进行模糊分类的判断, 检测所需要的时间也随之呈现出快速增长的趋势, 而 FDLDD 方法从全局出发, 只是在之前站场数判断方法的基础上增加了维数, 所以响应时间增加速度比模糊分类方法的时间要慢。图 13 (b) 表明在相同条件下, FDLDD 方法比模糊分类方法的准确性要高而误报次数少。FDLDD 方法是从管网的角度来解决泄漏

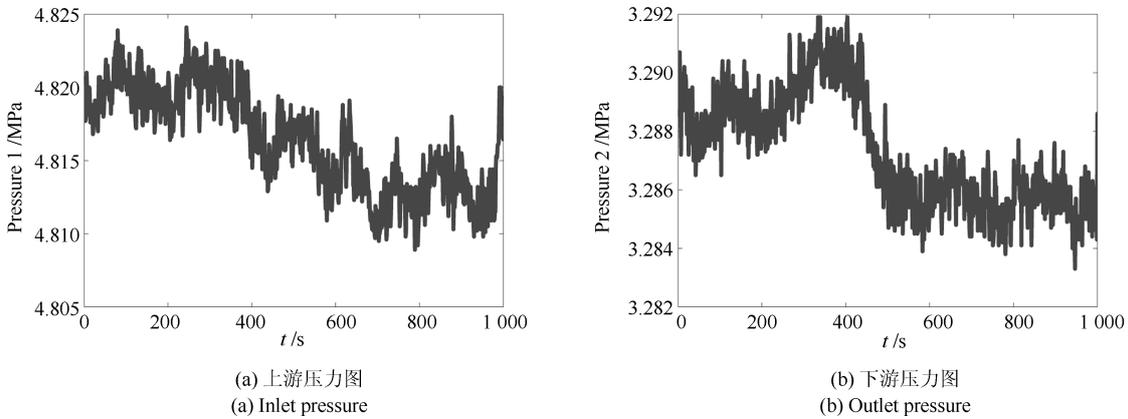


图 10 上下游压力曲线图  
Fig. 10 The curves of pipeline pressure

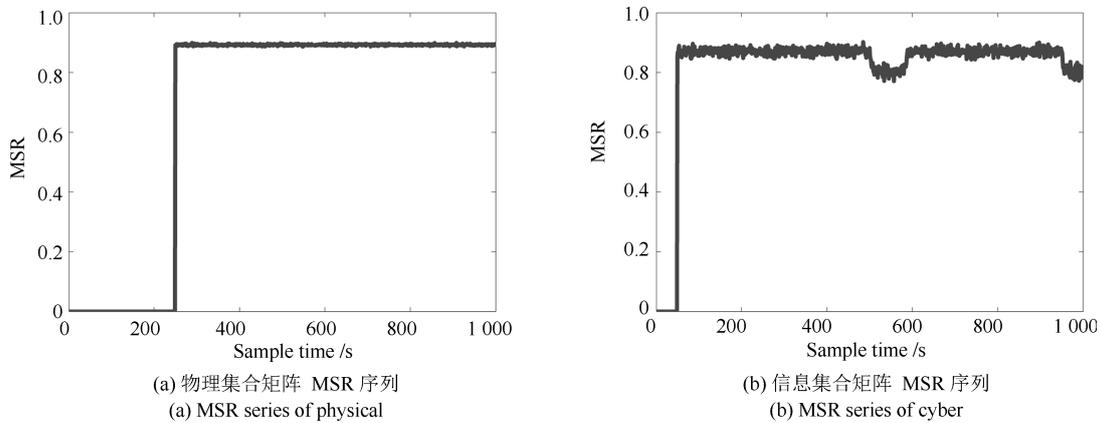


图 11 管网 MSR 时间序列  
Fig. 11 MSR series of pipeline network

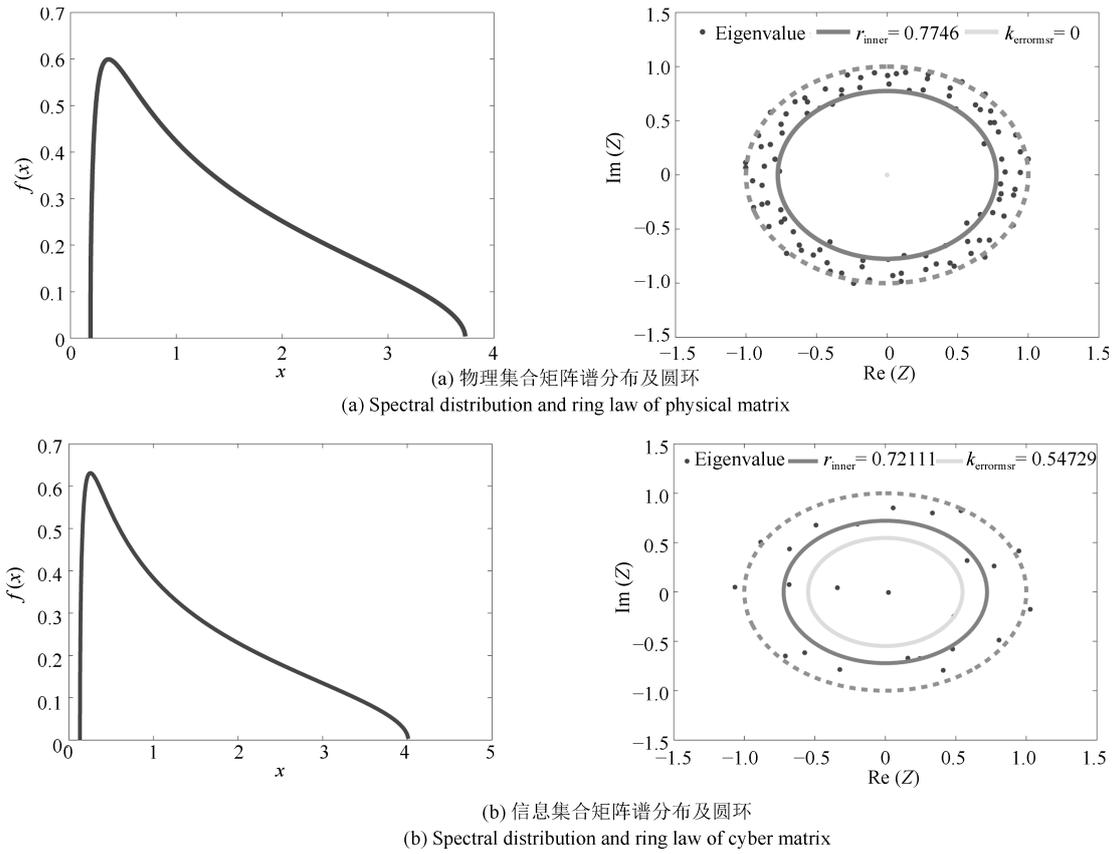


图 12 管网矩阵谱分布及圆环

Fig. 12 Spectral distribution and ring law of pipeline network matrix

