

基于 GMM 聚类的空调机组多未知模态 辨识方法研究

李冬辉, 何鹏林

(天津大学电气与自动化工程学院, 天津 300072)

摘 要: 智能建筑空调机组故障检测与诊断是保证建筑环境安全、舒适、节能的基本方法. 然而, 多未知模态辨识仍是其中的关键难点之一. 鉴于高斯混合模型(GMM)不受特定概率分布局限, 可在区分类别的基础上直接得出数据的统计分布, 具有优越的计算性能, 且能拟合任意连续分布, 本文提出一种基于 GMM 的空调机组多未知模态辨识方法. 仿真试验结果表明, GMM 聚类方法在空调机组运行模态辨识中具有较高的准确性与可靠性.

关键词: 空调机组; 运行模态; 高斯混合模型; 辨识方法

中图分类号: TP23 **文献标识码:** A **文章编号:** 0372-2112 (2014)10-2004-05

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2014.10.021

Research on Identification of Unknown Modes for Air-Conditioning Based on GMM

LI Dong-hui, HE Peng-lin

(School of Electrical and Automation Engineering, Tianjin University, Tianjin 300072, China)

Abstract: Fault detection and diagnosis for Air-Conditioning(AC) system in intelligent building is the guarantee of building safety, comfort and energy-saving. However, the identification of unknown modes for AC system is still one of the key difficulties. The Gaussian Mixture Model(GMM) is not confined to a specified probability distribution and the distribution of given data can be obtained on the basis of classification. With excellent computing performance any continuous distribution can be fitted via GMM. Hence, a new idea of clustering based on GMM to identify the unknown operation modes in AC is developed. The simulation experiment has qualified the accuracy and reliability of this GMM mode identification approach.

Key words: air-conditioning(AC) system; operation mode; Gaussian mixture model(GMM); identification approach

1 引言

优化空调故障检测与诊断技术可以使空调高效低碳运行, 提高智能建筑的节能环保性^[1]. 国内外专家学者在此方面已经取得了不少优秀成果, 如: Thomson M 等人提出以换热器的传热系数为依据来计算和监控热泵的实时运行状态^[2]; 湖南大学陈友明教授将主成分分析法应用到传感器的数据信号分析中以检测传感器中常见的故障^[3]等. 这些优秀成果在一定程度上对空调系统故障检测有一定的帮助. 但针对空调系统独有的参数漂移、多种模态、渐变和既有故障这三大类关键难点^[4], 尚未形成行之有效的方法. 本文针对多种模态这一关键难点, 结合智能建筑空调系统中相同单元在相同模态下具有相同的统计特性^[5~7], 提出比较单元相近统计特性进

行模态辨识的新思路, 着重研究空调机组故障诊断中运行模态的聚类分析问题. 与 K-均值聚类、K-中心聚类等方法相比, 高斯混合模型(GMM)不受特定概率分布的局限, 可在区分类别的基础上直接得出数据的统计分布, 具有良好的计算性能, 而且可通过增加模型分量来拟合任意连续分布, 是一种无监督聚类方法. 因此本文采用基于 GMM 聚类方法辨识空调机组故障诊断中的多未知模态: 首先采集单元数据, 选择利于区分模态的特征集, 并对其进行计算, 然后采用 EM(期望最大化)法估计模型的参数, 建立 GMM 模型以辨识单元模态.

2 空调机组运行模态分析

空调机组的主要任务是对房间空气温、湿度进行处理, 其控制结构如图 1 所示. 空调机组温度处理单元

有制冷、制热、通风三种运行模式,湿度处理单元有加湿、除湿、通风三种运行模式,空调机组的温、湿度模式基本相似,文中仅以温度处理单元为例展开分析.其中冷热模式下的模型结构基本相似,仅模型参数存在差异;通风模式仅与外界进行空气交换,与前两者的模型差异较大.

