

# 天然气管道泄漏检测中的知识发现

王秀芳, 王 岩, 岳茂兴

(东北石油大学 电气信息工程学院, 黑龙江 大庆 163318)

**摘 要:**随着国内外天然气管道建设的迅速发展,管道的安全运行在当前尤为重要。通过对天然气管线泄漏原因的分析,采用模态声发射法和负压波法对管线进行检测,目的是获取管道检测信息,通过预处理泄露信号数据,针对数据通过预处理以取消数据差异和冗余,采取挖掘算法实现天然气潜信息挖掘,在此采用了决策树分类、DBSCN 聚类分析、K 近邻算法完成知识发现,寻找管道传输的内在联系、发展趋势及潜在规则,实现天然气传输的事前控制和智能监测。

**关键词:**知识发现;天然气管道泄漏;潜信息挖掘;智能检测

**中图分类号:**TP31.1

**文献标识码:**A

**文章编号:**1673-629X(2012)04-0137-04

## Knowledge Discovery of Gas Pipeline Leak Detection

WANG Xiu-fang, WANG Yan, YUE Mao-xing

(School of Electrical Engineering & Information, Northeast Petroleum University, Daqing 163318, China)

**Abstract:** As rapid development of foreign and domestic natural gas pipeline construction, the safe operation of pipeline is particularly important in the current. Based on analyzing the cause of gas pipeline leak and using modal acoustic emission method and negative pressure wave method to detect the pipeline, to get pipeline detection information, through the pretreatment leak signal data, according to data mining algorithm realize natural gas potential information mining. Based on that use decision tree classification, DBSCN cluster analysis, K-nearest neighbor algorithm to realize the knowledge discovery process, looking for internal relations, development trend and potential rules, controlling the natural gas transmission beforehand and intelligent monitoring.

**Key words:** knowledge discovery; natural gas pipeline leak; latent information mining; intelligent detection

## 0 引 言

随着天然气气田的不断发展,国内外天然气管道建设得到了迅速发展,天然气管道工程是一个较为庞大的、复杂的系统工程,涉及输气站场、气田、管道、储气库及其用户,任何一处出现问题,都将会影响整个系统的运行,特别一出现问题,将会影响千家万户的正常生活,因此天然气管道的安全性在当前尤为重要<sup>[1]</sup>。

天然气管道泄漏的主要原因是腐蚀、断裂/裂缝、外部影响和材料缺陷,随着时间的推移,失效机理逐渐增强。根据国内外发生天然气管道事故可得知,由于管材失效及其腐蚀原因造成的事故占绝大多数。而这些事故大多也发生在早期建设的天然气管道上,主要原因是早期的材料性能较差和建设工艺的落后。

天然气管道泄漏以后的维修、抢救等处理费用往往需要付出很高的代价。因此,重视天然气泄漏预防,

客观、准确地分析泄漏的原因,进行超前的投入,也是避免经济损失的有效途径。

一方面可以从预防提高管线的设计质量、强化运行管理入手,使输气管道不断向高等级、大口径、厚管壁方向发展,同时采用优良的防腐材料、先进的施工技术和质检标准、完整的管理体系和 HSH 管理体系,来建设和管理新管道,使得管道事故趋于减少<sup>[2]</sup>。

另一方面,可以通过以往的泄漏数据,建立泄漏信号的模型数据库,对所有泄漏信息分类存储,并从大量的泄漏数据中挖掘泄漏潜规则,建立起管道的实时监测系统,做好管道的防护措施。

## 1 天然气泄漏检测方式

针对天然气管道泄漏的特点,在此主要采用了模态声发射法和负压波法<sup>[3]</sup>两种检测方式,初步粗略的定性是否有泄漏的发生,并且采集全部的传输信号用于数据挖掘工作使用。

### 1.1 声发射法

当管道发生泄漏时,流体通过裂纹或者腐蚀孔向外喷射形成声源,然后通过管道相互作用,声源向外辐

收稿日期:2011-09-15;修回日期:2011-12-20

基金项目:黑龙江省教育科学技术重点项目(12511z002)

作者简介:王秀芳(1967-),女,河北景县人,硕士生导师,博士后在读,从事无线通信、油气信息传输与处理方向的教学和研究工作;王岩(1987-),女,硕士,研究方向为人工智能、智能检测及数据挖掘。

