

# 基于分项计量数据对两种冷水机组 序列启停策略的识别算法\*

冯 帆, 李铮伟

(同济大学 机械工程学院, 上海 201804)

**摘要:** 冷水机组的序列控制方式对建筑能耗和室内热舒适有着显著的影响,如能对其进行自动识别和优化,对公共建筑节能具有重要的意义。从2008年开始,我国在全国范围内推行大型公共建筑的分项计量系统,冷热源系统的电耗是该系统的必要监测内容,该系统为自动识别冷水机组的启停策略提供了数据基础。因而,基于冷水机组的电耗数据,利用数据挖掘的算法来识别冷水机组序列启停方式的方法。该方法利用 OneR 分类算法来实现对以建筑即时冷负荷为指标的冷水机组序列控制方式(Q-based)和以压缩机功率为指标的控制方式(P-based)这两者的区分。最后,通过对位于上海市的一栋建筑进行模拟实验来验证该算法的可行性。

**关键词:** 冷水机组序列控制方式; 识别; 分项计量; 数据挖掘

**中图分类号:** TU831 **文献标志码:** A **文章编号:** 1673-7237(2016)02-0077-07

## An Algorithm Using Sub-Metering Data to Identify Two Chiller Sequencing Control Strategies

FENG Fan, LI Zheng-wei

(College of Mechanical Engineering, Tongji University, Shanghai 201804, China)

**Abstract:** For a multiple-chiller plant, chiller sequencing control significantly affects both building energy consumption and indoor temperature controls (indoor thermal comfort), hence the identification of chiller sequencing control is of great importance to energy saving of public buildings. Since 2008, the sub-metering systems of large public building have been applied for real-time monitoring and controlling in China and national level. To fill the gap, this paper proposes a method to identify the exact chiller sequencing control strategy and its related parameters. The proposed method uses a data-mining algorithms—OneR, to analyze the sub-metering data, aiming to discover the chiller sequencing control strategies. Finally, simulation experiments of several buildings located in Shanghai were conducted to validate the feasibility of the proposed method.

**Keywords:** chiller sequencing control; identification; sub-metering; data-mining

## 0 引言

大型公共建筑和国家机关办公建筑大都采用多冷水机组的系统,系统中冷水机之间彼此独立,以提高系统的灵活性和可靠性<sup>[1]</sup>。和单冷水机的系统相比,这样的系统在部分负荷情况下有着较高的能效<sup>[2]</sup>。多冷水机组系统的控制,除了要实现单个冷水机的负荷控制,还应考虑多个机组之间的群控策略。冷水机组的群控策略是根据建筑即时冷负荷来开启或关闭冷

水机组的控制方式,也即冷水机组序列启停控制。

目前,已经有多种冷水机组序列控制方式被开发并应用到实际系统中<sup>[3-4]</sup>,这些控制方式都根据直接或间接的冷负荷指标来启停冷水机组。根据所采用指标的不同,典型的控制方式有以下4种:以直接冷负荷为指标的控制方式(Q-based)、以冷冻水回水温度为指标的控制方式(T-based)、以旁通管流量为指标的控制方式(F-based)、以压缩机功率为指标的控制方式(P-based)。此外,还存在一些综合的控制方式<sup>[5]</sup>。在实际应用中,不同系统所采用的冷水机组序列控制方式不尽相同,

收稿日期:2015-07-11; 修回日期:2015-07-29

\* 基金项目:上海市浦江人才项目资助(15PJ1408100)

而分项计量系统暂时无法实现对系统序列控制方式信息的记录。造成这一现状的原因有很多,如建筑的物业管理人员缺乏相关的专业知识,自动记录上不能实现,而人工记录费时费力等<sup>[6]</sup>。但是,冷水机组序列控制方式的评价及优化却对这部分信息提出了要求。因此,本文提出了一种识别冷水机组序列运行方式的方法。

本文所提出的方法以分项计量数据为基础,利用数据挖掘算法来分析系统中的冷水机组序列控制方式。数据挖掘是一种从大量复杂数据中读取信息的方法,已经被成功应用在很多领域,如经济、通信、计算机科学等<sup>[7]</sup>。近些年来,也有很多研究将数据挖掘成功应用在建筑领域,如建筑能耗预测、故障检测和诊断以及优化控制等<sup>[8]</sup>。

实际系统中所应用的冷水机组序列控制方式有很多种,且每一种控制方式都有一个对应的参数空间。这就为序列控制方式的识别造成了较大的阻碍。为了简化问题,本文的识别对象仅仅为2种基本的典型控制方式 P-based 和 Q-based。基于这2种控制方式下的分项计量数据,用数据挖掘的算法从数据中解读其控制方式。根据 P-based 和 Q-based 控制方式的特征,本文选择了一种基于单一变量的分类规则识别算法(OneR)<sup>[9-11]</sup>来完成识别任务。结果表明,此方法能够有效地区分出 P-based 和 Q-based 这2种控制方式。

