

# 一种基于数据驱动的 CPS 建模方法研究

杨 帆 刘 彦 李仁发 段梦琴 谢国琪 黄 晶

(湖南大学嵌入式与网络计算省重点实验室 长沙 410082)

**摘 要** 信息-物理融合系统(CPS)由物理事件与计算系统两大部分组成,前者专注于处理连续的物理过程,与时间的流逝存在密切联系;后者则只能处理基于 0-1 机制的离散问题,两者存在本质上的差别.传统的建模方法一般需事先估计系统将要经历的状态及每一个状态转换所需的时间,这在复杂的 CPS(Cyber-Physical System)环境中难以预测,也没有真正意义上实现计算系统与物理事件之间的交互.因此,我们提出一种新的建模方法 DCDM:面向 CPS 的真实环境经传感器、网络等测试得出一些原始的离散数据,通过一定的数学方法(GMDH)从数据出发建立系统的连续模型,为了缩小离散数据与连续模型之间的误差,采用反馈控制的方式不断调整两者之间的差距,直至其减小到一定的范围之内.DCDM 从根本上改变了传统的建模方式,提出了一种从离散数据出发建立系统模型的思想,通过反馈控制实现物理事件与计算系统之间的深度融合.DCDM 具有以下 4 种优势:(1)建模对象容易获得且更加客观,能真实反映系统本身;(2)通过数学方法直接实现参数的筛选,去除冗余属性,减小弱影响因子对系统模型的影响;(3)相较于其他数据拟合的方法,DCDM 具有更加简单的复杂度,能节约大量的计算时间;(4)在离散数据与连续模型之间的误差方面,相较于其他的算法有着明显的改进.实验结果表明,在真实数据集中,DCDM 的执行速度快于当前最新的数据建模方法,且数据集越大优势越明显.

**关键词** 信息-物理融合系统;GMDH;连续系统;离散系统;反馈控制;物联网

**中图法分类号** TP301; TP391 **DOI 号** 10.11897/SP.J.1016.2016.00961

## A Modeling Method Research Based on Data in Cyber-Physical System

YANG Fan LIU Yan LI Ren-Fa DUAN Meng-Qin XIE Guo-Qi HUANG Jing

(Key Laboratory for Embedded and Network Computing, Hunan University, Changsha 410082)

**Abstract** Physical process and computing system are two aspects of CPS. The former focuses on processing continuous objects which requires time-dependent models, and the latter takes its attention in handling the discrete problems (0-1 formalism) in most cases, so it exists great difference in modeling method of this two aspects. Notice that the traditional modeling methods focused on states transition in general, which is difficult to predict in such a complicated environment of CPS, it didn't truly realize the interaction between the two aspects of CPS. In this paper, we obtained some discrete data from the real environment, and presented a new modeling mechanism called DCDM which started from the original data to build a continuous model through a mathematical method (GMDH). To decrease the relative error between them, we took the feedback control to the original data to adjust the error. Furthermore, from the adjustment of new discrete data, our method established the continuous modeling again by repeating the above steps until the error turns to an acceptable range. DCDM has fundamentally changed the traditional modeling style

收稿日期:2015-04-03;在线出版日期:2015-11-19. 本课题得到国家自然科学基金(61173036,61300037)资助. 杨 帆,男,1985 年生,博士研究生,中国计算机学会(CCF)会员,主要研究方向为嵌入式系统、计算机系统结构. E-mail: yangfanf117@126.com. 刘 彦(通信作者),男,1979 年生,博士,中国计算机学会(CCF)会员,主要研究方向为嵌入式计算系统. E-mail: liuyan@hnu.edu.cn. 李仁发,男,1956 年生,博士,教授,博士生导师,中国计算机学会(CCF)杰出会员,主要研究领域为计算机系统结构、嵌入式系统、计算机网络. 段梦琴,女,1989 年生,硕士研究生,主要研究方向为嵌入式系统. 谢国琪,男,1983 年生,博士,主要研究方向为计算机系统结构、容错计算. 黄 晶,男,1986 年生,博士研究生,主要研究方向为嵌入式系统.