射能量形成声波,这就是管道泄漏声发射现象<sup>[4]</sup>。对这些因泄漏引起的声发射信号进行采集和分析处理,就可以对泄漏以及其位置进行判断。

当管道出现泄漏时,管道中的流体被扰动,接收换能器上的电压将发生明显变化。通过采集若干个泄漏点电压变化量,描绘出泄漏点与电压变化量的关系曲线,并求出曲线对应的方程。用这种方法,可以根据接收换能器上检测仪表电压的变化立即发现泄漏,进而根据拟合曲线或方程确定泄漏点的位置。根据声速在介质中传播速度的公式可知,随着液体密度的增大,其声速也将增大。例如,声波在水中的传播速度大约是在空气中的5倍。由于原油的密度比水大,因而原油中的声速远远大于在空气中的声速,所以利用超声波实时监测输油管道的运行响应速度快、灵敏度高。

## 1.2 负压波法

当管道发生泄漏事故时,在泄漏处立即有物质损失,并引起局部密度减小,进而造成压力降低。泄漏时产生的减压波被称为负压波。设置在泄漏点两端的传感器根据压力信号的变化和泄漏产生的负压波传播到上下游的时间差,就可以确定泄漏位置。该方法灵敏准确,无需建立管线的数学模型,原理简单,适用性很强。但它要求泄漏的发生是快速突发性的,对微小缓慢泄漏不是很有效。

基于负压波的传播理论,提出了两种定位方法:能够快速捕捉负压波前峰到达压力测量点的波形特征点的微分算法,并基于此种算法进行漏点定位;将极性相关引入漏点定位技术,通过确定相关函数峰值点的方法,进行漏点定位。这两种定位方法是对泄漏时的压力时间序列分别从微分和积分、从瞬态和稳态两方面进行处理,提取特征值。这两种方法配合使用,能够提高泄漏点定位的准确度。

目前,负压波法在国内输油管道上进行了多次试验,取得了令人满意的效果,但在输气管道上的试验并不多。有研究表明,负压波法完全适合于气体管道的泄漏检测,ICI公司曾经使用负压波法在乙烯管道上进行过成功的试验。使用压力波法时,应当选用只对负压波敏感的压力传感器(因为泄漏不会产生正压波),传感器应当尽量靠近管道,而且要设定合适的闭值,这样可以更好地抑制噪音。

通过上述的两种方法,可以从不同的数据源采集到大量的天然气传输应力波,负压波信号数据,从产生的有用数据、管理存储的数据中提取出人们可信的、隐藏的、事先未知的、有用的信息。进行空间挖掘,以揭示蕴含在天然气管道传输背后的本质规律,以及各个参数的内在的联系和发展趋势,来进行天然气管道泄漏的预测,使得事故减少,防护加强。

## 2 基于管道监测数据的知识发现

对天然气管道数据进行挖掘,首先要搜集海量的数据信息,其中可能存在大量的噪声、冗余数据、稀疏数据或不完全数据等。因此对搜集和提取的大量天然气管道的类型繁杂的数据量,首先要进行初步的预处理<sup>[5]</sup>。

要取消来自各个不同数据源的压力、声波等指标数据中存在的错误和不一致,进行数据值的标准化,使其具备统一的格式,也就是所谓的数据清洗<sup>[6]</sup>。对清洗后的信号数据进行集成,即来自多个管线的数据合并存放在一个统一的数据存储中,在此必须解决模式集成、冗余数据和数值冲突等方面的问题。在对传输数据内容理解的基础上,寻找数据的有用特征,以缩减数据规模,从而在尽可能保持数据原貌的前提下最大限度地精简数据量<sup>[7]</sup>,通过这样的数据选择过程,可以使得数据的规律性和现在特性更加明显。

### 2.1 决策树初步分类

对于做好初步预处理的负压波信号和声波信号数据,可以根据这两种监测方式的特点(也可理解为属性),画出初步的决策树,对其进行分类<sup>[8]</sup>。首先根据压力是否下降,可以把没有压力下降的,标记为正常;若有下降,再根据声波频率是否发生迅速的增长,来进行进一步的判断,如果有,则为异常;否则,标记为正常,如图1所示。

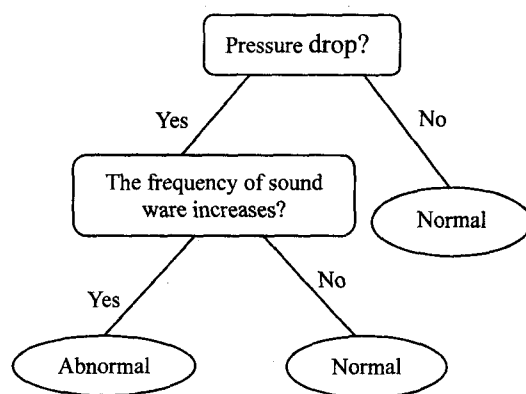


图1 分类决策树

从决策树中,可以提取到的 IF—THEN 规则是:

R1: IF Pressure drop AND the frequency of sound waves increases THEN Data is abnormal.