## 1 冷水机组的序列启停控制方式

在本文中,主要研究以冷负荷为指标的序列控制方式 Q-based 和以压缩机功率为指标的冷水机组序列控制方式(P-based)这2种控制方式。

### 1.1 以冷负荷为指标的序列控制方式(Q-based)

以冷负荷为指标的控制方式(以下统称为 Q-based)是直接以建筑即时冷负荷值为指标,来决定开启或关闭机组。图1所示是一个多冷水机系统的一次回路部分。每个冷水机组都和一个定转速水泵连接,构成一个冷水机单元,各个冷水机单元之间并联连接,彼此独立。旁通管将一次冷冻水回路与二次回路分开,平衡两边的流量。建筑即时冷负荷由下式计算得:

$$Q = c_w m_p (T_{rn} - T_{sup}) \quad (1)$$

式中:  $c_w$  为冷冻水的比热容;

$m_p$  为一次系统中冷冻水的质量流量;

$T_{rn}$  为冷冻水回水温度;

$T_{sup}$  为冷冻水供水温度。

对于多冷水机组的系统,系统所能提供的最大供冷量由开启的冷水机组决定,将不同冷水机组组合方式按照其最大供冷量进行排序,可得到一个冷水机组序列。假设一个系统可以有  $n$  种组合方式,每种组合方式记为  $St_i (i=1, 2, 3, \dots, n)$ 。然后,根据冷负荷大小选



图1 多冷水机系统的一次冷水环路系统图

择相应的开启方式,这叫做冷水机组的序列控制。其基本原理如图2所示。图中横坐标为建筑即时冷负荷,纵坐标分别为总能耗和冷水机组的组合方式,曲线为开启方式为  $St_i$  时的总的性能曲线。

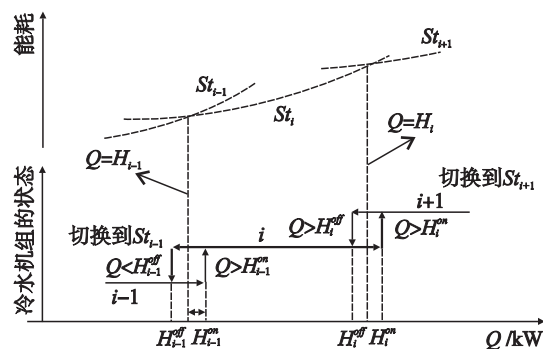


图2 冷水机组序列控制原理图

以  $St_i$  和  $St_{i+1}$  总的性能曲线的交点对应的冷负荷(记为  $H_i$ )为开关机的阈值,当冷负荷  $Q > H_i$  时,选择  $St_{i+1}$  的开启方式;反之,当  $Q < H_i$  时,选择  $St_i$  的开启方式。为了避免当负荷在开关机阈值附近波动会导致频繁开关机的行为,使开机阈值  $H_i^{on}$  稍微大于关机阈值  $H_i^{off}$ ,两者之差称为死区,记为  $d$ 。如下式:

$$H_i^{on} = Cap_i + (1/2)d \quad (2)$$

$$H_i^{off} = Cap_i - (1/2)d \quad (3)$$

此外,采用一个时间限制  $T_l$  来避免测量噪音或者其他因素的干扰<sup>[8]</sup>。

(1)开机:当建筑即时冷负荷大于开机阈值  $H_i^{on}$ ,并且持续时间超过时间限制  $T_l$  时,进入下一个开机状态  $St_{i+1}$ 。

(2)关机:当建筑即时冷负荷小于关机阈值  $H_i^{off}$ ,并且持续时间超过时间限制  $T_l$  时,返回上一个开机状态  $St_i$ 。

### 1.2 以压缩机功率为指标的冷水机组序列控制方式(P-based)

以压缩机功率为冷负荷的指标的控制方式(以下简称为 P-based),与 Q-based 的控制方式类似。在这个控制方式中,由 Q-based 的开关机阈值结合 COP 曲线算出相应的功率,即为 P-based 的开关机阈值。然后,

根据直接测量得到的压缩机功率  $P$  来决定开机状态。 $P$ -based 的开关机策略如下:

(1) 开机: 当压缩机功率大于开机阈值, 并且持续时间超过时间限制  $T_1$  时, 进入下一个开机状态  $St_{i+1}$ ;

(2) 关机: 当压缩机功率小于关机阈值, 并且持续时间超过时间限制  $T_1$  时, 返回上一个开机状态  $St_i$ 。

## 2 冷水机组序列控制方式的识别

冷水机组控制方式的识别是根据冷水机组的运行数据, 利用数据挖掘算法得出其控制方式及其相关信息的过程。该过程主要有两部分: 输入数据准备和数据处理。

### 2.1 输入数据描述

$P$ -based 和  $Q$ -based 这两种序列控制方式涉及的数据主要有压缩机功率  $P$ 、建筑即时冷负荷  $Q$ 、冷水机组开机状态  $St$ 、时间日期等。因而, 控制方式的识别以这些数据为输入。

