

基于关联规则的流程工业报警参数的影响分析

尚永涛, 杜长河, 杨 斌, 于 灏, 李秀福

(青岛高校信息产业有限公司, 山东 青岛 266071)

摘要:流程工业信息化系统中为保证安全生产会设置许多报警参数,这样在信息系统数据库中存在着大量的报警记录。本文从数据挖掘的角度出发,对报警记录运用关联规则算法模型挖掘,找出多参数同时报警的关联规则^[1],并根据参数报警的先后顺序,将报警规则作为重点参数报警的原因分析指导和预警条件^[2],使运行人员对参数报警有更深入的认识,指导系统和设备的安全运行。

关键词:流程工业;参数报警;关联规则

中图分类号: F062.1 **文献标志码:** A **文章编号:** 1671-1807(2014)10-0159-03

随着工业信息化进程的推进,多数流程工业企业都已经建立了以实时数据库为基础平台的企业能源信息管理系统,在该系统中用户可以对实时生产数据进行监测与分析。近年来随着数据挖掘技术的快速发展,为了有效挖掘出企业能源管理系统中大量的数据与信息,将数据挖掘技术融入到生产管理系统中,从而获取更多的数据知识,有效指导企业的生产运行,提高了信息管理系统的应用价值,同时为企业的生产运行节约了成本,提高管理效率和经济效益。

报警模块是企业能源管理系统中的必要模块,随着生产的不断运行,在实时数据库中产生大量的历史报警记录。利用关联规则模型找到报警参数之间的关联影响关系,找出哪些报警参数一直在同步发生,有可能是哪些报警参数影响了其他参数的报警发生等,通过关联规则模型找出这些关联关系,为生产系统和设备的状态监测与故障分析提供了依据^[3]。本文基于上述实际的应用需求,考虑以已经实施的企业能源管理系统的数据库平台为基础,建立了报警参数的关联规则模型,并进行关联报警模块开发,运用到实际生产中去。

1 关联规则模型

关联规则挖掘(Association rule mining)是数据挖掘中最活跃的研究方法之一,可以用来发现事情之间的相互联系,最早是 Agrawal 等人于 1993 年首先提出,为了挖掘顾客交易数据库中不同商品之间的关

联规则问题,以后诸多的研究人员对关联规则的挖掘问题进行了大量的研究^[4]。

关联规则揭示了事务之间的某种关系,支持度和置信度总是伴随着关联规则存在的,它们是对关联规则必要的补充。

对某条关联规则而言,如 $A \rightarrow B$ ($\text{support} = 45\%$, $\text{confidence} = 90\%$)。

其中 $\text{support} = 45\%$ 为支持度,是在所有的事务中 A 和 B 同时出现的概率。而 $\text{confidence} = 90\%$ 为置信度,是所有事务中,在出现 A 的情况下出现 B 的概率,即条件概率。

同时满足最小支持度阈值和最小置信度阈值的规则称为强规则。给定一个事务集 D,计算所有关联规则的支持度和置信度,将满足最小支持度阈值和最小置信度阈值的规则提取出来,最终得到最优关联规则^[5]。

模型步骤:①获取某一时间段的模型样本数据;②将模型数据转化为 0-1 矩阵;③将得到的 0-1 矩阵代入到 Apriori 模型中;④根据设定的支持度、置信度和最大事务集阈值,得到最优关联规则^[6];⑤根据因果支持度阈值得到最优因果关联规则;⑥计算最优因果关联规则预警时间^[7]。

2 流程工业参数报警的关联规则实现

流程工业指主要生产过程为连续生产(或较长一段时间连续生产)的工业,特点是原材料通过一系列

收稿日期: 2014-07-29

基金项目:科技型中小企业技术创新基金(12C26213714122)

作者简介:尚永涛(1984—),男,山东寿光人,青岛高校信息产业有限公司,数学算法研究员,管理科学与工程硕士,研究方向:能源数据挖掘与应用。

连续的化学反应和物理变化得到要求的产品,其加工顺序是固定不变的,生产设施按工艺流程布置,生产过程连续不间断,而且机理复杂、数据信息量大、处理难度较大,对管理控制的协调性、实时性、可靠性要求较高。

