

基于支持向量机的抽油机井功图识别研究

李春生¹, 苏晓伟¹, 魏 军², 王丽丽²

(1. 东北石油大学 计算机与信息技术学院, 黑龙江 大庆 163318;
2. 克拉玛依职业技术学院 信息工程系, 新疆 克拉玛依 833600)

摘 要:抽油机的异常情况会使油田的产油效率降低,而不同的异常类型对应的抽油机示功图特征也各不相同,因此造成的损害程度也不同。针对以上问题,文中提出了一种抽油机井功图识别模型,该方法将支持向量机(SVM)用于抽油机井功图识别。首先利用改进的矢量曲线数据压缩方法(ICVDC)对抽油机井下示功图进行特征数据提取,在此基础上,采用“一对一”分类法建立基于支持向量机的井下示功图分类模型,进而对不同特征的示功图进行分类识别,并与其他识别分类模型进行了识别分类效果对比。实验结果表明,该方法分类准确度高,有效地解决了示功图的识别和分类问题,方便对油井设备等进行进一步的故障分析处理,从而大大提高抽油机的性能与效率,以此来达到油田提高采收率的目的。

关键词:示功图;支持向量机;特征提取;分类;识别

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2014)08-0215-04

doi:10.3969/j.issn.1673-629X.2014.08.051

Research on Diagrams Identification of Pumping Unit Based on Support Vector Machine

LI Chun-sheng¹, SU Xiao-wei¹, WEI Jun², WANG Li-li²

(1. College of Computer and Information Technology, Northeast Petroleum University,
Daqing 163318, China;
2. Department of Information Engineering, Karamay Vocational and Technical College,
Karamay 833600, China)

Abstract: The pumping unit anomalies will reduce oil production efficiency, while the different abnormal types corresponding to the characteristics of oil pumping machine indicator diagram also are not identical, so the damage degree is different. For above problem, put forward a well pumping unit work diagram recognition model, the method uses Support Vector Machine (SVM) for pumping unit well figure identification. First use the Improved Curve of Vector Data Compression (ICVDC) method to extract characteristics of the data of pumping unit downhole indicator diagram, on this basis, using the “one-against-one” classification to establish the downhole indicator diagram classification model based on SVM, with different features to identify the classification of the indicator diagram, and compare with other recognition classification model in classification effect. The experimental results show that, the method for the classification has high accuracy, effectively solving the problem of identification and diagnosis of the diagram, which is convenient for further analyzing and handling the fault of oil well equipment, thus greatly improving the performance and efficiency of the pumping unit, in order to achieve the purpose of oil field recovery improved.

Key words: indicator diagram; support vector machine; feature extraction; classification; identification

0 引言

目前,油田现场普遍采用的判断抽油机井工况的方法是分析泵功图的形状,而获取泵功图的主要方法是利用井上测得的悬点功图数据进行推导得到的^[1]。

通过对泵功图的分析可以判断抽油井的工作是否正常,深井泵在井下工作是否存在异常以及杆、泵参数组合与油井是否适应。但在通常情况下,对井上悬点功图进行分类判断是由巡井的工作人员根据自己的工作

收稿日期:2013-10-16

修回日期:2014-01-26

网络出版时间:2014-05-21

基金项目:黑龙江省教育科学技术研究项目(12511010)

作者简介:李春生(1960-),男,河北定州人,教授,博士生导师,研究方向为人工智能及其应用、模式识别与人工智能;苏晓伟(1987-),女,河北石家庄人,硕士研究生,研究方向为信息智能分析与处理。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20140525.1242.006.html>

经验来人工识别的,由于受多方面的限制这种方法已不能满足油田现代化生产的需要,所以先后出现了很多方法对功图进行识别。例如,网格法、直观几何法、傅里叶位置法、矩特征识别法等^[2-4],但这些方法都不能很好地描述各功图之间的细微差别。近年来人们已经开始尝试使用一些人工智能领域的方法对泵功图识别问题进行研究^[5]。利用神经网络方法进行模式识别已经取得了一些成果,但在样本容量有限的情况下,神经网络模式识别方法存在网络参数难以合理确定、容易陷入局部最小、网络学习过程容易发生振荡以及泛化能力不强等问题^[6]。