理论上,  $P$ -based 和  $Q$ -based 这两种控制方式是高度相关的, 这是由于  $Q$  与  $P$  之间有着确定的关系, 即:

$$Q = P \times COP \quad (4)$$

但是, 由于测量误差等不确定性的存在, 以及部分负荷率下  $COP$  的动态特性,  $P$  与  $Q$  的测量值之间的关系变得不确定, 使得不同控制策略的控制性能存在差异。考虑  $P$  和  $Q$  之间不确定性后这两者的关系:

$$Q_{meas} = P_{meas} COP_n + \Delta_{dl} \quad (5)$$

式中:  $Q_{meas}$  为冷负荷的测量值;

$P_{meas}$  为冷水机功率的测量值;

$COP_n$  为冷水机的名义  $COP$ ;

$\Delta_{dl}$  为  $P$  和  $Q$  之间的不确定性。

### 2.2 序列控制方式的识别方法

#### 2.2.1 数据挖掘算法

从数据挖掘的角度, 可以把  $P$ -based 和  $Q$ -based 控制策略看作一种分类问题。以  $Q$ -based 为例,  $Q$ -based 控制策略的决策变量是建筑即时冷负荷  $Q$ ,  $Q$  是数值型变量; 响应变量是开机状态  $St$ ,  $St$  取值可以为  $St_1, St_2, \dots, St_n$ , 即视作可以分为类。根据基于  $Q$  的判定条件来决定  $St$  的取值, 例如当时  $Q < H_i$  时,  $St = St_i$ 。而对控制方式进行识别便成为了以  $Q$  为决策变量, 以  $St$  为响应变量的分类规则识别问题。

通过对比多种分类算法, 本文最终决定选用 OneR 算法。这是因为这种分类问题的决策变量只有一个, 而 OneR 正是基于单一决策变量的分类算法。对于决策变量为单一数值型变量, 响应变量为离散性变量的分类问题, OneR 算法将决策变量的取值区间分为若干

① MATLAB 是美国 Mathworks 公司出品的商业数学软件, 用于算法开发、数据可视化、数据分析和数值计算等的高级技术计算语言和交互环境。本文所用的版本是 MATLAB R2013b。

段, 每一段  $I_i$  对应一个响应值  $St_i$ 。本文所用到的 OneR 算法和常规的有些许不同, 因为冷水机组在进行控制时是确定的, 其序列也是已知的, 因而, 其决策变量最终被分成的段数也是预先可知的。

最终, 基于分项计量数据  $\{Q, St\}$ , 利用 OneR 算法可以得出一套分类规则, 即:

当  $Q \in I_i$  时,

$$St = St_i \quad (i=1, 2, \dots, k) \quad (6)$$

在本文中, 利用 Matlab<sup>①</sup>提供的工具箱 Statistics and Machine Learning Toolbox 中的函数 fitctree 来实现 OneR 算法。该函数是基于 CART 决策树的一个分类算法。因为, 本质上来讲, 本文所用的 OneR 算法也是一种分类树算法, 只是其分类变量只有一个。同时, 由于本文所提出的方法计算的终止条件是当区间  $I$  被分成了  $n$  段, 在 Matlab 中可以通过设定 fitctree 函数的参数 MaxNumSplits 的值来控制最终区间的段数。

#### 2.2.2 识别方法

以压缩机功率  $P$  和冷水机组的开机状态  $St$  数据为一组变量,  $Q$  和  $St$  为另外一组变量, 分别利用 OneR 算法进行规则识别, 可以得到两组分类规则, 记为  $Rule_p$  和  $Rule_q$ 。

通过比较这两组规则的误差(再代入误差 resubstitution error, 是分类规则基于训练数据的误差)  $E_p$  和  $E_q$  来判断系统所采用的控制方式。

$$E = \frac{N_{error}}{N_{all}} \quad (7)$$

式中:  $N_{all}$  为用于规则训练的总样本数;

$N_{error}$  为所得规则分类错误的样本数。

在冷水机组序列控制所得的运行数据中, 大部分时间内功率  $P$  都为 0 kW, 见图 3, 图中数据是一栋建筑一周的实测数据。从图中可以看出, 有将近一半时间内的功率值  $P$  都是 0 kW。这是因为在夜间等非工作时段内, 冷水机组都处于关闭状态。在压缩机功率  $P$  为 0 kW 时, 冷负荷  $Q$  并不为零, 但由于冷负荷  $Q$  和序列控制的逻辑无关, 因此, 也将这部分  $Q$  取值为零。这么大规模的零值样本并没有提供任何关于序列控制方式的有用信息, 甚至还会影响序列控制方式的识别。因而, 在将运行数据输入 OneR 算法之前先将这部分数据剔除, 以得到更为准确的误差  $E_p$  和  $E_q$ 。