流程工业常常处于十分恶劣的生产环境。因此,生产安全可靠是流程工业的首要任务,对一些关键设备和关键生产过程必须有故障报警^[8]。由于报警系统只是被动的显示报警参数的报警信息,不能提前得到设备报警的信息^[9]。如果我们能够提前判断设备运行状态并及时做出操作调整,就会减少参数报警的出现,起到保护运行设备和保障生产的作用^[10]。

因此,本文建立了关联规则模型,找出报警参数的关联参数及预警时间,可以提前对关键报警参数进行预警,并及时做出调整。

2.1 关联规则模型数据

从数据库获取的样本数据是报警参数的指标名称、开始时间、结束时间和报警时长,为了便于模型计算,需要将这些数据转换成 0—1 矩阵。具体步骤如下:首先找到样本数据中所有报警开始时间的最早时间和所有报警结束时间的最晚时间,将其定义为模型的时间区间,然后考虑到参数报警有一定的时间延迟,因此需对时间区间进行切片,步长为 $h=180s$,最后根据报警参数得到 $(EndTime-BeginTime)/h$ 行 N 列的 0—1 矩阵,其中 N 为报警指标参数的个数。

2.2 模型规则

筛选规则:选取客户关注的报警参数在关联规则右侧,及客户关注的报警参数不会出现在箭头左侧的

关联规则。
因果规则:根据关联规则,找出箭头前面的所有报警参数的报警时间都在目标参数报警开始时间之前的,并记录一次因果计数,最后得到的因果次数占总次数的比例即为因果支持度。

2.3 关联规则模型过程实现

将整理后的 0—1 矩阵代入到关联规则模型中去,设定支持度和置信度阈值及最大事务集个数 n ,同时根据目标客户关注的报警参数,并根据筛选规则,最后得到最优关联规则。

模型得到的最优关联规则只是表示它们之间的关联关系,即该事务发生的概率,没有显现出前后之间的因果关系,根据因果规则及人工经验,可以找出箭头前后具有因果关系的关联规则,用于参数报警预测。若计算得到的因果支持度大于设定的因果支持度阈值,则该关联规则是具有前后因果关系的,删除那些无效的关联规则,最后得到最优关联规则,并计算规则中的所有报警时差,根据正态分布得到报警时差的 95% 置信区间,作为未来报警参数预测时间区间。

3 案例分析

以某发电厂为例,选取 2013 年 1 月 10 日—2013 年 6 月 10 日所有报警参数的报警记录,运用 Matlab 软件编程,设定支持度阈值为 0.25,置信度阈值为 0.9,因果支持度阈值为 0.75,最大事务集个数为 5,运行程序,得到的运行结果如下。

表 1 最优因果关联规则

序号	关联规则	支持度	置信度	因果支持度	95% 置信区间 (S)	
1	20→16	0.36	0.94	0.85	79.11	147.36
2	13→16	0.28	0.95	0.81	248.44	347.75
3	13,20→16	0.28	0.95	0.79	268.56	375.06

表 2 关联规则参数结果

序号	关联规则	预警时间 (S)	
1	'2#炉炉膛密相区上部温度'→'2#炉主蒸汽压力'	79.11	147.36
2	'2#炉汽包水位'→'2#炉主蒸汽压力'	248.44	347.75
3	'2#炉炉膛密相区上部温度' '2#炉汽包水位'→'2#炉主蒸汽压力'	268.56	375.06

根据表 2 中的目标报警参数的预警时间,当目标报警参数前面的关联报警参数发生报警的时候,预示着在未来多长时间内,有 95% 的概率目标报警参数

发生报警。操作人员可以在目标参数发生报警之前,对设备关联参数及时做出调整,避免发生一些不必要的损失。

4 结论

锅炉主蒸汽压力控制是锅炉控制的重要环节,关系到蒸汽质量,生产安全的重中之重。影响锅炉蒸汽压力参数的变量如:炉温、水温、液位等则属于热惯性大、反应迟缓的慢对象,这大大增加了控制的难度。根据蒸汽压力控制难等特点,通过关联规则模型,找出这些报警参数的关联关系及预警时间,可以提前对参数做出调整,使得锅炉始终在最佳状态下运行,从而达到最大的经济效益和使用效益。