文中提出一种基于支持向量机的抽油机井功图识别方法,它可以在较少的训练样本条件下取得很强的推广能力,通过对抽油机井功图信息的学习,可以有效对功图类型做出准确的识别、分类,便于对抽油机故障做出准确的诊断,将损失降到最低。

1 多分类支持向量机

多分类支持向量机是在二分类支持向量机的基础上,对二分类支持向量机进行组合,最终将多分类问题转化为二分类的问题^[7]。目前,实现支持向量机的多分类的方法主要有:“一对一”分类法、“一对多”分类法^[8]。文中主要采用“一对一”分类法。

“一对一”分类法是将多类问题转化为两类分类问题。对于一个 k 类分类问题,其中任意两类训练样本构造一个支持向量机分类器,所有不同的两类组合共同构成 $k(k-1)/2$ 个二值子分类器,将其他的每个类别单独作为反例进行训练,从而每个类别对应 k 个子分类器,在分类过程中,对 k 个类别中的任意一个测试样本进行分类时,将每个测试样本分别输入 $k(k-1)/2$ 个子分类器中,采用投票法统计所有 $k(k-1)/2$ 个子分类器的分类结果,最后得票最多的类别即为该测试样本的类别^[9]。为区分第 i 类和第 j 类样本数据,需构造子分类器,并将属于第 i 类的样本标记为正,将属于第 j 类的样本标记为负,需求解一个形式如下的优化问题^[10]:

$$\begin{aligned} \min_{w^i, b^i, \xi_i} \quad & \frac{1}{2} (w^i)^T w^i + C \sum_i \xi_i^i \quad (1) \\ \text{s. t.} \quad & (w^i)^T \varphi(x_i) + b^i \geq 1 - \xi_i^i, \text{ if } y_i = i \\ & (w^i)^T \varphi(x_i) + b^i \geq -1 + \xi_i^i, \text{ if } y_i = j \\ & \xi_i^i \geq 0 \end{aligned}$$

从而得到一个分类决策函数,该函数的形式如下所示:

$$f_{ij}(x) = \text{sgn}((w^i)^T \varphi(x_i) + b_{ij}) \quad (2)$$

其中, $i, j = 1, 2, \dots, m$, 并且有 $i \neq j$ 。

“一对多”分类法是将训练样本为 k 个类别的多分类问题转化为 k 个两类问题。即在训练某一类别时,将该类样本标为正类,将除第 i 类以外的其他所有 $k-1$ 类样本作为负类,以此来构建出 k 个二值支持向量量子分类器^[11]。在测试阶段中,将测试样本数据分别输入 k 个子分类器中,对决策函数值进行计算,并将结果输出函数值最大(即决策函数值为正)的子分类器所对应的正类作为测试样本的所属类别。其中,对第 i 个子分类器的训练构成对应的如下优化问题^[12]:

$$\begin{aligned} \min_{w_i, b_i, \xi_i} \quad & \frac{1}{2} \|w_i\|^2 + C \sum_{j=1}^l \xi_{i,j} \quad (3) \\ \text{s. t.} \quad & (w^i)^T \varphi(x_j) + b^i \geq 1 - \xi_j^i, \text{ if } y_j = i \\ & (w^i)^T \varphi(x_j) + b^i \geq -1 + \xi_j^i, \text{ if } y_j \neq i \\ & \xi_j^i \geq 0, j = 1, 2, \dots, l \end{aligned}$$

很显然,对于 k 个类别的多分类问题,对 k 个优化问题求解,可得到以下 k 个分类决策函数:

$$\begin{cases} f_1(x) = (w_1)^T \varphi(x) + b_1 \\ \dots \\ f_m(x) = (w_m)^T \varphi(x) + b_m \end{cases} \quad (4)$$

判别任意一个测试样本 x_i 的类别 y_i 的方法是,将样本点 x_i 分别代入上述的 k 个分类决策函数中,计算得到最大分类函数值 $f_i(x)$ 的那个类,作为样本点 x_i 的类别,如下式所示^[13]:

$$\text{Class}(x) = \text{argmax}((w_i)^T \varphi(x) + b_i) \quad (5)$$

2 泵示功图分类模型的构建

2.1 泵示功图的基本类型

泵示功图也称为井下示功图,它是对抽油杆柱最下端载荷与位移变化的规律以及相互关系的描述^[14]。通常情况下,泵示功图是由悬点功图经过公式推导得到的。由于井下情况复杂,泵示功图表现出不同的形状特征,其中包括几种典型的泵功图,如图 1 所示。