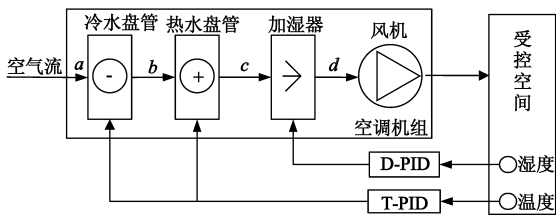


图1 空调机组温湿度控制结构

3 温度处理单元模型的建立

严格来说,温度处理单元是一个大时间常数且具有纯滞后的控制对象.本着突出关键因素,忽略次要因素的原则,暂不考虑控制对象的纯滞后特性.

如图 2 所示,空气温度处理单元主要由温度传感器、冷热水阀及盘管组成.以制冷为例,空气流入带有盘管的空气管道,与盘管中冷水介质进行热交换,以此降低空气温度.调节盘管阀门开度,改变介质流量可以调节温度的变化幅度.空气管道内的冷水介质与空气满足能量守恒定律.

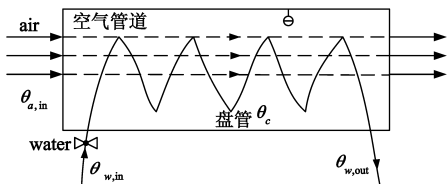


图2 温度处理单元结构示意图

盘管热量变化率为 $m_c c_c \frac{d\theta_c}{dt}$. 其中 m_c 为盘管质量 (kg), c_c 为换热比热容 (J/kg·°C), θ_c 为盘管温度 (°C), 盘管外空气温度与盘管温度相近.

单位时间内空气放出的热量为 $k_a A_a (\theta_{a,in} - \theta_c)$. 其中 k_a 为传热系数 (W/m²·°C), A_a 为换热面积 (m²), $\theta_{a,in}$ 为进风温度 (°C).

单位时间内冷水介质吸收的热量为 $L \rho c (\theta_{w,in} - \theta_{w,out})$, 其中 L 为介质流量 (m³), 与阀门开度成正比, 可近似视为阀门开度, $\theta_{w,in}$ 、 $\theta_{w,out}$ 为入口、出口处介质的温度 (°C), ρ 为介质密度 (kg/m³), c 为介质定压比热容 (J/kg·°C).

根据能量守恒, 可得出空调机组温度处理单元数学模型:

$$m_c c_c \frac{d\theta_c}{dt} = k_a A_a (\theta_{a,in} - \theta_c) - L \rho c (\theta_{w,in} - \theta_{w,out}) \quad (1)$$

经变换:

$$m_c c_c \frac{d\theta_c}{dt} + k_a A_a \theta_c = k_a A_a \theta_{a,in} + L \rho c (\theta_{w,out} - \theta_{w,in}) \quad (2)$$

上式为温度处理单元的微分方程式. 式中阀门开度 L 是输入量, 盘管温度 θ_c 是单元的输出量, $k_a A_a \theta_{a,in}$ 视作干扰. 得到温度处理单元数学模型为:

$$G_T(s) = \frac{K_1}{T_1 s + 1} \quad (3)$$

式中 $K_1 = \frac{\rho c (\theta_{w,out} - \theta_{w,in})}{k_a A_a}$; $T_1 = \frac{m_c c_c}{k_a A_a}$

假设采样周期为 T , 采用反向差分变换离散化得:

$$\theta_c(k) = \frac{T_1}{T_1 + T} \theta_c(k-1) + \frac{K_1 T_1}{T_1 + T} L(k) \quad (4)$$

经整定: $K_1 \approx 25$, $T_1 \approx 140s$, 采样周期取 $T = 1s$.

以上分析忽略滞后, 若系统为滞后系统, 滞后时间为 τ_0 , 其差分方程则为

$$\theta_c(k) = [T_1 / (T_1 + T)] \theta_c(k - k_0) + [K_1 T / (T_1 + T)] L(k)$$

式中 $k_0 = \tau_0 / T$. 为方便分析, 下文忽略滞后.