with the proposal of a new modeling direction which models the system from discrete data, and deeply realizes the integration between the discrete system and continuous system through feedback. DCDM has the following four advantages: (1) the modeling object is easy to be obtained and objective to reflect the system itself; (2) DCDM can directly select the suitable parameters through the GMDH to remove the redundancy attributes and minimize the factors impact on the system models; (3) Compared to other data fitting methods, DCDM has a simpler complexity to save a lot of computing time; (4) There is a significant improvement in the error between the discrete data and the continuous model. Experimental results show that the implementation of DCDM is faster than the current data modeling method, and it is more obvious with the increase of data.

**Keywords** cyber-physical system; GMDH; continuous system; discrete system; feedback control; Internet of Things

## 1 引 言

信息-物理融合系统(CPS)自2006年提出以来,国内外众多专家学者进行了深入的研究,使其成为当今计算机前沿交叉学科的重点研究内容,具有广泛的应用前景<sup>[1]</sup>. CPS主要包含信息系统与物理系统两个方面,前者涉及计算系统与网络部分,专注于处理离散的计算过程、网络通信及反馈控制,最终目的都是将问题转换为计算机可以处理的问题,其核心思想是离散的;后者则主要描述真实的物理世界,包括高山流水、天气、阳光、化学反应、机械运动、环境变化等等,面向对象是整个真实的物理环境,其核心思想是连续的<sup>[2]</sup>. 对于如何设计实时、可靠的CPS系统,解决信息与物理之间存在的矛盾就成了专家学者共同面临的难题<sup>[3]</sup>. 模型驱动开发架构(Model Driven Architecture, MDA)能够在早期阶段对系统进行分析与验证,有助于保证系统的质量属性,并有效控制开发与成本<sup>[4]</sup>. 但随着CPS研究的深入,一些在通用计算机领域很少受到关注的系统建模问题逐渐凸显出来. 在CPS中,离散的计算过程与连续的物理过程并存于系统之中,难以用单一模型刻画两者之间的融合特征,且CPS系统具有明显的时空性和动态非确定性,现有的建模语义都难以适应<sup>[5]</sup>,发展新的建模语言及方式是推动CPS发展的重要途径.

图1简单描述了基于MDA的开发流程. 金字塔的最底层主要包括CPS的处理对象:人体、计算机、网络、物理世界等等,MDA开发流程的第1步是将这些关联对象分类,从复杂的环境中加以抽象,

并提取相关的特征向量;第2步建立相应的模型,鉴于CPS所涉及的面向对象,其模型一般包括连续模型与离散模型两大类,一般情况下使用状态机、PETRI网、AADL等建模语言描述信息部分,而采用微积分方程的方式来描述物理事件. 在此之后运用一定的建模理论如离散模型中插入连续变量,连续模型分解为多个离散状态,发展新的建模语言一体化描述离散连续系统等方式将模型统一起来. 最后也就是金字塔的顶层,在统一模型的基础上设计实现整个系统,并提前对系统进行分析验证,包括可靠性验证、实时性分析、安全分析等等,以避免给系统带来重复设计与操作<sup>[6]</sup>. 从设计流程可以看出基于MDA的开发方法其关键问题在于采用何种建模方式来建立系统模型.

目前,从信息的角度与物理的角度考虑建模的思维方式是完全不一样的<sup>[7]</sup>. 众所周知,信息科学处理的是离散问题,最本质的特征是将一系列的问题抽象成0/1模式,在此基础上再运用计算机的离散运算加以处理,信息系统建模语言发展迅速,时间自动机、PETRI网、AADL、UML等无一例外都高度依赖于计算机科学的发展<sup>[8]</sup>. 另一方面,物理事件是连续的,它们是不不断变化的,其建模方式一般是建立关于时间的微积分方程,从立方方程解方程的过程中寻找规律及相关属性.