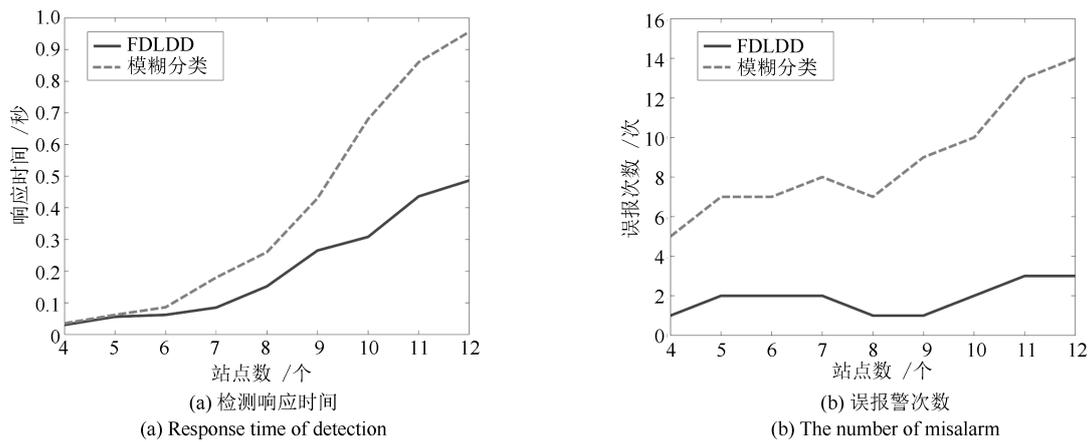


图 13 两种方法响应时间与误报警次数的对比

Fig. 13 The comparison of response time and misalarm number of the methods

检测的问题, 相较于基于单条管道的检测方法, 它能够充分利用管网的物理数据和信息数据, 采用模糊决策的方法将二者进行有效地融合, 减少由于正常工况调整操作带来的影响. 因此, 从以上分析可以看出本文提出的方法在检测响应时间和误报次数方面的性能均优于模糊分类的方法.

### 3.2 实例验证

为了能够测试本文提出方法的有效性以及检测该方法的实际效果和是否能够有效地屏蔽误报警, 选择在某管网现场进行应用测试. 图 14 为放油测试时管网的压力下降曲线图.

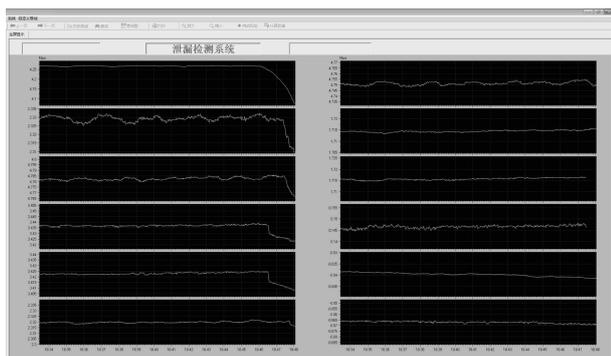


图 14 管网压力示意图

Fig. 14 The pressure of pipeline network

现场共进行了两天的测试, 一共进行泄漏模拟放油测试 19 次, 其中停输保压 3 次, 稳态单点 10 次, 全线降量 2 次, 全线增量 2 次, 稳态多点 2 次, 总共放油时长 3 小时 34 分 31 秒. 工况调整 30 次. FDLDD 方法针对上述任意一种工况都可以准确地进行报警. 在工况调整的过程中, 只有一次由于通讯网络原因所采集的物理信息不完整而出现误报.

#### 4 结论

本文针对管网的复杂工况、众多种类不同的设备和运行状态量的特点, 从大维数据驱动的角度出发对油气管网建立信息物理系统模型, 实现了整个管网泵、阀门等运行物理数据和压力、流量等信息数据的融合. 为了判断管网整体的运行情况, 通过管网的大维物理数据和信息数据矩阵的谱分布函数以及圆环变化对管网的状态进行实时监控, 并且利用模糊决策及时判断管网运行状态. 当管道管网任意一点发生泄漏时, 都可以准确报警, 并降低由于工况调整所引起的误报. 通过一系列仿真及实际测试说明在管网发生异常时, FDLDD 方法能够及时地检测异常情况并进行报警.