4 基于 GMM 聚类系统建立

如图 3 所示为基于 GMM 的空调机组不同单元模式的聚类系统结构, 该系统由训练阶段和分类阶段两部分构成. 在训练阶段, 首先用 Kalman 滤波计算出特征参数, 然后进行 EM 迭代, 最后得出 GMM; 在分类阶段, 根据训练阶段确定的 GMM, 依据 Bayes 最大后验概率 (MAP) 判别准则将该单元采集到的所有数据点分到相应的高斯分布, 进行模式辨识.

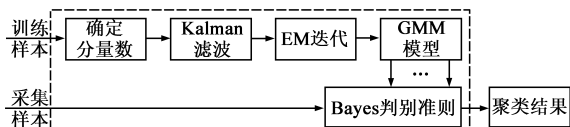


图3 基于GMM空调机组聚类系统结构

4.1 特征参数选择与计算

选择合适的特征参数是产生正确聚类结果的基础. 根据温度处理单元的分析结果, 将上述差分方程转换成如下状态空间表达式:

$$\begin{aligned} \mathbf{x}_k &= \mathbf{E} \mathbf{x}_{k-1} \\ \theta_k &= [\theta_{k-1}, L_k] \mathbf{x}_k \end{aligned} \quad (5)$$

式中 $\mathbf{x}_k = [T_1 / (T_1 + T), K_1 T / (T_1 + T)]^T$ 为状态向量, L_k 表示 k 时刻的输入阀门开度, θ_k 是 k 时刻的输出空气温度. 分析每种模式下的状态分布 \mathbf{x}_k : 通风模式下, L_k 为零, 输出 $\theta_k = \theta_{k-1} = \dots = \theta_0$, 于是 $\mathbf{x}_k = [1, 0]^T$; 制冷、制热模式下增益 K_1 有正负之分, 并且由于冷热介质传热比热容不同导致 T_1 存在差异. 所以每种模式下

全局变量 x_k 的分布必定不同,宜将其作为特征参数.

因为该特征参数无法直接通过各种传感器执行器中的数据采集得到.基于该单元是线性系统,其值可通过 Kalman 滤波得到,这里不再赘述.

4.2 基于 EM 算法的 GMM 参数确定

基于 GMM 聚类的一系列优点,本文采用 GMM 对空调机组进行模态的聚类分析.GMM 的表达式为:

$$f(x) = \sum_{k=1}^K \alpha_k f_k(x; \beta_k) = \sum_{k=1}^K \alpha_k p(x; \mu_k, \Sigma_k) \quad (6)$$

其中 α_k 为权值,满足 $\sum_{k=1}^K \alpha_k = 1, 0 \leq \alpha_k \leq 1; p(x; \mu_k, \Sigma_k)$ 表示第 k 个高斯密度函数,其概率分布为:

$$p(x; \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{d/2} \sqrt{|\Sigma_k|}} \cdot \exp(-1/2(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k)) \quad (7)$$

现已知特征参数,要得到 GMM,即根据特征参数,估算出 GMM 的权值 α_k 、均值 μ_k 和协方差 Σ_k ,这一过程可以采用 EM 算法完成.

EM 算法通过假设某特征数据的类别简化了问题,再根据迭代得到 α_k, μ_k 和 Σ_k 的最大似然估计值.主要步骤如下^[8-10]:

E-步:引入隐含变量,即 x_i 源于第 k 类的后验概率 $p(k | x_i, \Phi^{(t)})$. 现已知 $X = \{x_1, x_2, \dots, x_n\}$,上一步得到的估计参数 $\Phi^{(t)} = (\alpha^{(t)}, \beta^{(t)})$ 或初始参数值 $\Phi^{(0)}$,初始参数值可随机选取也可采用其他简单聚类得到.其中 $\beta^{(t)} = (\mu^{(t)}, \Sigma^{(t)})$. 则隐含变量为:

$$p(k | x_i, \Phi^{(t)}) = \frac{\alpha_k^{(t)} p_k(x_i | \beta_k^{(t)})}{\sum_{j=1}^K \alpha_j^{(t)} p_j(x_i | \beta_j^{(t)})} \quad (8)$$