总之,CPS是一个混合系统,既包含离散的计算过程,也需要处理连续变化的物理事件,因此发展一种新的CPS建模方法解决当前离散与连续建模之间的问题,统一CPS的建模方式是一件非常有意义的研究工作.

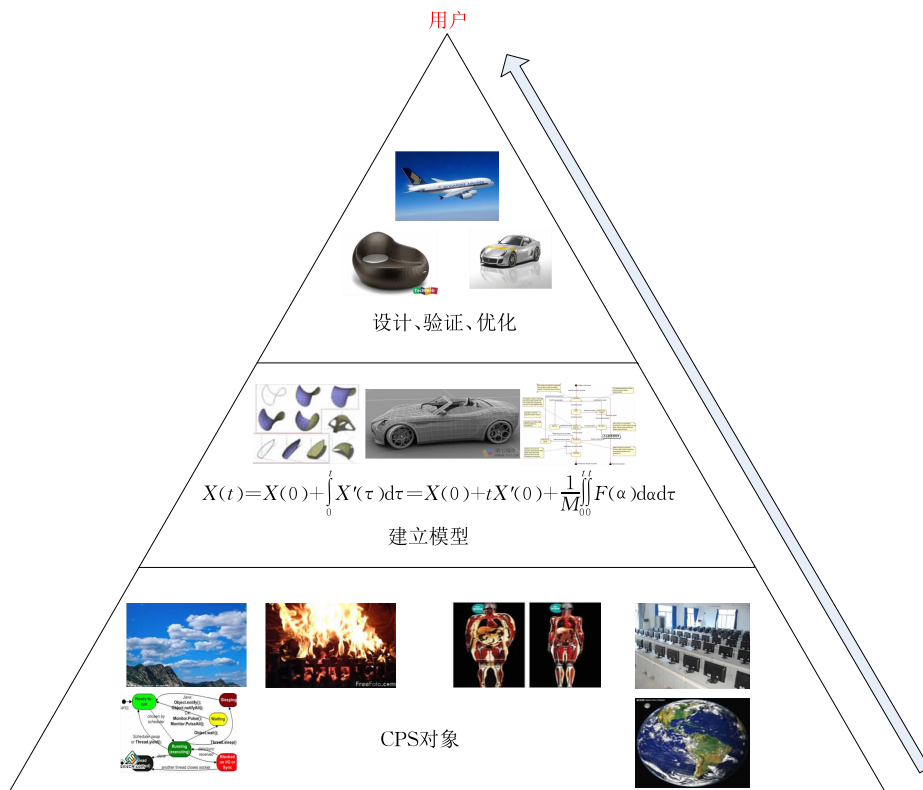


图 1 基于 MDA 的 CPS 设计流程

## 2 我们的主要工作

当前解决 CPS 的建模问题主要为发展新模型理论及在现有建模语言基础上扩展相关属性两种方式. 对于前者, 由于现有的计算机都采用冯诺依曼体系架构, 几乎所有的编程语言都缺乏对于时间及并发的表示, 如我们坚持采用这一发展模式, 势必打破现有的计算机体系架构, 由此而带来的问题是非常多的; 对于后者, 众多的研究学者主要关注与已有建模语言的结合, 在 PETRI 网、自动机、Ptolemy 等建模语言中添加相关因子, 这一实现方式并没有从根本上解决离散-连续问题, 且扩展之后并没有一个统一的标准. 本文我们提出将 CPS 的建模问题转化为离散-连续问题, 建立模型之前, 我们并不需要了解 CPS 系统所面临的环境如何复杂, 也不需要分析系统即将经历的状态, 这些在现有建模语言中是必须且分析过程冗繁而又复杂的. 从系统所经历的真实物理环境出发, 测试获得系统属性的真实数据及相关限制, 这样既能真实反映系统的运行状况又较容易获得, 在此基础上建立系统连续模型从而作进一步的分析处理. 我们的方法解决了常规的 CPS 建模中离散连续系统分开处理的方式, 提出