参考文献

[1] ISAM ELAYYADI, SALIMA BENBERNOU, MOURAD OUZIRI, MUHAMMAD YOUNAS. A tensor-based distributed discovery of missing association rules on the cloud[J]. Future Generation Computer Systems, 2013, 3: 49—56.
[2] PANNAPA CHANGPETCH, DENNIS K J LIN. Selection of

multinomial logit models via association rules analysis[J]. WIREs Comp Stat, 2012, 51: 68—77.
[3] 毛国君, 殷立娟, 王石. 数据挖掘原理与算法[M]. 北京: 清华大学出版社, 2007: 91—125.
[4] 张小刚. 关联规则挖掘及其在复杂工业过程控制中的应用研究[D]. 长沙: 湖南大学, 2002.
[5] 马盈仓. 挖掘关联规则中 Apriori 算法的改进[J]. 计算机应用与软件, 2004, 21(11): 82—83.
[6] 郑元兵, 孙才新, 李剑, 等. 变压器故障特征量可信度的关联规则分析[J]. 高压电技术, 2012, 38(1): 82—87.
[7] 郭崇慧, 张震. 基于组合评价方法的关联规则兴趣度评价[J]. 情报学报, 2011, 30(4): 353—360.
[8] 刘红, 吴四. 多维关联规则数据挖掘在船舶价格影响因素分析中的应用[J]. 上海海事大学学报, 2013(4): 31—37.
[9] 苏立, 南海鹏, 等. 关联规则分类的数据流挖掘方法在水电机组故障诊断的应用[J]. 西安理工大学学报, 2012(1): 106—110.
[10] 吕杰林, 陈是维. 基于相关性度量的关联规则挖掘[J]. 浙江大学学报: 理学版, 2012(3): 892—896.

The Impact Analysis of Process Industry's Alarm Parameters Based on Association Rules

SHANG Yong-tao, DU Chang-he, YANG Bin, YU Hao, LI Xiu-fu

(Qingdao Gaoxiao Information Industry Co., Ltd, Qingdao Shandong 266071, China)

Abstract: The informatization system of process industry will be set many alarm parameters in order to ensure the safety in production, so that the database of information system save a lot of alarm record. This article from the perspective of data mining, to the alarm record using the association rules mining algorithm model, find out the association rules that alarm simultaneously, and according to the order of alarm parameters, the alarm rules as a guide for the analysis of the causes of the key parameters of the alarm and warning conditions, make the parameter alarm operators have a more in-depth understanding, guiding the safe running of the systems and equipment.

Key words: process industry; parameters of the alarm; association rules

(上接第 120 页)

[4] 郭庆旺, 贾俊雪. 中国全要素生产率的估算: 1979—2004[J]. 经济研究, 2005(6): 51—60.
[5] 刘秉镰, 李清彬. 中国城市全要素生产率的动态实证分析: 1990—2006[J]. 南开经济研究, 2009(3): 139—152.
[6] 韩先锋, 师萍, 宋文飞. 中国工业行业全要素生产率增长的实证分析[J]. 西安电子科技大学学报: 社会科学版, 2010, 20(1): 23—27.
[7] 张晓霞. 基于 DEA 的中国银行业全要素生产率测度[D]. 长

春: 东北师范大学, 2012.
[8] 戴平生, 陈建宝. 我国省域工业的全要素生产率变动及其影响因素实证分析[J]. 中国经济问题, 2009(3): 54—59.
[9] 魏科龄. 数据包络分析[M]. 北京: 科学出版社, 2004: 1.
[10] CAVES L R, CHRISTENSEN W E, DIEWERT. The economic theory of index numbers and the measurement of input, output, and productivity[J]. Journal of Econometrics, 1982, 50: 1393—1414.

Estimation and Comparison of Total Factor Productivity in the Yangtze River Delta Region's Industry

XUE Hong-fang, XIE Shou-hong

(School of Business, Jiangnan University, Wuxi Jiangsu 214122, China)

Abstract: This paper calculates the total factor productivity of the Yangtze River Delta Region's Industry from 1998 to 2011 by DEA-Malmquist, divides the 25 industries into high-tfp industries, middle-tfp industries and low-tfp industries. Studies have shown that, the average growth rate of Total Factor Productivity in Yangtze River Delta Region's Industry, during the years between 1998 and 20011 is 9.3% and the main source of Total Factor Productivity growth is technological progress. Technical efficiency is the most eminent factor to separate the 3 sorts of industries. Low-tfp industries accounted for a heavy proportion of Industrial output, is the focus of future work.

Key words: total factor productivity; industry; The Yangtze River Delta Region; Malmquist index