M-步:将上一步得到的隐含变量代入三个参数的最大似然估计公式可以得到第 $(t+1)$ 次迭代的 GMM 参数:

$$\alpha_k^{(t+1)} = \frac{1}{n} \sum_{i=1}^n p(k | x_i, \Phi^{(t)}) \quad (9)$$

$$\mu_k^{(t+1)} = \frac{1}{n \alpha_k^{(t+1)}} \sum_{i=1}^n x_i p(k | x_i, \Phi^{(t)}) \quad (10)$$

$$\Sigma_k^{(t+1)} = \frac{1}{n \alpha_k^{(t+1)}} \sum_{i=1}^n (x_i - \mu_k^{(t+1)})^2 p(k | x_i, \Phi^{(t)}) \quad (11)$$

此时判断终止条件 $\|\Phi_k^{(t+1)} - \Phi_k^{(t)}\| < \epsilon$, 其中 ϵ 设置为足够小的值.若满足,迭代终止;否则返回继续执行 E-步.

综上分析,在已知 GMM 分量数和参数初始值的基础上,确定此模型的参数 $(\alpha_k, \mu_k, \Sigma_k)$,即通过迭代运算就可得到 GMM.

4.3 模态辨识

GMM 模型建立后,将数据输入 GMM,采用 Bayes 最

大后验概率(MAP)准则,即可得到该数据产生于哪一具体模态,原理如下:

假设有一组样本数据集,但未知其中数据样本点来自何种模态,该问题则可以通过下面的式子来确定:

$$k^* = \arg \operatorname{Max}_{k \in K} P(k | X) \quad (12)$$

如图 3,训练阶段得到 GMM 来表示空调机组单元的概率分布情况.已知该单元的特征参数集为 $X = \{x_1, x_2, \dots, x_n\}$,则每个数据样本点来自第 k 个高斯分量的后验概率为:

$$P(\alpha_k, \beta_k | x_i) = \frac{\alpha_k p(x_i | \beta_k)}{\sum_{j=1}^G \alpha_j p(x_i | \beta_j)} \quad (13)$$

其中, α_k 为第 k 个高斯分布密度的先验概率; $p(x_i | \beta_k)$ 为特征向量 x_i 属于第 k 类的概率密度.样本 x_i 属于某类的后验概率密度最大, x_i 就属于该类,因此聚类结果可以由最大后验概率准则给出,即

$$k^* = \arg \operatorname{Max}_{1 \sim K} [P(\alpha_k, \beta_k | x_i)] \quad (14)$$

其中 k^* 表示聚类结果,即 x_i 属于后验概率最大的高斯分布.

5 空调机组模态辨识仿真

前文给出了基于 GMM 的空调机组单元的聚类系统,现在 MATLAB 环境下设计程序验证本文提出的基于 GMM 聚类方法的可行性与准确性.

本次仿真数据来自 Honeywell 智能大厦中央空调系统模拟实验平台,如图 4 所示,该平台能够再现智能建筑中的空调运行状况.本文中各算法均是在一台 CPU 为 2.83GHz,内存为 4.00GB 的 Win Server2008 服务器上实现的未经优化的 Matlab 程序,软件版本为: Matlab7.8.0.347.从模拟实验平台每种模态下采集 300 组数据,共 900 组数据作为训练样本进行模型训练,由前文可知系统分量数为 3.本文 EM 算法设定迭代终止条件为连续的参数估计值距离不超过 0.1%,即 $\epsilon = 0.001$,经过 574 次迭代后 EM 算法收敛,由此得到 GMM 模型的参数如表 1 所示.