此外, 由于冷水机组在不同状态下运行的时间不同, 即  $St$  取不同值的频数不同, 如表 1, 表中数据来自一栋建筑一年的计量数据。在用 OneR 算法进行规则识别时, 这种差别会误导识别的结果。因而, 在将运行数据输入 OneR 算法之前需要对每个  $St$  取值的样本数量进行平衡处理, 使每个  $St$  取值的样本数量相当。



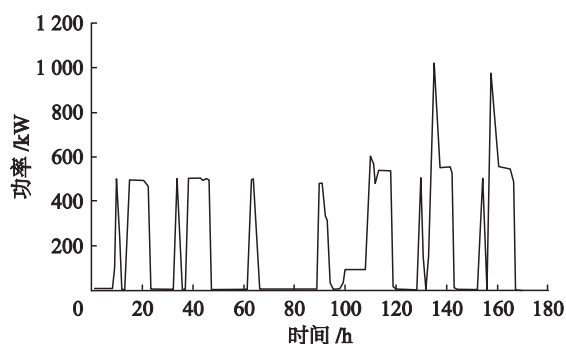


图3 冷水机组电耗的实测数据

表1 冷水机组运行状态的统计结果

$St$	1	2	3	4	5
样本频率/%	0.282	0.389	0.153	0.171	0.005

本文用到的平衡方法是以样本数量中最大的一个  $St$  值为标准,其他  $St$  值的样本直接进行自复制,使得每个  $St$  值的样本数量相等。

这就是本文所提出的用 OneR 算法来区分 Q-based 和 P-based 控制方式,其流程见图4。用这个方法区别 Q-base 和 P-based 控制方式有一个前提:如果系统采用 P-based 与 Q-based 中的某一种控制方式,则该种控制方式所对应的误差  $E$  会比较小。例如,若系统采用 P-based,分别以  $\{P, St\}$  和  $\{Q, St\}$  为 OneR 输入数据得出两组分类规则  $Rule_p$  和  $Rule_q$ ,其相应的误差  $E_p < E_q$ 。然而,这个前提不一定成立,这就会导致本文所提出的方法不一定可行,因而,需要对其进行可行性验证。

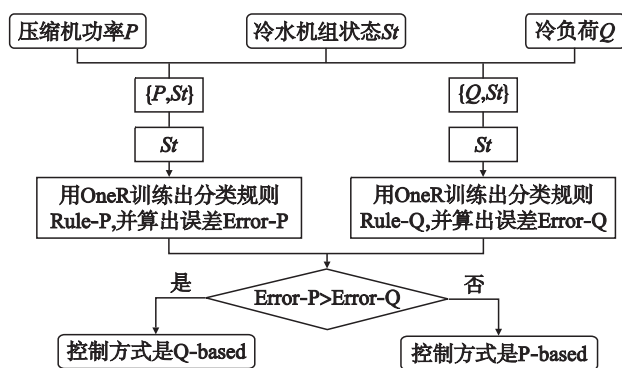


图4 识别算法的流程图

### 2.2.3 识别方法的可行性分析

当一个系统采用 P-based 控制方式时,用上述算法所得出分类规则  $Rule_p$  的误差是由死区  $d$  和时间限制  $T$  导致的。而  $Rule_q$  在  $Rule_p$  的基础上,还受  $P$  与  $Q$  之间的不确定性的影响。下面通过模拟实验来分析这两者对  $Rule$  的误差的影响。

#### 2.2.3.1 死区 $d$ 和时间限制 $T$

通过模拟实验来说明死区  $d$  和时间限制  $T$  对 OneR 算法所得分类规则误差的影响。冷负荷  $Q$  的输入数据为(0, 8 000 kW)之间的 500 个随机值,  $St$  取值个数为 6,即冷水机组序列数为 6,通过 Matlab 编程实现 Q-based 的控制。在进行死区  $d$  和时间限制  $T$  的参数设定时,图5设这两者均为 0,图6设死区  $d=50$  kW,时间限制  $T=30$  min。可以看出,在没有死区  $d$  和时间限制  $T$  的情况下,每个区间内的数据都是纯净的,反之,各个区间之间的数据存在交叉。以上两个模拟实验得到的数据  $\{Q_a, St_a\}$  和  $\{Q_b, St_b\}$  为 OneR 输入数据,得到的规则  $Rule_a$  和  $Rule_b$  的误差分别为 0 和 0.053 可以看出,死区  $d$  和时间限制  $T$  的确会影响所得规则的误差。