#### References

- 1 Wang Gui-Zeng, Ye Hao. *Leak Detection and Positioning in Fluid Transmit Pipeline*. Beijing: Tsinghua University Press, 2010.  
(王桂增, 叶昊. 流体输送管道的泄漏检测与定位. 北京: 清华大学出版社, 2010.)
- 2 Zhang Y, Chen S L, Li J, Jin S J. Leak detection monitoring system of long distance oil pipeline based on dynamic pressure transmitter. *Measurement*, 2014, **49**: 382–389
- 3 Feng J, Zhang H G. Diagnosis and localization of pipeline leak based on fuzzy decision-making method. *Acta Automatica Sinica*, 2005, **31**(3): 484–490
- 4 Liu Jin-Hai, Feng Jian. Research on leak fault intelligent detection method for fluid pipeline based on fuzzy classification. *Chinese Journal of Scientific Instrument*, 2011, **32**(1): 26–32  
(刘金海, 冯健. 基于模糊分类的流体管道泄漏故障智能检测方法研究. 仪器仪表学报, 2011, **32**(1): 26–32)
- 5 Huang Y C, Lin C C, Yeh H D. An optimization approach to leak detection in pipe networks using simulated annealing. *Water Resources Management*, 2015, **29**(11): 4185–4201
- 6 Ozevin D, Harding J. Novel leak localization in pressurized pipeline networks using acoustic emission and geometric connectivity. *International Journal of Pressure Vessels and Piping*, 2012, **92**: 63–69
- 7 Kim Y, Lee S J, Park T, Lee G, Suh J C, Lee J M. Robust leak detection and its localization using interval estimation for water distribution network. *Computers and Chemical Engineering*, 2016, **92**: 1–17
- 8 Hou Zhong-Sheng, Xu Jian-Xin. On data-driven control theory: the state of the art and perspective. *Acta Automatica Sinica*, 2009, **35**(6): 650–667  
(侯忠生, 许建新. 数据驱动控制理论及方法的回顾和展望. 自动化学报, 2009, **35**(6): 650–667)
- 9 Wang Hong, Chai Tian-You, Ding Jin-Liang, Brown M. Data driven fault diagnosis and fault tolerant control: some advances and possible new directions. *Acta Automatica Sinica*, 2009, **35**(6): 739–747  
(王宏, 柴天佑, 丁进良, 布朗马丁. 数据驱动的故障诊断与容错控制: 进展与可能的方向. 自动化学报, 2009, **35**(6): 739–747)
- 10 Dai Wei, Chai Tian-You. Data-driven optimal operational control of complex grinding processes. *Acta Automatica Sinica*, 2014, **40**(9): 2005–2014  
(代伟, 柴天佑. 数据驱动的复杂磨矿过程运行优化控制方法. 自动化学报, 2014, **40**(9): 2005–2014)
- 11 Wang Zhong-Jie, Xie Lu-Lu. Cyber-physical systems: a survey. *Acta Automatica Sinica*, 2011, **37**(10): 1157–1166  
(王中杰, 谢璐璐. 信息物理融合系统研究综述. 自动化学报, 2011, **37**(10): 1157–1166)
- 12 Wen Jing-Rong, Wu Mu-Qing, Su Jing-Fang. Cyber-physical system. *Acta Automatica Sinica*, 2012, **38**(4): 507–517  
(温景容, 武穆清, 宿景芳. 信息物理融合系统. 自动化学报, 2012, **38**(4): 507–517)
- 13 Khaitan S K, McCalley J D. Design techniques and applications of cyberphysical systems: a survey. *IEEE Systems Journal*, 2015, **9**(2): 350–365
- 14 Bai Z D, Silverstein J W. *Spectral Analysis of Large Dimensional Random Matrices* (Second edition). New York: Springer-Verlag, 2010.
- 15 Liu Qiang, Qin S J. Perspectives on big data modeling of process industries. *Acta Automatica Sinica*, 2016, **42**(2): 161–171  
(刘强, 秦泗钊. 过程工业大数据建模研究展望. 自动化学报, 2016, **42**(2): 161–171)
- 16 Stubbs S, Zhang J, Morris J. Fault detection in dynamic processes using a simplified monitoring-specific CVA state space modelling approach. *Computers and Chemical Engineering*, 2012, **41**: 77–87