R2: IF Pressure drop AND the frequency of sound waves not increase THEN Data is normal.

R3: IF Pressure not drop THEN Data is normal.

对于分到两个类别中的数据信号:非正常信号(也就是所谓的泄漏信号)和正常输气信号,需要分类进行进一步的数据处理,以获取更多的挖掘资料。

### 2.2 基于 DBSCN 的聚类分析

对于分到两个类别中的数据信号,需要进行聚

类<sup>[9]</sup>分析,下面以泄漏信号为例,进行分析。

将收取到的全部泄漏信号,归为各个泄漏点,并分类标记,主要有三类包括在内:泄漏点由距离函数和指定的距离参数  $Eps$  确定的核心泄漏点,该区域内点的个数超过给定的阈值  $MinPts$ ;在核心区的某个邻域内的边界泄漏点;噪声泄漏点。接下来将噪声泄漏点删除。然后为距离在  $Eps$  内的全部核心点赋予一条边,在  $Eps$  内的核心点形成了一个簇,即可作为一种泄漏类型的判别标准。因此,可以将全部的泄漏情况标记为三类情况的发生,如图2所示。

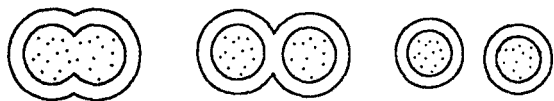


图2 相交、相接、相离的核心点

(1)核心泄漏点相交,两种泄漏类型出现交集状况,两个内圈交集为泄漏,与内圈相邻的外环,为泄漏预测区域,当检验信号达到此邻域内,就要进行预测警报,有发生泄漏的倾向。

(2)核心泄漏点相连接,两者泄漏无交集,但预测泄漏环有相交的区域。

(3)核心泄漏点相离,两者没有任何交集,这种泄漏类型分明,便于识别。

当获取未知信号若位于某类外环区时,则需对该区进行实时的警戒,有发生泄漏的可能性,在警戒区内,离核心区越近,泄露的可能性越大;反之,则越小。

### 2.3 连续信号 $K$ 近邻判别算法

持续接收获取的输气信号,依顺序先取第一个未知信号  $Sig1$ ,找到在存储数据中与该信号最邻近的  $K$  点,显然未知样本周围的点个数越少,区域也就越大;反之,则区域越小。首先要确定以  $Sig1$  为中心,以规约测定的参量  $Radius$  为半径的区域内是否可以全部包括最邻近  $K$  点,若不满足则该点归为正常信号处理<sup>[10]</sup>。在满足的前提下,  $K$  近邻有可能发生多类交集现象,此时要标记为多类点。依据所属类来判断第一个信号是否泄漏,并对  $Sig1$  进行负压波、声波及类型信息的存取。

(1)正常输气点:在以接收信号  $Sig$  为中心、 $Radius$  为半径的区域内不能包括  $K$  近邻点,归为该类,如图3所示。

(2)预泄漏点:当该信号落入泄漏预测区域,并且它的  $Radius$  区域内包括了泄漏区的  $K$  点,则该点可归为预泄漏点,要加强对该点所处的管道进行防护,有泄漏的可能性发生。