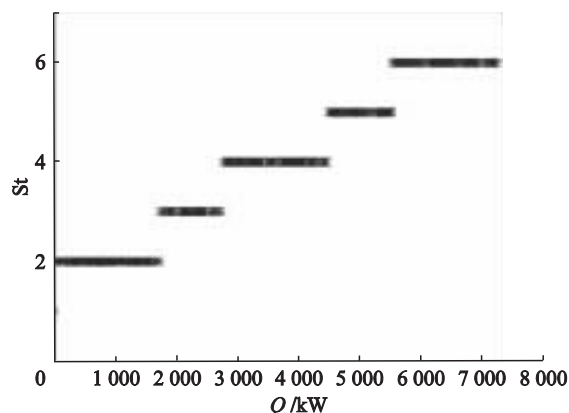


图5 无死区和时间制的 Q-based

#### 2.2.3.2 $P$ 与 $Q$ 之间的不确定性

同样,通过模拟说明  $P$  与  $Q$  之间的不确定性对分类规则误差 Error 的影响。对于采用 Q-based 控制的系统,在没有死区  $d$  和时间限制  $T$  的情况下,所得的输出数据为图5所示。需要进一步分析系统中的  $\{P, St\}$  的数据结构。当  $Q$  和  $P$  之间关系确定时,  $P$  和  $St$  的分布模式如图7所示,可见,每个区间内的数据都是纯净的。而当  $Q$  和  $P$  之间存在不确定性时,在 2.1 中对不确定性有详细描述,此处通过附加一个(0, 50 kW)的随机数来模拟这种不确定性,所得  $P'$  和  $St$  的分布模式见图8所示。可以看出,每个区间内的数据变得不再纯净。以上两个模拟实验得到的数据  $\{P, St\}$  和  $\{P', St\}$  为 OneR 输入数据,得到的规则  $Rule_p$  和  $Rule_{p'}$  的误差分别为 0 和 0.186。可以看出,  $P$  与  $Q$  之间的不确定性的确会影响所得规则的误差。

由上述分析可知,在采用 Q-based 控制的系统中,以  $\{P, St\}$  为输入数据所得到的分类规则  $Rule_p$  的误差来源有两部分:死区  $d$  和时间限制  $T$  的影响,  $P$  和  $Q$  之间不确定的影响。而以  $\{Q, St\}$  为输入数据所得到的

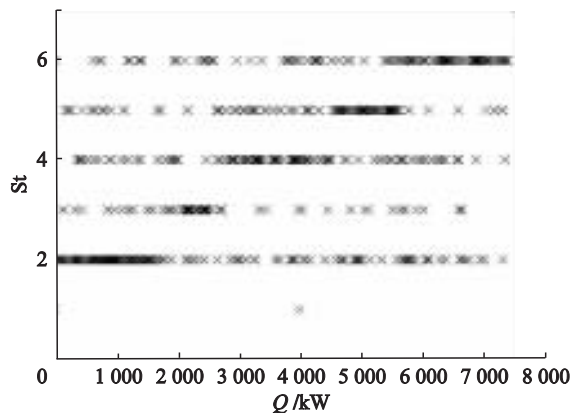


图6 有死区和时间限制的 Q-based

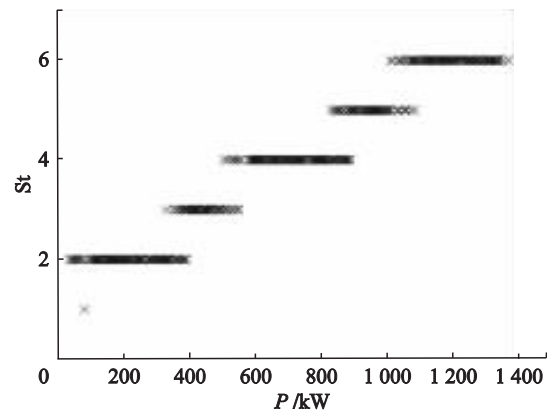


图8 P和Q之间存在不确定性

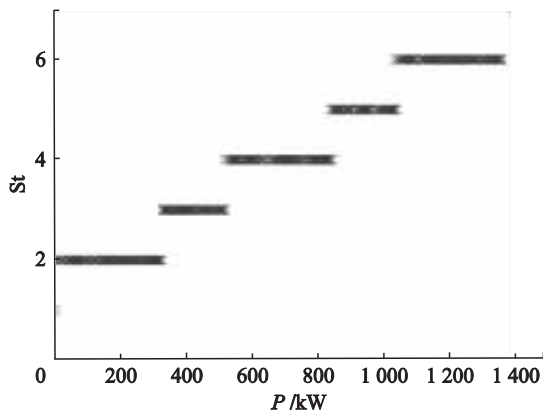


图7 P和Q之间关系确定

$Rule_Q$  的误差只受死区  $d$  和时间限制  $T$  的影响。因而,要使得  $Rule_P$  和  $Rule_Q$  的误差有明显的差异,则  $P$  与  $Q$  之间的不确定性对误差的影响要相对显著,即与死区  $d$  和时间限制  $T$  的影响大小相当或者大于后者。这个前