了一种新的建模分析方法, 为相关的研究人员提供一种新的思路:

(1) 离散-连续-离散建模方法. 将其应用到 CPS 的建模问题中能解决离散连续系统分开处理之后合成的问题, 且问题的本质最后又回到离散系统, 这方便我们应用计算机科学的相关知识加以分析处理.

(2) 客观式建模. 对于 CPS 系统, 我们不需分析系统将要经历的状态也不知系统所处的物理环境, 却通过真实的数据更加准确地反映系统本身. 该建模方法完全根据真实的数据建模, 客观真实地反映系统本身.

(3) 反馈控制. 为了减少离散系统与连续系统之间的误差, 我们实现了它们之间的反馈控制. 通过传感器等工具获得真实的测试数据, 从数据出发建立系统连续模型, 以连续模型为基准调整离散数据, 然后重新建立系统的连续模型, 通过多次反馈控制及相互限制, 在一定范围内控制离散连续模型之间的误差.

在复杂的 CPS 环境中, 相对于准确估计系统将要经历的状态, 通过工具获得大量的真实数据是简单易行的, 而且提出的方法通过两模型之间的反馈控制不断缩小误差, 使得建立的模型更加精确.

### 3 相关研究

随着 CPS 对工业进步及社会发展重要性影响的日益增强,2007 年,美国总统科学技术顾问委员会将 CPS 列为国家战略发展的顶级研究领域.在此之后,欧洲各发达国家也陆续将大量人力物力投入到 CPS 的研究中来.解决 CPS 的关键问题在于如何建立准确的模型,在此之前已有大量的研究工作存在(见总结性文献[9-10]),PETRI 网、UML、时间自动机、AADL 等建模语言发展迅速也大量应用到 CPS 的研究分析中.文献[11]在面向对象 PETRI 网的基础上提出了针对 CPS 建模的专有模型,扩展描述了传感器节点模型、执行器模型及控制器模型,该模型缩小了系统的状态空间,但与面向对象的编程语言一样并没有从本质上解决时间所带来的影响.文献[12]在随机 PETRI 网的基础上提出了一种概率模型用来描述 CPS 系统中存在的恶意节点,并对 CPS 中的攻击入侵行为做检测分析,利用响应系统(IDRS)来实时回应运行时的恶意事件,该模型主要侧重验证系统的可靠性,对 CPS 其他性能的分析需进一步的扩展研究.文献[13]在 UML、MDA 及面向目标设计语言的基础上提出了一种 CPS 模型用来检测石油的分布情况,该方法提高了石油控制应用的开发效率,更好地保证了发生石油泄漏时的应用控制.这一模型定义复杂,在 CPS 服务模型中需包含资源模型的类型及定义、资源利用、资源约束、资源到服务的映射等等,并集成资源模型,所有这些在模型的建立过程中显得复杂而又难以实现.文献[14]在文献[13]的基础上解决了对资源形式化表达的相关问题,描述了一个上下文相关的以资源为中心的 CPS 模型.

除了以上对现有模型语言的进一步研究,一些新的建模语言也得到了广泛应用.文献[15]提出一种扩展的 AADL 语言用来描述 CPS 系统(CPSADL),该语言将带有上下文环境组件及控制组件的物理实体添加到物理事件的描述及交互过程中,所有的物理组件都包含连续的端口,在 PDA (Push-Down Automata)的基础上对设计模型的实时性进行验证.文献[16]在此基础上对 CPS 模型的流延迟特性作进一步的分析,并将其转换为 PDA 来评估模拟其延迟特性.这一模型的提出实现了对物理事件的分析验证,但缺少连续系统与信息系统之间的真正交互.混合系统理论<sup>[17-18]</sup>作为 CPS