(3)泄漏点:未知信号  $Sig$  点落入了核心泄漏区(它的  $Radius$  区域内必包括了泄漏区的  $K$  点),为泄漏

点。

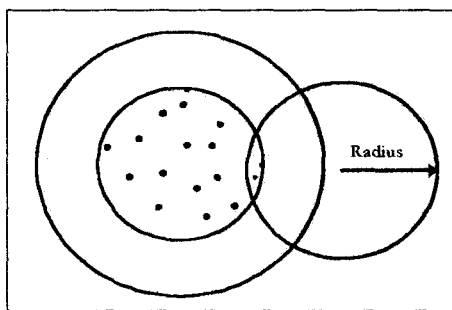


图3 未知信号  $K$  近邻判别

(4)多类预泄漏点:  $Sig$  在多个类的泄漏预测区域中(它的  $Radius$  区域内包括来自多个类的泄漏区的  $K$  点)。

(5)多类泄漏点:未知信号  $Sig$  点落入了核心泄漏区( $Radius$  区域包括多个类的  $K$  点泄漏区)。

接着对第二个传递过的信号  $Sig2$  进行分析。首先将  $Sig2$  与  $Sig1$  来进行近似度匹配,若近似度达到设定的指数,就可以将其归为与  $Sig1$  同种类型;若不足,则取出  $Sig2$ ,进行同  $Sig1$  的  $K$  近邻算法过程。接下来对第三个传过来的信号  $Sig3$  也进行相同处理,首先与  $Sig2$  进行近似匹配,如达到某设定的近似指数  $AppIndex$ ,则与  $Sig2$  归为同种类型;如果不同,再与  $Sig1$  匹配,直到有匹配上的截止,若所有匹配都不能成功,则进行  $K$  近邻算法过程<sup>[11]</sup>。以后接收的每个  $Sig$  信号都依据上述操作过程处理。

通过上述的三种挖掘算法的使用,在达到对天然气管道传输数据量进行实时监控的同时,预测及其避免一些常态泄漏情况的发生,但无法避免特殊情况的发生。根据算法的使用,需要确定的是涉及的特定参量,例如:未知信号区域半径  $Radius$ 、 $K$  值等。才能把从定性判定转为定量化检验,使泄漏监测更加精确。

## 3 实验设计

根据所获得的大庆天然气管道公司各站点的气节点压力,可知输气管道的压力的变动范围在 0.00 ~ 2.14,而站点的一般节点压力在 0.80 ~ 1.20 内变动。依据多类点压力在短时间骤升或骤降,同时声波信号出现突兀尖峰状态,便可以测定该状态为泄漏情况的发生<sup>[12]</sup>。

首先实验针对一种泄漏情况利用 Matlab 对产生的负压波信号进行模拟。并利用小波变换对原始信号进行降噪处理,产生频谱图,如图4所示。

声波信号进行小波变换后产生的频谱图,如图5所示。

通过声波信号的频谱图,可以看到一个明显的尖峰信号,它反映输气管道发生了泄露情况。



图 4 负压波频谱图

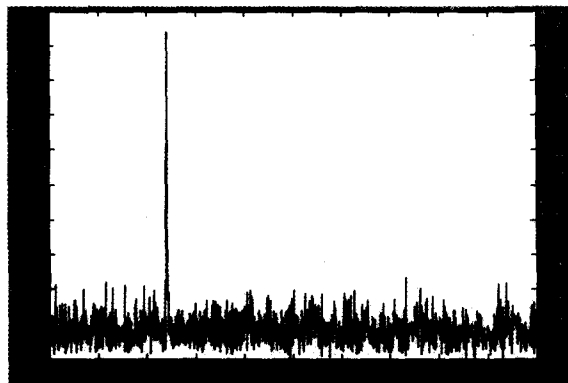


图 5 声波频谱图

由此种泄漏情况综合,规约为 DBSCN 聚类中的泄漏,其特点:当负压波频谱骤变达 0.06,声波频谱骤变 600 单位时即可判为泄漏。该泄漏点为中心泄漏点,确定泄漏区,该内部点超过给定的阈值 MinPts(此值应为实际管道实验测定),据此确定以距离参数 Eps(多聚类点的密集区)为半径形成的人为破坏泄漏簇。人为泄漏簇的外环预泄漏区的环径也是设定值 Ring。

当接收到的第一个未知信号 Sig1,可分类的具体情况如表 1 所示。

表 1 泄漏参数分类判别

NP	NP_app	AE	AE_app	泄漏位置
0.057	95%	583	97.2%	泄漏区
0.049	81.7%	583	97.2%	半泄漏区
0.057	95%	508	84.7%	半泄漏区
0.049	81.7%	508	84.7%	预泄漏区
0.041	68.3%	378	63%	未泄漏区

表注:负压波频谱骤变范围:NP(Negative pressure);负压波近似度:NP\_app;声波频谱骤变范围:AE(Acoustic emission);负压波近似度:AE\_app。

当 Sig1 的负压波和声波的近似指数均达 90% 时,该点为人为破坏泄漏簇的泄漏点;若只有一个近似指数满足条件,另一个近似指数达 70%~90% 时,可看成半泄漏点,此时已有发生泄漏的可能性,应加强此管线的测试检查工作;当两个近似指数均在 70%~90% 变化时,Sig1 落入预泄漏区,加强事前控制操作;在 0~70% 范围时为未泄漏区,近似度越小,越安全;反之,越大则越危险。