提是否成立,则需要对这两者的影响进行定量分析。

2.2.3.3 对两类误差的定量分析

采用上海市一栋建筑的实时数据来进行模拟实验,对上述两类误差进行准确的定量分析。该建筑为位于上海市长宁区的古北国际财富中心二期,其冷冻水系统有 2 个,分别用于商业区和办公区,在此仅选用办公区系统进行研究。该冷冻水系统的详细参数见表 2。由表 2 可以得出,该冷水机组的开启方式总共有 5 种,记冷水机 HC-F500GSG 为 S,冷水机 HC-F800GSG 为 B,则 5 种开启方式为: S1B0, S0B1, S1B1, S0B2, S1B2 (字母后的数字为该型号机组开启的台数),其对应的最大制冷量为: 1 758, 2 813, 45 71, 5 626, 7 384 kW。在此,采用一年的冷水机组电耗数据(从 2013 年 11 月到 2014 年 10 月),其计量频率为 15 min。在进行分项计量数据分析之前,先对数据进行预处理,以减少数据中的噪音、错误项以及缺失项。

表 2 冷水机组详细信息

厂家	型号	台数	制冷量 /kW	功率 /kW	名义 COP	冷冻水流量 /(m <sup>3</sup> /h)	冷冻水供水温度 /℃
日立	HC-F500GSG	1	1 758	314	5.60	252.0	6
HITACHI	HC-F800GSG	2	2 813	504	5.58	403.2	

对于死区  $d$  和时间限制  $T$ ,由于这两者对所得分类规则的误差的影响会随其取值变化而变化。因而,在这部分的研究中,对死区  $d$  和时间限制  $T$  均取多个值进行探索,这两者的取值见表 3。其中,取  $d=0$  kW 和  $T=0$  min 作定量分析的基准。

表 3 死区  $d$  和时间限制  $T$  的设定值

参数	取值范围	单位
死区 $d$	[0,15,20,25,30]	kW
时间限制 $T_i$	[0,15,30]	min

以实际计量数据  $P$  为输入,用 Matlab 模拟 P-based 控制过程,参数  $d$  和  $T$  从表 3 中取值,其中参数  $d$  的

取值约为单个冷水机组名义制冷量的 10%。由于该分项计量系统对于冷负荷的计量不能满足分析的要求,因而,此处冷负荷  $Q$  用下式间接计算得出。

$$Q_{meas}=P_{meas}COP_n+\Delta_{dl} \tag{8}$$

根据文献[4]中对不确定性的分析可知,该不确定性  $\Delta_{dl}$  可以用正态分布来进行描述,该正态分布的数字特征分别为  $\mu=0, \sigma=300$ ,即  $\Delta_{dl} \sim N(0, 300^2)$ 。

模拟之后得出 15 个结果,分别绘出  $\{P, St\}$  和  $\{Q, St\}$  的散点,见图 9 和图 10。对比图 9 和图 10 可以看出, P-based 控制下的  $\{P, St\}$  和  $\{Q, St\}$  这两组数据间的模式有一些区别,这和前边描述的一致。

分别以 15 组  $\{P, St\}$  和  $\{Q, St\}$  为 OneR 分类算法的输入数据,得到 15 组分类规则,其误差见表 4,图 11 和图 12。