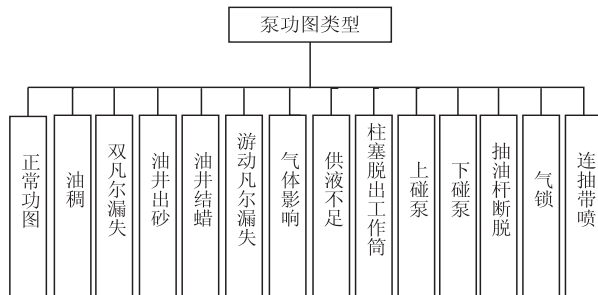


图 1 泵功图类型

2.2 功图特征提取方法

在示功图识别过程中,特征提取是一个关键问题。特征提取的过程是通过映射的方法用低维空间来表示高维空间的功图样本,映射之后的特征为功图的提取

特征^[15],对示功图提取有效特征,以便建立功图识别模型。文中是将示功图看作一条封闭的二值曲线,进而使用一种对曲线提取特征点的方法-改进的矢量曲线数据压缩方法(ICVDC)^[16],对功图曲线进行特征点提取。

ICVDC 方法是从曲线的一端开始,首先记录端点值,依次计算曲线上相邻两点之间线段的斜率,如果后一线段斜率与前一线段斜率之差的绝对值小于某个阈值,则这两个线段合为一条线段,从而在两线段连接点处的点作为冗余点被压缩;下一步再用压缩冗余点后所形成线段的斜率与下一相邻线段的斜率相减,如果结果值的绝对值大于某个阈值,则保留后一线段,并且记录两线段的交点,对该线段用同样的方法继续与下一线段计算斜率差的绝对值,并判断是否小于或大于阈值,以此方法继续求直到起始点为止,则记录的点值作为功图曲线的特征点。

如图2所示的功图曲线,首先记录 A_1 点的值,取线段 $\overline{A_1A_2}$ 的斜率

$$K_{12} = \frac{Y_2 - Y_1}{X_2 - X_1} = \frac{\Delta Y_{12}}{\Delta X_{12}} \quad (6)$$

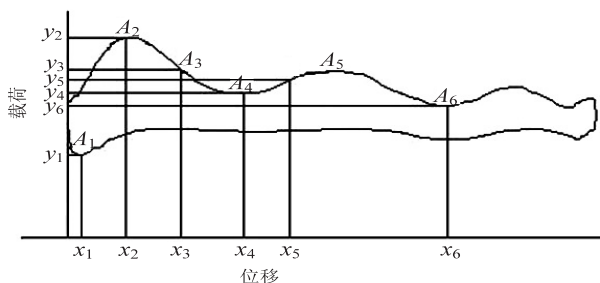


图2 改进的矢量曲线数据压缩过程

取线段 $\overline{A_2A_3}$ 的斜率

$$K_{23} = \frac{Y_3 - Y_2}{X_3 - X_2} = \frac{\Delta Y_{23}}{\Delta X_{23}} \quad (7)$$

假如取阈值为 ε , 如果有

$$K_{23} - K_{12} = \frac{\Delta Y_{23}}{\Delta X_{23}} - \frac{\Delta Y_{12}}{\Delta X_{12}} \leq \varepsilon \quad (8)$$

则以线段 $\overline{A_1A_3}$ 取代曲线 A_1A_3 , 求出线段 $\overline{A_1A_3}$ 的斜率。

若有 $K_{23} - K_{12} = \frac{\Delta Y_{23}}{\Delta X_{23}} - \frac{\Delta Y_{12}}{\Delta X_{12}} > \varepsilon$, 则以线段 $\overline{A_2A_3}$

取代曲线 A_2A_3 , 并记录 A_2 点的值, 对于曲线的其余部分用同样的方法用折线拟合, 一直到终点, 则 A_1, A_2, \dots 为功图曲线特征点。