的根本依据,在 CPS 的建模过程中也发挥了重要作用.文献[19]改变了传统的混合系统中无限的时间范围内都做离线模型检测的做法,定期对混合系统模型做在线模型检测,每一个时间段,都会在一个固定的周期范围内对混合系统模型进行检查.这一方法虽将模型检测从离线提到在线范围内,但由于设定的是周期检查,并没有真正意义上的实现系统检测的实时性.另外文献[20]使用混合自动机的形式化理论描述分析 CPS 的设计模式以保证系统在有通信故障的情况下的安全性.文献[21]在混合系统理论的基础上设计和实现了一个称为 HybridSim 的集成建模及仿真工具链,用来对 CPS 进行建模及分析.

以上提出的所有 CPS 建模方法,在建立模型之前,研究人员都需掌握其运行环境及所经状态.另外有一些方法则从不同的角度出发,文献[22-23]使用经验数据用来对 CPS 通信过程中的主要属性(自相似、非平稳性)进行描述,从 CPS 所具有的这些复杂特性出发,提出了一种与分形几何相似的统计物理启发式模型,根据系统特征从数据中找出普遍规律并建立模型.CPS 这一根据数据建模的方式,不失为一个好的研究方向.

### 4 一种面向 CPS 的建模机制

建模问题成了 CPS 发展的瓶颈所在<sup>[24-27]</sup>.但不容忽视的是,随着 CPS 复杂度的增加,面临的环境日益复杂,CPS 系统将要经历的状态、系统将在何时完成、会给外部环境带来或受其怎样的影响,所有这些都是难以预测的.不可否认的是,传统的建模方法在建立系统模型之前需取得一些先验知识<sup>[18,28-29]</sup>,例如,加速度等于物体各个方向所受力的矢量和除以物体的质量,而速度则是加速度的积分:

加速度:

$$x''(t) = \frac{F(t) + F^+(t) - F^-(t)}{M}.$$

速度:

$$\forall t > 0, x'(t) = X'(0) + \frac{1}{M} \int_0^t (F(\tau) + F^+(\tau) - F^-(\tau)) d\tau.$$

在获得速度之前,需事先计算物体所受各个方向的力的矢量和  $F(t)$ ,所受正方向的力  $F^+(t)$  包括顺风力、机械动力、推力等等,而负方向的力  $F^-(t)$  则包括路面阻力、摩擦力、其他与物体前进方向相反的力等等,另外计算速度也需计算物体的初始加速

度及质量的大小,所有这些在复杂的物理环境中都难以获得.描述系统的状态也是如此,很难限定系统何时开始何时结束,甚至都难以保证系统在这种模式下能准确的执行.我们提出的方法与之前的建模方法不同,不需预计系统将要经历的状态而从一系列测试获得的离散数据出发建立系统的连续模型,以汽车信息-物理融合系统(ACPS)为例,图 2 给出了建模方法的总体流程:首先,汽车在行驶过程中,通过外部设备如传感器、网络等测量获得大量的离散数据,这些数据看起来是杂乱无章且没有规律的;在特征提取这一环节通过大量的经验分析及数学方

法,去掉大部分的冗余数据,并提取出一些有用的能反映系统属性本质特征的数据;接着,从留下的有用的离散数据出发根据一定的数学方法(方法将在接下来的章节详细介绍)建立系统的连续模型;最后将连续模型与离散模型类比,反过来调节控制离散数据,若它们之间存在较大误差,则修改影响因子较大的离散数据后重新建立系统连续模型,直到两者之间的误差小于一定的阈值.对于非功能性属性的验证,也可以从获得的离散数据出发并通过反馈控制加以实现,本方法主要讨论 CPS 的建模问题,集中于处理离散-连续-离散模型.