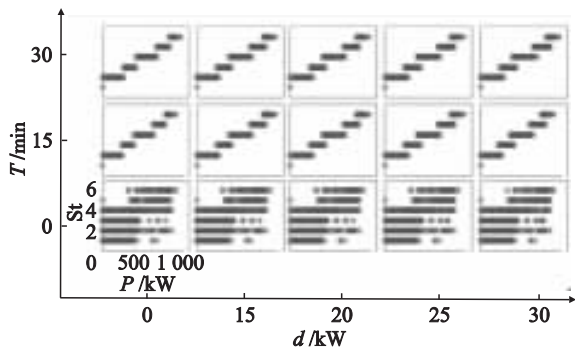


图9 冷水机组电耗  $P$  和机组运行状态  $St$  的数据散点图

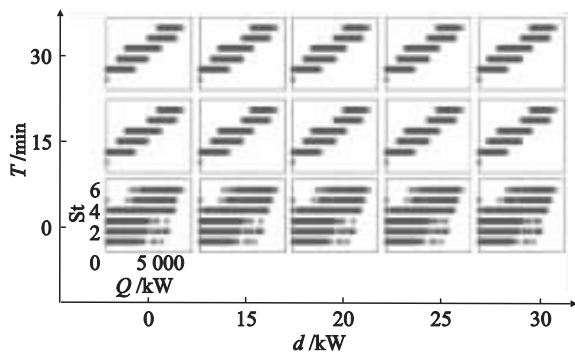


图10 冷负荷  $Q$  和机组运行状态  $St$  的数据散点图

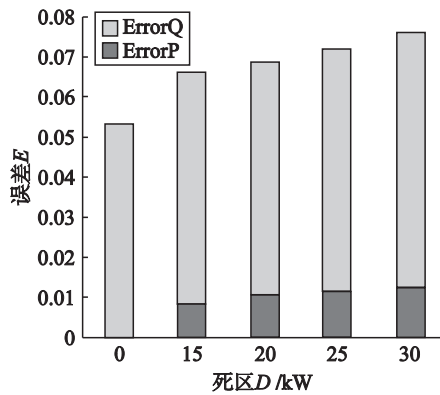


图11 时间限制  $T=0$  min 和  $T=15$  min 时 P-based 和 Q-based 的误差

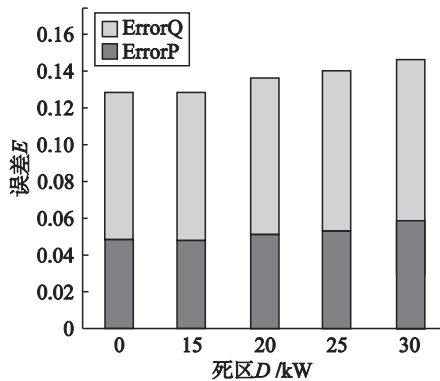


图12 时间限制  $T=30$  min 时 P-based 和 Q-based 的误差

表4中的%代表  $E_Q$  和  $E_P$  相比时  $E_Q$  的相对大小,dif则代表两者的绝对差值。

表4 P-based 和 Q-based 的误差

		死区 $d$ /kW					
		0	15	20	25	30	
时间限制 $T_l$ / min	0	$E_p$	0.000	0.008	0.010	0.011	0.012
		$E_q$	0.053	0.058	0.059	0.061	0.064
		dif	0.053	0.050	0.049	0.050	0.052
		%	—	6.200	4.871	4.800	4.336
	15	$E_p$	0.000	0.008	0.010	0.011	0.012
		$E_q$	0.053	0.058	0.059	0.061	0.064
		diff	0.053	0.050	0.049	0.050	0.052
		%	—	6.200	4.871	4.800	4.336
	30	$E_p$	0.047	0.047	0.050	0.052	0.058
		$E_q$	0.081	0.082	0.087	0.089	0.089
		dif	0.034	0.035	0.037	0.037	0.031
		%	0.721	0.732	0.744	0.708	0.533

死区  $d$  和时间限制  $T_l$  对所得分类规则误差的影响可以通过单独分析  $E_Q$  或  $E_P$  得出。当  $T_l$  取定值时,死区  $d$  对所得误差  $E$  虽然有影响,但是影响很小。而当死区  $d$  为定值时分析  $T_l$  对误差  $E$  的影响会发现,当  $T_l=0$  min 和  $T_l=15$  min 时结果完全相同,这是因为所用数据的计量间隔为 15 min。而对比  $T_l=30$  min 时的结果会发现,时间限制  $T_l$  对结果有很显著的影响。例如, $d=15$  kW,  $T_l=15$  min 和  $d=15$  kW,  $T_l=30$  min 时,  $E_P$  分别为 0.008 和 0.047,  $E_Q$  分别为 0.058 和 0.082。

$P$  和  $Q$  之间不确定性对所得分类规则误差的影响可以通过对比  $E_Q$  和  $E_P$  得出。尽管  $E_Q$  和  $E_P$  之差会随着死区  $d$  和时间限制  $T_l$  取值的变化而发生变化,但是总体而言,  $E_Q$  都明显大于  $E_P$ , 即相对差值在 50% 以上。这也就说明了,当采用 P-based 控制的系统,通过用 OneR 算法分析  $\{P, St\}$  和  $\{Q, St\}$  这两组数据,所得的分类规则  $Rule_P$  和  $Rule_Q$  所对应的误差  $E_Q$  和  $E_P$  之间有明显的差别。这样就证实了通过 OneR 算法来区分 P-based 和 Q-based 的可行性。

在验证了用本文所提出的算法识别 P-based 和 Q-based 这 2 种控制方式的可行性后,需要进一步研究,用该算法对控制方式的一些参数设定的识别。根据第三部分的介绍,冷水机组序列控制方式的设定参数主要有:开机阈值  $H_i^m$  和关机阈值  $H_i^d$ ,死区  $d$  和时间限制  $T$ 。而用本文所提出的方法得出的分类规则正好对应控制方式中的开关机阈值,而对于死区  $d$  和时间限制  $T$  这两个参数,该方法尚无法进行识别。

同样,当死区  $d$  和时间限制  $T$  取不同值时,P-based 控制方式的结果  $\{P, St\}$  数据也会相应变化,因而用 OneR 算法所识别出的开机阈值也会不同。此处,只分析该算法在最差的情况,即  $d=30$  kW 和  $T=30$  min 时,所得到的开关机阈值的准确度。当  $d=30$  kW 和  $T=30$  min



时,OneR 算法所得到的  $Rule_p$  的误差为 0.058,该分类规则如下。

当  $0 < P < 364.4$  kW 时,  $St=2$ ;  
 当  $364.4 < P < 518$  kW 时,  $St=3$ ;  
 当  $518 < P < 605$  kW 时,  $St=4$ ;  
 当  $605 < P < 1\,029.8$  kW 时,  $St=5$ ;  
 当  $P > 1\,029.8$  kW 时,  $St=6$ 。

开关机阈值  $H$  的计算结果和真实值见表 5。从中可以看出,在最差的情况下,OneR 算法的识别结果还是比较好的,其中,有 2 个值得相对误差小于 1%。而在实际中,死区  $d$  和时间限制  $T$  的设定值一般都相对小一些,则所得到的开关机阈值  $H$  的结果将会更加准确。