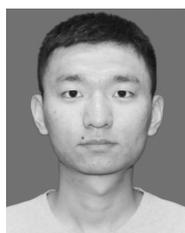
- 17 Fan Ji-Cong, Wang You-Qing, Qin S J. Combined indices for ICA and their applications to multivariate process fault diagnosis. *Acta Automatica Sinica*, 2013, **39**(5): 494–501 (樊继聪, 王友清, 秦泗钊. 联合指标独立成分分析在多变量过程故障诊断中的应用. *自动化学报*, 2013, **39**(5): 494–501)
- 18 Yan Xue-Jun, Zhao Chun-Xia, Yuan Xia. 2DPCA-SIFT: an efficient local feature descriptor. *Acta Automatica Sinica*, 2014, **40**(4): 675–682 (颜雪军, 赵春霞, 袁夏. 2DPCA-SIFT: 一种有效的局部特征描述方法. *自动化学报*, 2014, **40**(4): 675–682)
- 19 Zhang Hua-Guang, He Xi-Qin. *Theory and Application of Fuzzy Self adaptive Control*. Beijing: Beijing University of Aeronautics and Astronautics Press, 2002. (张化光, 何希勤. 模糊自适应控制理论及其应用. 北京: 北京航空航天大学出版社, 2002.)
- 20 Zhang H G, Zhang J L, Yang G H, Luo Y H. Leader-based optimal coordination control for the consensus problem of multiagent differential games via fuzzy adaptive dynamic programming. *IEEE Transactions on Fuzzy Systems*, 2015, **23**(1): 152–163
- 21 Yang F S, Zhang H G, Wang Y C. An enhanced input-delay approach to sampled-data stabilization of T-S fuzzy systems via mixed convex combination. *Nonlinear Dynamics*, 2014, **75**(3): 101–512
- 22 Yang Gui-Jun, Jiang Zhao-Hui, Gui Wei-Hua, Yang Chun-Hua, Xie Yong-Fang. Fuzzy synthesis evaluation method for position state of blast furnace cohesive zone based on entropy weight extension theory. *Acta Automatica Sinica*, 2015, **41**(1): 75–83 (杨贵军, 蒋朝辉, 桂卫华, 杨春华, 谢永芳. 基于熵权-可拓理论的高炉软熔带位置状态模糊综合评判方法. *自动化学报*, 2015, **41**(1): 75–83)
- 23 Guionnet A, Krishnapur M, Zeitouni O. The single ring theorem. *Annals of mathematics*, 2011, **174**(2): 1189–1217
- 24 Benaych-Georges F, Rochet J. Outliers in the single ring theorem. *Probability Theory and Related Fields*, 2016, **165**(1–2): 313–363



马大中 东北大学信息科学与工程学院副教授. 主要研究方向为故障诊断, 容错控制, 能源管理系统以及分布式发电系统、微网和能源互联网的优化与控制. 本文通信作者.

E-mail: madazhong@ise.neu.edu.cn

(MA Da-Zhong Associate professor at the College of Information Science and Engineering, Northeastern University. His research interest covers fault diagnosis, fault-tolerant control, energy management systems, and control and optimization of distributed generation systems, microgrids and energy internet. Corresponding author of this paper.)



胡旭光 东北大学信息科学与工程学院博士研究生. 主要研究方向为基于数据驱动故障诊断, 信息物理系统的建模及优化控制.

E-mail: 1501004@stu.neu.edu.cn

(HU Xu-Guang Ph.D. candidate at the College of Information Science and Engineering, Northeastern University.

His research interest covers fault diagnosis based on data-driven, modeling and optimal control of cyber-physical system.)



孙秋野 东北大学信息科学与工程学院教授. 主要研究方向为网络控制技术, 分布式控制技术, 分布式优化分析及其在能源互联网, 微网, 配电网等领域相关应用. E-mail: sunqiuye@mail.neu.edu.cn

(SUN Qiu-Ye Professor at the College of Information Science and Engineering, Northeastern University.

His research interest covers network control technology, distributed control technology, distributed optimization analysis and various applications in energy internet, microgrid, power distribution network.)