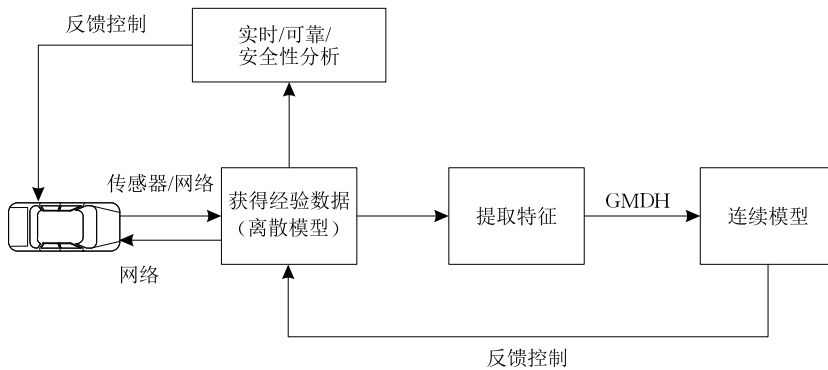


图 2 CPS 中一种新的建模机制

## 5 建模方法及其改进

### 5.1 离散-连续建模方法

从离散数据出发建立系统连续模型有很多方法可以选择,其中数据处理的分组方法(GMDH)能快速建立系统模型,具有自组织和全局选优等特性,非常适合非线性数据的拟合<sup>[30-31]</sup>.在测量获得大量非线性的离散数据之后,将分析对象作为因变量,其他因子作为自变量,GMDH 主要分 3 步来建立系统连续模型:

#### (1) 回归计算

一个离散数据构成的矩阵(CPS 系统测试数据)如表 1 所示.

表 1 离散数据

$y_1$	$x_{11}$	$x_{12}$	...	$x_{1m}$
$y_2$	$x_{21}$	$x_{22}$	...	$x_{2m}$
$y_3$	$x_{31}$	$x_{32}$	...	$x_{3m}$
...	...	...	...	...
$y_{nt}$	$x_{nt,1}$	$x_{nt,2}$	...	$x_{nt,m}$
...	...	...	...	...
$y_n$	$x_{n1}$	$x_{n2}$	...	$x_{nm}$

设  $y$  为因变量(分析属性(a)), $x$  为自变量(影响因子(b)),该矩阵共包含  $m$  个自变量及  $n$  个数据点,且将其分为训练样本(A)和测试样本(B)两个部分,其中训练样本包含  $nt$  个数据集.对每一对  $x_i$  和  $x_j$  作如下多项式回归处理:

$$y = A + Bx_i + Cx_j + Dx_i^2 + Ex_j^2 + Fx_ix_j, \\ i \neq j, 1 \leq (i, j) \leq m.$$

多项式回归处理之后,可以得到  $k = \binom{m}{2} = m(m-1)/2$  个更高阶的回归多项式,设:

$$z = a_0 + b_0x_i + c_0x_j + d_0x_i^2 + e_0x_j^2 + f_0x_ix_j.$$

那么(b)部分在数据拟合后形成一个新的矩阵(表 2).

表 2 拟合后自变量构成的新矩阵

$z_1$	$z_2$	...	$z_k$
$z_{11}(x_{11}, x_{12})$	$z_{12}(x_{11}, x_{13})$	...	$z_{1k}(x_{1(m-1)}, x_{1m})$
$z_{21}(x_{21}, x_{22})$	$z_{22}(x_{21}, x_{23})$	...	$z_{2k}(x_{2(m-1)}, x_{2m})$
...	...	...	...
$z_{nt,1}(x_{nt,1}, x_{nt,2})$	$z_{nt,2}(x_{nt,1}, x_{nt,3})$	...	$z_{nt,k}(x_{nt,(m-1)}, x_{nt,m})$

#### (2) 替代计算

将  $x$  用  $z$  来替换

使用测试样本(B)中的因变量  $y$  计算如下均方

根值(RMS),将  $r_j$  作为迭代结束的阈值.