另从实验平台上采集 900 组数据,在经过 Kalman 滤波后得到数据的实际分布如图 5 所示.将新采集的样

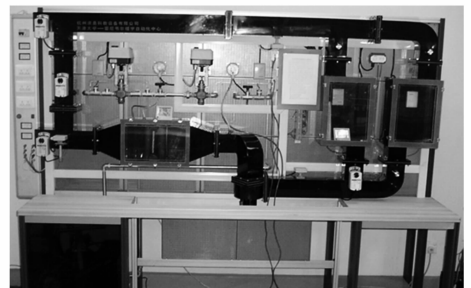


图4 Honeywell智能大厦中央空调系统模拟实验平台

本数据经 GMM 系统聚类后得到的结果如图 6 所示.

表 1 迭代终止时 GMM 参数

模式	α_k	μ_k	Σ_k
制热	0.3354	(1.0008, 0.2545)	[0.0078, -0.0007; -0.0007, 0.0080]
通风	0.3341	(1.0052, 0.0153)	[0.0098, -0.0003; -0.0003, 0.0100]
制冷	0.3305	(0.9845, -0.1803)	[0.0102, -0.0005; -0.0005, 0.0080]

图中“+”、“·”、“*”依次表示依据 GMM 聚类辨识出的制热、通风、制冷模态下的样本. 图中的较大的“●”表示聚类中心 $\mu = (\mu_1, \mu_2)$, 制热模态的聚类中心为(1.0032, 0.2523), 通风模态的聚类中心为(1.0072, 0.0150), 制冷模态的聚类中心为(0.9845, -0.1859). 三者都满足 $\mu_1 \approx 1$, 非常接近理想中心. 聚类中心外的三个圈分别表示 $1 - \sigma, 2 - \sigma, 3 - \sigma$ 的边界, σ 为数据的标准差, $\sigma = \text{sqrt}(\Sigma)$. 计算得数值分布在 $(\mu - \sigma, \mu + \sigma)$ 中

的概率为 0.6526; 数值分布在 $(\mu - 2\sigma, \mu + 2\sigma)$ 中的概率为 0.9504; 数值分布在 $(\mu - 3\sigma, \mu + 3\sigma)$ 中的概率为 0.9973, 可见图中的数据分布满足正常分布, 基本全部位于 $(\mu - 3\sigma, \mu + 3\sigma)$ 范围内.

作为比较, 图 7 所示为采用 K-均值聚类得出的聚类结果, K-均值聚类是一种硬聚类, 从图中可以看出, 此时全部数据点被分成三大块, 从上到下依次为制热模态、通风模态、制冷模态, 三者之间的界限清晰, 但界限两侧数据点没有相互融合的现象, 与实际分布偏差很大.

通过以上比较不难发现, 采用 GMM 聚类产生的结果更加贴合实际分布, 对模态辨识准确率更高, 更加可靠. 所以当辨识出每个单元当前处于哪种运行模态, 将相同运行模态下的单元统计特性进行比较, 才能根据统计特性变化诊断出是否在一个或多个单元发生故障.

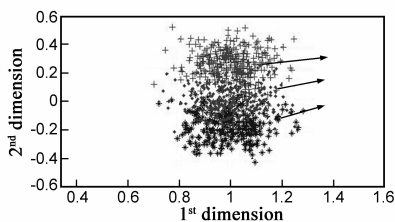


图5 各模态下的实际数据分布

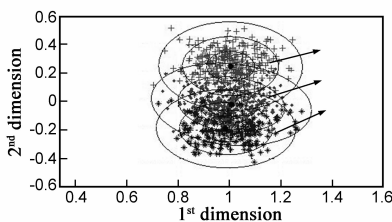


图6 GMM聚类辨识结果

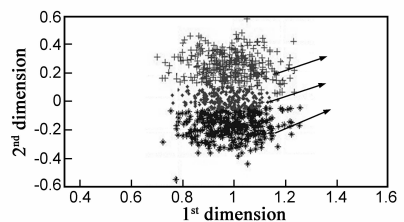


图7 K-均值聚类辨识结果

6 结论

目前, 多未知模态辨识仍是空调机组故障检测与诊断的关键难点之一. 针对本难点及 GMM 的一系列优点, 本文提出了一种基于 GMM 聚类的空调机组多未知模态辨识方法, 并在 MATLAB 环境下设计程序对该方法进行验证. 验证结果表明, 与其他聚类方法相比, 基于 GMM 的聚类方法能够有效准确地对空调机组多未知模态进行辨识, 从而证明本文方法具有很高的可行性与可靠性. 本文提出的基于 GMM 聚类的多未知模态辨识方法可用于智能建筑空调机组故障检测与诊断.