2.3 基于支持向量机的示功图分类模型的建立

示功图分类模型建立的步骤如图3所示。

1) 数据预处理。从油厂采集功图数据, 将数据预处理, 在数据预处理阶段包括功图曲线归一化及数据

点数统一两个方面。功图曲线归一化是由于不同的井, 围成功图曲线的载荷和位移的数量级可能不同, 需要将功图曲线的载荷和位移的值转变成具有相同量纲和相同数量级的数据, 文中用到的数据归一化的方法为最大最小法, 将数据归一化到 $[0, 1]$ 范围内。由于围成功图曲线的数据点不同, 为了便于计算, 文中采用二次差值的方法, 将功图数据点统一成 360 个点。

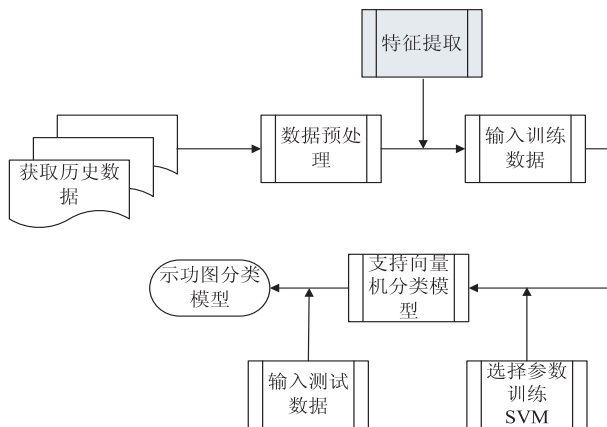


图3 示功图分类模型建立过程

2) 特征点提取。特征提取方法有很多, 比如直观几何特征提取法、网格法、矩特征识别法等, 文中采用改进的矢量曲线数据压缩方法对功图曲线提取特征点。该方法取斜率为压缩冗余点的条件, 斜率是一个相对量, 比其他以点对线段的距离为压缩冗余条件的方法更为合理。

3) 选择支持向量机的参数: 核函数和惩罚系数 c 。在支持向量机中有一个重要的概念-核函数, 核函数把高维空间里的内积运算转换为原始空间中的核函数计算, 避免了“维数灾难”的问题。目前常用的核函数主要有^[17]:

(1) 线性核函数:

$$K(x_i, x_j) = (x_i, x_j)$$

(2) 多项式核函数:

$$K(x_i, x_j) = [(x_i, x_j) + 1]^n \quad n = 1, 2, \dots$$

(3) 径向基核函数:

$$K_\gamma(x_i, x_j) = \exp\{-\gamma |x_i - x_j|^2\}$$

式中, γ 为参数。

(4) Sigmoid 核函数:

$$K(x_i, x_j) = \tanh(\gamma(x_i \cdot x_j) + c)$$

式中, $\gamma > 0, c > 0$ 。

文中选择径向基函数作为支持向量机的核函数, 径向基核函数的参数 γ 与 c 的选择文中主要采用网格搜索法, 通过对训练样本进行分组交叉验证寻找支持向量机的最优参数。采用“一对一”分类法, 将2)提取到的功图曲线的特征点作为 SVM 模型的输入, 模型的输出为功图类型, 得到支持向量机分类模型。

4)用测试样本对支持向量分类模型进行测试,最后得到示功图分类模型。

3 仿真结果分析

3.1 数据准备

训练及测试样本数据的组成结构如表 1 所示,是由 150 幅泵功图数据共同构成的。从中任取 60 组数据作为测试数据。将训练数据和测试数据转化成标准格式,即将训练数据和测试数据同时缩放至[0,1],将载荷与位移点数统一成 360 个点。

表 1 训练及测试样本

项目	数目
正常功图	5
油稠	10
双凡尔漏失	12
游动凡尔漏失	10
油井结蜡	8
油井出砂	15
气体影响	20
供液不足	8
柱塞脱出工作筒	12
上碰泵	10
下碰泵	12
抽油杆断脱	5
气锁	10
连抽带喷	13

3.2 实验结果及分析

在进行训练时,分别尝试了三种特征提取方法,在径向基核函数条件下,实验结果如表 2 所示。

表 2 泵功图分类实验结果

特征名称	训练样本数	测试样本数	惩罚因子 c	γ	归类正确数	正确识别率/%
几何特征法	90	60	0.5	2.0	50	83.333 3
不变矩特征法	90	60	0.5	2.0	52	86.666 7
ICVDC	90	60	0.5	2.0	55	91.666 7