$$r_j = \left[ \frac{\sum_{i=nt+1}^n (y_i - z_{ij})^2}{\sum_{i=nt+1}^n y_i^2} \right]^{\frac{1}{2}}, j = 1, 2, \dots, k.$$

从数据集(表 1)可以看出测试样本(B)为下标  $nt+1$  到  $n$  的  $n-nt-1$  条数据,在计算得出均方根值  $r_j$  之后,选定一个经验值  $L$ ,删除满足条件  $r_j \geq L$  对应的列,假设有  $k'$  列满足,将这  $k'$  列数据标记为  $Z_k$ ,将剩下的  $k'$  列数据替代数据集中(b)部分,得到如下新的矩阵(表 3),且满足  $k < m$ :

表 3 新数据集

$y$	$z_1, z_2, z_3, \dots, z_k$

(3) 检验计算

在第 2 步中取最小的  $r_j$  值,将其标记为  $R_{\min}$ ,使用  $z$  替换  $x$  之后,重复第 1 步和第 2 步直至下次产生的  $R_{\min}$  值大于本次  $R_{\min}$ ,得到多项式如下:

$$\bar{y}_1 = a + \sum_{i=1}^m b_i x_{1i} + \sum_{i=1}^m \sum_{j=1}^m c_{ij} x_{1i} x_{1j} + \dots$$

$$\bar{y}_2 = a + \sum_{i=1}^m b_i x_{2i} + \sum_{i=1}^m \sum_{j=1}^m c_{ij} x_{2i} x_{2j} + \dots$$

$$\bar{y}_n = a + \sum_{i=1}^m b_i x_{ni} + \sum_{i=1}^m \sum_{j=1}^m c_{ij} x_{ni} x_{nj} + \dots$$

为了得到最终的线性回归方程,在计算过程中需保存每一次迭代的回归系数.随着数据结构的发展,树结构成了较好的选择(图 3):假设在计算过程中,保存的系数树结构为图 4 所示.

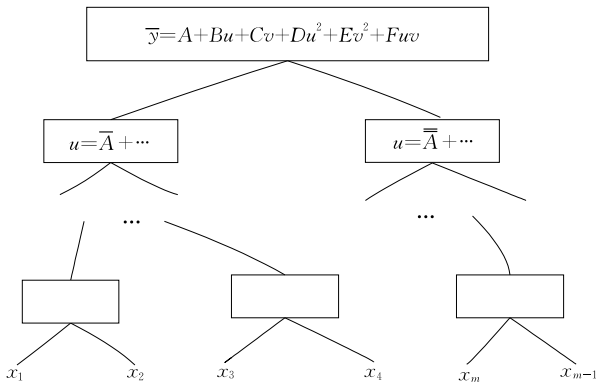


图 3 回归系数的保存结构

那么线性回归方程表示为

$$y = 1 + u_1 + 2u_2 + 2u_1^2 + 4u_2^2 + 2u_1 u_2,$$

其中

$$u_1 = 2 + 6x_1 + x_2 + 3x_1^2 + 2x_2^2 + x_1 x_2,$$

$$u_2 = 1 + x_4 + 2x_3^2 + 3x_4^2 + x_3 x_4.$$

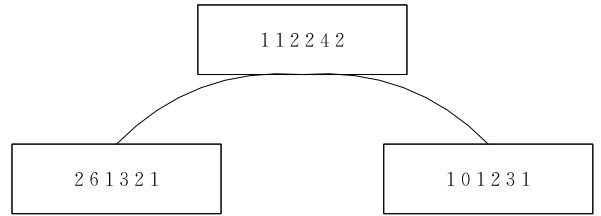


图 4 系数树结构

回归计算是从一个  $N$  阶矩阵中两两配对构建新的自变量,易知其时间复杂度为  $O(N^2)$ . 替代计算针对测试样本计算各个因变量的均方根值,其自变量的个数为  $C_N^2$ ,因此其时间复杂度为  $O(m \times C_N^2)$ ,假设  $L$  为总的迭代次数,  $F_j (j=1, 2, \dots, L)$  为每一次迭代所得的最优矩阵,易知其最大迭代次数不会超过  $N$ ,即  $L \leq N$ ,检验计算的时间复杂度为  $O(N \times m \times C_N^2)$ [32].