对 Sig1 近似度接近 70%,可以做进一步 K 近邻实验检测,并判断在其约定 Radius 的阈值区域半径

内,能否包含 K 泄漏点。如有,则应采取措施控制薄弱点;否则,可归为暂时安全区。依据此方法依次接收其他未知信号进行检测。

#### 4 知识发现过程的评价

针对天然气管道实现数据挖掘,可以提取出管道信息中隐藏的、事先未知的、有价值的信息量,还可以揭示蕴含在天然气管道中的客观本质规律,以及各个参数的内在的联系和发展趋势,实现对天然气管道泄漏的预测,减少事故的发生几率,缩减了经济的损失。

但是数据挖掘的基础是海量数据,这也是一个算法实现的关键,来自多源数据量的搜集也是算法实现的困难之处<sup>[13]</sup>。数据挖掘过程需要多次扫描数据库<sup>[14]</sup>。这也就造成算法有个致命的性能瓶颈:数据量太多,算法必须耗费大量的时间处理;多次扫描数据库,需要很大的 I/O 负载,在时间、空间上都需要付出很大的代价<sup>[15]</sup>。

#### 5 结束语

文中对天然气管道的泄漏原因及其检测泄漏的方式进行简要概述,接着对知识发现和实验验证过程有了细致清晰的介绍,并对该挖掘算法进行了评价。天然气数据挖掘过程设计了由浅入深的挖掘规则:先由决策树进行数据量的初步分类,接着对数据实现密度聚类分析,最后实现未知点的 K 近邻归类判别。输气数据挖掘如何提高挖掘效率、定量化挖掘参量和实际应用到现场等问题都将是下一步研究的热点问题。现有的算法远远不能满足人们对挖掘系统快速及时响应的需求,如何提高挖掘过程的效率、与用户进行交互生成可视化结果等等都是以后研究工作的重点和难点。

#### 参考文献:

- [1] 胡灯明,骆 晖. 国内外天然气管道事故分析[J]. 石油工业技术监督,2009(9):8-12.
- [2] 李 刚,王世泽,郭新江. 天然气常见事故预防与处理[M]. 北京:中国石化出版社,2008.
- [3] Taghvaei M, Beck S B M, Staszewski M J. Leak detection in pipelines using cepstrum analysis[J]. Measurement Science and Technology, 2006, 17:367-372.
- [4] 王效东,黄 坤,朱小华,等. 油气管道泄漏检测技术发展现状[J]. 管道技术与设备,2008(1):24-25.
- [5] Han Jiawei, Kamber M. Data Mining Concepts and Techniques[M]. 2nd ed. Beijing:China Machine Press,2008:206-209.
- [6] Tan Pangning, Steinbach M, Kumar V. Introduction to Data Mining[M]. [s. l.]:Posts & Telecom Press,2006:307-319.

(下转第 143 页)

算法,由此可得出,改进蚁群算法在解决盲均衡问题中性能有了一定的改善。

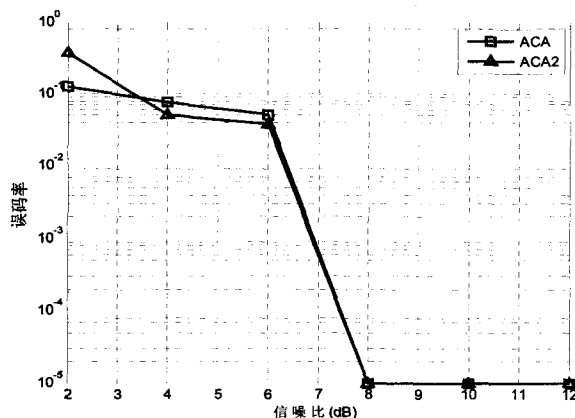


图2 改进蚁群算法和基本蚁群不同信噪比下误码率比较

试验2:不同迭代次数下代价函数曲线。  
仿真结果如图3所示。

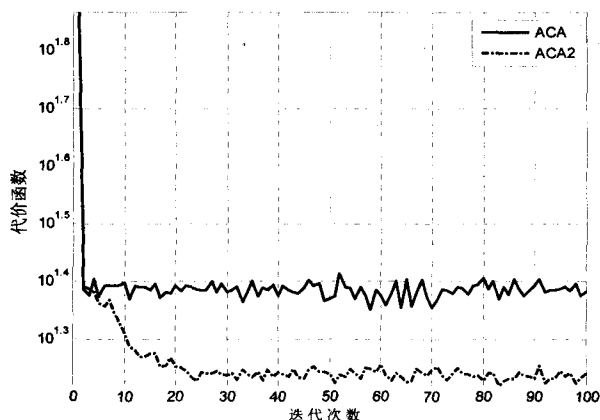


图3 代价函数平均值曲线

图3为信噪比为7dB时代价函数随迭代次数变化的曲线,从图3可以看出,随着迭代次数的增加,代价函数均在逐渐减小,但是改进蚁群算法的收敛速度明显优于基本蚁群算法,表现了良好的收敛性。