从表 2 可以看出:使用改进的矢量曲线数据压缩方法对功图进行特征提取,准确度优于其他方法。

3.3 与其他方法的性能比较

BP 神经网络是一种按误差逆传播算法训练的多层前馈网络,是目前应用最广泛的神经网络模型之一。BP 网络能学习和存贮大量的输入-输出模式映射关系,不需要事前揭示描述这种映射关系的数学方程。它的学习规则是使用最速下降法,通过反向传播来不断调整网络的权值和阈值,使网络的误差平方和最小^[18]。但 BP 神经网络存在缺点,例如学习速度很

慢、网络训练失败的可能性较大等。

文中将相同样本下的支持向量机分类方法与 BP 神经网络分类方法做了比较,结果如表 3 所示。

表 3 支持向量机分类方法与 BP 神经网络分类方法性能比较

分类方法	测试样本	正确样本数	正确识别率/%
BP 神经网络	60	53	83.33
支持向量机	60	56	93.33

从表 3 可以看出,在相同样本情况下,支持向量机分类方法优于 BP 神经网络分类方法,分类正确率较高。由此,文中提出了基于支持向量机的抽油机井功图识别模型,克服了神经网络过学习、局部极小值等问题。

4 结束语

文中提出了基于支持向量机的抽油机井功图识别模型,采用改进的矢量曲线数据压缩方法对泵功图曲线进行特征提取,利用多分类支持向量机,建立了基于支持向量机的抽油机井功图识别模型。通过实验结果表明,文中提出的基于支持向量机的方法能更好地对功图图形做出分类识别,是一种有效的预测方法,具有很好的应用价值。

参考文献:

[1] 刘 炜. 基于支持向量机的泵功图工况识别[J]. 价值工程,2010,29(17):156-157.

[2] 刘 炜,刘宏昭. 三种泵功图特征提取方法及比较[J]. 煤炭技术,2010,29(9):202-204.

[3] 刘 炜. 基于支持向量机的抽油机示功图工况判别[D]. 西安:西安理工大学,2009.

[4] 岳继红,左维军,李洪伟. 抽油机井示功图矩特征识别方法[J]. 油气田地面工程,1999,18(3):15-16.

[5] 冯娜娜. 抽油机井泵况智能诊断方法研究[D]. 秦皇岛:燕山大学,2006.

[6] 王 凯. 基于多分类支持向量机的有杆抽油泵故障诊断研究[J]. 西安石油大学学报:自然科学版,2010,25(1):91-95.

[7] Weston J, Watkins C. Multi-class support vector machines [C]//Proceedings of ESANN99. Brussels:[s. n.],1999:41-83.

[8] 李 健,范万春,何 驰. 基于多分类支持向量机的网络入侵检测技术[J]. 计算机应用,2005,25(7):1551-1553.

[9] Burges C J C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery,1998,2(2):121-167.

[10] Wu Wei, Meng Yangyang. Comparing differect feature extraction methods of pump dynamograph based on support vector

果观察仔细便能发现,结果图片中每一个被标记的小块在源图像沙子中都能发现。

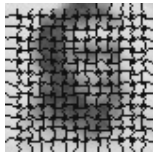
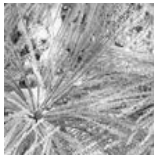


图 4 有色线条标记来自源图像的块示意图

此外每个像素的 R 、 G 、 B 分量作为传输约束的实验结果如图 5 所示。通过对比分析可得,纹理传输过程在满足纹理合成约束的基础上,能够选择的源图像和目标图像对应的传输约束度量是多样的,并不是单一的,不同的约束度量,传输得到的效果也不一样,这种传输约束度量的选择多样性,为图像风格^[14]的传输渲染提供了更多的可能。



(a) 源图像



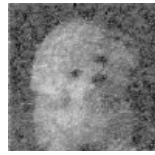
(b) 目标图像



(c) R 约束结果



(d) G 约束结果



(e) B 约束结果

图 5 实验结果图

4 结束语

文中提出了一种改进的块拼接纹理合成技术,对以往纹理合成图片在接缝处不连续和不平滑的现象进行了改进,并且将这种改进的纹理合成技术应用到纹理传输中,此外对纹理传输过程中两幅图像的传输约束度量的多样性进行了研究,相应地得到了不同风格的传输结果。由于文中改进的块拼接技术,增加了计