表 5 开关机阈值  $H$  的计算结果和真实值比较

项目	$H_1$	$H_2$	$H_3$	$H_4$
计算结果/kW	364.4	518	605	1\,029.8
真实值/kW	304	514	818	1\,028
绝对误差	60.4	4	213	1.8
相对误差/%	19.87	0.78	26.04	0.18

### 3 结论

本文提出了一种基于大型公共建筑分项计量数据,利用数据挖掘算法来识别冷水机组序列控制方式的方法。并结合一栋建筑——古北财富中心的分项计量数据,通过模拟实验对该方法的可行性进行验证。结果表明,当建筑中冷水系统所采用的控制方式为 P-based 和 Q-based 两者中的一个时,用本文所提出的方法能够有效地识别出具体应用的控制方式。

本文所提出的算法仅仅局限于 Q-based 和 P-based 这 2 种序列控制方式。但是,除了这 2 种序列控制方式外,还有多种控制方式被提出和应用到实际中去。例如以回水温度为冷负荷指标的 T-based 和以旁通管水流量为冷负荷指标的 F-based 等。由于后 2 种控制方式本身所具有的一些特点,如对于 T-based,开关机阈

值  $H_i$  并不是一系列逐渐增长的值,这就导致本文所提出的算法不再适用。下一步工作拟对该问题进行研究。

#### 参考文献:

- [1] M. Schwedler, A. Yates. Multi-Chiller-System Design and Control, Application engineering manual (SYS-APM001-EN) [M]. TRANE, La Cross, Wisconsin, USA, 2001.
- [2] ASHARE. ASHARE handbook—HVAC system and equipment, SI edition [M]. ASHRAE Inc, Atlanta, GA, 2011.
- [3] Honeywell. Engineering Manual of automatic control for commercial building, SI edition [M]. Honeywell, Minneapolis, MN, 1997.
- [4] Yundan Liao, Gongsheng Huang, Yongjun Sun, et al. Uncertainty analysis for chiller sequencing control [J]. Energy and Buildings, December 2014, 85: 187-198.
- [5] Yongjun Sun, Shengwei Wang, Fu Xiao. In situ performance comparison and evaluation of three chiller sequencing control strategies in a super high-rise building [J]. Energy and Buildings 2013, 61: 333-343.
- [6] 中华人民共和国住房和城乡建设部.关于印发《国家机关办公建筑及大型公共建筑能耗监测系统(以下简称监测系统)分项能耗数据采集技术导则》的通知 [L]. 建科[2008]114 号, 2008-06-24.
- [7] Pang-Ning Tan, Michael Steinbach, Vipin Kumar. Introduction to data mining [M]. Addison-Wesley Longman Publishing Boston, MA, USA, 2005.
- [8] Cheng Fan, Fu Xiao, Chengchu Yan. A framework for knowledge discovery in massive building automation data and its application in building diagnostics [J]. Automation in Construction, February 2015, 50: 81-90.
- [9] Holte R. C. Very Simple Classification Rules Perform Well on Most Commonly Used Datasets [J]. Machine Learning, 1993, 11(1): 63-90.
- [10] Ian H. Witten, Eibe Frank, Mark A. Hall. Data mining: Practical Machine Learning Tools and Techniques [M]. 3rd edition. Beijing: China Machine Press, 2012: 86-90.
- [11] Craig G. Nevill-Manning, Geoffrey Holmes, Ian H. Witten. The development of Holte's 1R classifier [J]. Proceedings of Second New Zealand International Two-Stream Conference, 1995, 239-242.

作者简介: 冯帆(1993),男,西安人,毕业于同济大学,建筑环境与设备工程专业,研究方向:建筑能耗(18817598306@163.com)。

指导教师: 李铮伟(1981),男,同济大学助理教授,研究方向:建筑能源系统的优化运行及故障诊断等。

(上接第 58 页)最大,其主要原因是在模拟分析时未考虑遮阳措施,夏季西晒对建筑能耗影响较大,因此,西向宜选择 Low-E 中空玻璃及合理的遮阳措施。

影响办公建筑全年能耗的因素有很多,窗墙比仅是其中重要因素之一,在今后建筑设计中,应当综合考虑选择合适的窗墙比,既能满足建筑节能,又能够有很好的采光和通风效果,提高人们的办公舒适性。

#### 参考文献:

- [1] DGJ32J 96—2010,公共建筑节能设计标准[S].
- [2] 张吉超,李庆磊,房大兵,等.窗墙比和体形系数对办公建筑能耗的影响[J].建筑节能,2011,(9):47-49.
- [3] 王秀丽.民用建筑节能设计的要点解析[J].价值工程,2010,(34):53-54.

作者简介: 黄金美(1984),女,云南宣威人,工程师,主要从事建筑节能技术研究(280515952@qq.com)。