## 4 结束语

文中针对蚁群算法的结构和盲均衡问题的特点,

修正了基本蚁群算法转移概率使其得到了简化且其更容易实现并行计算。并通过仿真实验证明了这种方法的可行性。并且仿真结果表明这种简化的算法有着非常好的检测性能,它的误码率特性、收敛速度都表现了优于文献算法的结果。

## 参考文献:

- [1] Dorio M, Maniezzo V, Colomi A. Positive Feedback as a Search Strategy[R]. Dipartimento di Elettronica, Politecnico di Milano, Italy: [s. n.], 1990.
- [2] Bell J E, McMullen P R. Ant colony optimization techniques for the vehicle routing problem[J]. Advanced Engineering Informatics, 2004, 18(1): 41-48.
- [3] Wang J F, Liu J H, Zhong F Y. A novel ant colony algorithm for assembly sequence planning[J]. Int J of Advanced Manufacturing Technology, 2005, 25(11): 1137-1143.
- [4] Giannakis B, Mendel J M. Identification of nonminimum phase system using higher order statistics[J]. IEEE Transaction on Acoustic Speech and Signal Processing, 1989, 37(3): 360-377.
- [5] Godard D N. Self-recovering equalization and carrier tracking in two dimensional data communication systems[J]. IEEE Trans on Commun, 1980, 28(11): 1867-1875.
- [6] 张志涌, Bai Erwei. SIMO 含公零信道直接盲序列检测[J]. 电子学报, 2005, 33(4): 671-675.
- [7] 杨磊, 于舒娟. 基于精英策略的逆向蚁群优化盲检测算法[J]. 计算机技术与发展, 2010, 20(12): 90-93.
- [8] Gleiser P, Danon L. Community structure in jazz[J]. Advances in Complex Systems, 2003(6): 565-573.
- [9] 殷志锋, 蔡子亮, 田亚菲. 面向多用户检测的蚁群算法及其改进[J]. 计算机工程与设计, 2007, 28(7): 1511-1513.
- [10] 张元敏, 殷志锋, 周雅. 蚁群算法在多用户检测中的应用及其改进[J]. 信号处理, 2007, 23(4A): 338-342.
- [11] 段海滨. 蚁群算法原理及其应用[M]. 北京: 科学出版社, 2005.
- [12] 李士勇. 蚁群算法及其应用[M]. 哈尔滨: 哈尔滨工业大学出版社, 2004.

(上接第140页)

- [7] 张云海, 龚铃. 数据挖掘原理与技术[M]. 北京: 电子工业出版社, 2004: 50-54.
- [8] 王苗, 柴瑞敏. 一种改进的决策树分类属性选择方法[J]. 计算机工程与应用, 2010, 46(8): 127-129.
- [9] Holte R C. Very simple classification rules perform well on most commonly used datasets[J]. Machine learning, 1998, 11(1): 63-91.
- [10] 曹志宇, 张忠林. 快速查找初始聚类中心的 K-means 算法[J]. 兰州交通大学学报, 2009, 28(6): 15-18.
- [11] Huang Z. Extensions to the k-means algorithm for clustering

large data sets with categorical values[J]. Data Mining and Knowledge Discovery II, 1998, 3(2): 283-304.

- [12] 秦楠, 万江文. 基于神经网络的城市天然气管网泄漏检测融合数据研究[D]. 北京: 北京邮电大学, 2009.
- [13] 于冬, 顾培亮, 李晨光. 基于数据挖掘技术的铁路货运安全数据管理系统[J]. 中国铁道科学, 2004(4): 114-116.
- [14] 刘新亮, 姜同强, 左敏, 等. 数据仓库与数据挖掘技术在矿山信息化中的应用[J]. 金属矿山, 2009(2): 121-127.
- [15] 袁玉波, 杨传胜, 黄廷祝, 等. 数据挖掘与最优化技术及其应用[M]. 北京: 科学出版社, 2007: 100-103.