算量,算法运行时间有所增加,因此,接下来的研究致力于在合成较好图片的同时提高算法运行的效率,降低计算的复杂度。

参考文献:

- [1] 徐晓刚,鲍虎军,马利庄. 纹理合成技术研究[J]. 计算机研究与发展,2002,39(11):1405-1411.
- [2] 陈云飞. 基于样图的纹理合成技术研究[J]. 计算机工程与应用,2004,40(9):42-45.
- [3] Efros A A, Freeman W T. Image quilting for texture synthesis and transfer[C]//Proceedings of SIGGRAPH. Los Angeles:ACM Press,2001:341-347.
- [4] 林 定,陈崇成,唐丽玉,等. 基于 Image Quilting 算法的纹理合成[J]. 系统仿真学报,2008,20(S):381-384.
- [5] Wei Liyi, Levoy M. Fast texture synthesis using tree-structured vector quantization [C]//Proceedings of SIGGRAPH. [s. l.]:[s. n.],2000:479-488.
- [6] Long J, Mould D. Improved Image quilting [C]//Proceedings of graphics interface. [s. l.]:[s. n.],2007:257-264.
- [7] 郎 瑶. 一种改进的 Sobel 边缘检测算子[J]. 广西轻工业,2011(5):59-60.
- [8] Efros A A, Leung T K. Texture synthesis by non-parametric sampling [C]//Proc of international conference on computer vision. Greece:ACM Press,1999:1033-1038.
- [9] 冯 欣,李 花. 纹理合成技术综述[J]. 计算机与现代化,2006(4):16-19.
- [10] 陈 虹,葛庆平. 使用可变大小的邻域进行纹理合成[J]. 微计算机信息,2006(06S):225-226.
- [11] 钱小燕,肖 亮,吴慧中. 快速风格迁移[J]. 计算机工程,2006,32(21):15-17.
- [12] 郭 捷,施鹏飞. 基于亮度和纹理特征的可见水印技术[J]. 红外与激光工程,2003,32(1):92-95.
- [13] 王 东,周世生,桑贤生. 基于纹理传输的烙画风格图像仿真[J]. 系统仿真学报,2010,22(12):2929-2933.
- [14] 徐文华,魏宝刚,潘云鹤. 基于纹理合成的图像艺术风格学习[J]. 武汉大学学报(工学版),2003,36(3):115-119.

(上接第 218 页)

machine [J]. Advances in Automation and Robotics, 2011 (2):501-506.

- [11] 吴换妮. 基于支持向量机的多类模式识别模型[D]. 重庆:重庆大学,2007.
- [12] Osuna E, Freund R, Girosit F. Training support vector machines: an application to face detection [C]//Proc of IEEE computer society conference on computer vision and pattern recognition. San Juan:IEEE,1997:130-136.
- [13] 郑勇涛,刘玉树. 支持向量机解决多分类问题研究[J]. 计算机工程与应用,2005,41(23):190-192.

- [14] 高银中. 用示功图计算抽油机井井口产液量方法研究 [D]. 北京:北京交通大学,2007.
- [15] 翟永杰. 基于支持向量机的故障智能诊断方法研究 [D]. 北京:华北电力大学,2004.
- [16] 刘可晶. 一种改进的矢量曲线数据压缩算法[J]. 甘肃科学学报,2005,17(3):112-115.
- [17] 李 卓,刘 斌,刘铁男,等. 基于支持向量机的抽油机故障诊断研究[J]. 微计算机信息,2006,22(7):209-211.
- [18] 邢志伟,张 辉. 基于支持向量机的飞机地面结冰型分类预测[J]. 计算机技术与发展,2012,22(6):247-250.

基于支持向量机的抽油机井功图识别研究

作者:	李春生 , 苏晓伟 , 魏军 , 王丽丽 , LI Chun-sheng , SU Xiao-wei , WEI Jun , WANG Li-li
作者单位:	李春生, 苏晓伟, LI Chun-sheng, SU Xiao-wei (东北石油大学 计算机与信息技术学院, 黑龙江 大庆, 163318) , 魏军, 王丽丽, WEI Jun, WANG Li-li (克拉玛依职业技术学院 信息工程系, 新疆 克拉玛依, 833600)
刊名:	计算机技术与发展 
英文刊名:	Computer Technology and Development
年, 卷(期):	2014(8)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_wjfz201408051.aspx