参考文献

- [1] 曹茂春, 齐雄. 大型智能建筑控制节能和管理节能技术探讨[J]. 智能建筑与城市信息, 2013(3): 29 - 33.
- [2] Thomson M, Twigg P M, Majeed B A. Statistical process control based fault detection of CHP unit[J]. Control Engineering Practice, 2000, 8(8): 13 - 20.
- [3] 陈友明, 郝小礼. 建筑能源管理与控制系统中传感器故障及其检测与诊断[J]. 暖通空调, 2004, 34(2): 83 - 88.
- [4] 周东华, 胡艳艳. 动态系统的故障诊断技术[J]. 自动化学报, 2009, 35(6): 748 - 758.
Zhou Dong-Hua, Hu Yan-Yan. Fault diagnosis techniques for dynamic systems[J]. Acta Automatica Sinica, 2009, 35(6): 748

- 758. (in Chinese)
- [5] Anwer Bashi, Vesselin P Jilkov, X Rong Li. Fault detection for systems with multiple unknown modes and similar units and its application to HVAC[J]. IEEE Transactions on Control System, 2011, 19(5): 957 - 968.
- [6] Anwer Bashi, Vesselin P Jilkov, X Rong Li. Fault detection for systems with multiple unknown modes and similar units-part I [A]. Proceedings of 12th International Conference on Information Fusion[C]. Seattle: WA, 2009. 732 - 739.
- [7] Anwer Bashi, Vesselin P Jilkov, X Rong Li. Fault detection for systems with multiple unknown modes and similar units-part II [A]. Proceedings of 12th International Conference on Information Fusion[C]. Seattle: WA, 2009. 740 - 747.
- [8] 李文斌, 刘椿年, 陈焱瑛. 基于混合高斯模型的电子邮件多过滤器融合方法[J]. 电子学报, 2006, 34(2): 247 - 251.
Li Wenbin, Liu Chunnian, Chen Yiyang. Combining multiple email filters of naive bayes based on GMM[J]. Acta Electronica Sinica, 2006, 34(2): 247 - 252. (in Chinese)
- [9] 梁岩, 鲍长春, 夏丙寅, 等. 用基于高斯混合模型的压缩域语音增强方法[J]. 电子学报, 2012, 40(10): 2031 - 2038.
Liang Yan, Bao Changchun, Xia Binyin, et al. Compressed domain speech enhancement based on gaussian mixture model [J]. Acta Electronica Sinica, 2012, 40(10): 2031 - 2038. (in Chinese)
- [10] 谢昭, 高隼. 基于高斯统计模型的场景分类及约束机制

新方法[J].电子学报,2009,37(4):733-738.

Xie Zhao, Gao Jun. Anovel method for scene categorization

with constraint mechanism based on gaussian statistical model

[J]. Acta Electronica Sinica, 2009, 37(4): 733-738. (in Chinese)

作者简介



李冬辉 男, 1962年11月出生, 黑龙江伊春人, 教授、博士生导师. 主要从事故障诊断技术、楼宇自动化、电力电子应用及计算机控制等方面的研究.

E-mail: lidonghui@tju.edu.cn



何鹏林 男, 1988年5月出生, 甘肃武威人, 硕士研究生. 2012年毕业于天津大学电气与自动化工程学院自动化系. 主要从事楼宇自动化及嵌入式控制系统方面的研究.

E-mail: hel_tju@163.com