综上所述,从离散数据出发,通过 GMDH 非线性拟合方法建立了 CPS 系统的连续模型,但离散数据与连续模型之间不可能完全的一致,它们之间势必存在一定的误差,这就需对上述拟合方式提出改进,以实现模型的准确性.

5.2 建模方法的优化

模型优化通过离散数据与连续模型之间的反馈控制来实现,缩小两者之间的误差,使得所建立的模型更加精确.为了真正意义上减少离散系统与连续系统中存在的误差,模型优化从 GMDH 算法本身出发:第 1 次执行 GMDH 时,选定分析对象  $y$  作为因变量,其他属性作为自变量并建立关于  $y$  的多项式方程,在众多的自变量中存在一些对  $y$  值影响较大的因子.以汽车 CPS 系统中自适应巡航控制系统(ACC)为例,当分析汽车速度受其他属性影响之时(以速度作为因变量),加速度、气缸数、马力等因子比生产年限、品牌等属性更能影响当前汽车的速度.我们称这些对因变量影响较大的属性为强因子.第 1 次执行 GMDH 建立连续模型后,在众多的强因子中选择一个作为因变量,而之前的分析对象则作为自变量(例如:加速度为  $y$ ,速度为  $x_1$ ),再次建立系统的连续模型,将旧因变量的值(速度)代入新的多项式方程,计算新因变量值(加速度)与原始值对比并加以修改.重复上述操作,并选定新的强因子作为因变量,继续对比修改,直到所建立的连续模型与初始的测量数据之间的误差小于一定的阈值,模型优化算法流程如图 5 所示.

假设计算所得新因变量的值(速度)为  $y_c$ ,离散

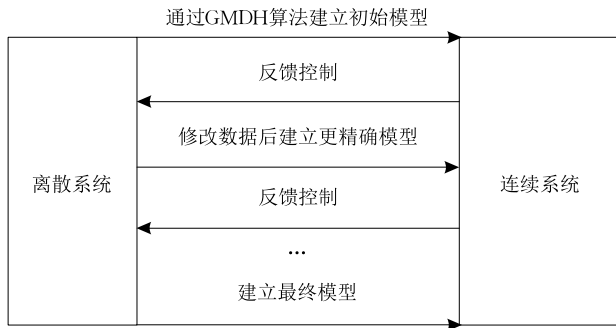


图 5 反馈控制及模型优化

数据中存在的因变量(速度)初始值为  $y_o, y_c, y_o$  为  $m$  维空间中的两点,判断其相似性最直接的方式是计算其欧式距离:

$$d = \sqrt{(y_{o1} - y_{c1})^2 + (y_{o2} - y_{c2})^2 + \dots + (y_{om} - y_{cm})^2} < T_s.$$

迭代终止条件为  $y_o$  与  $y_c$  之间的欧式距离小于一定的阈值  $T_s$ ,该阈值的设定既可以根据经验值判断,也可以为了达到某一属性或要求而设定.由于欧氏距离的计算与数据集的选择有关,在实验过程中我们可根据运行时间及参数的变化而选择一定的经验值.

图 6 给出了离散-连续-离散建模方法及优化过程(DCDM),它通过不断的反馈控制使得模型更加精确.随着 CPS 复杂度的增加,分析物理环境及所处状态是非常复杂的,从数据出发建立系统模型是一个新的研究方向,特别在大数据时代环境背景下,通过数据分析系统的相关属性也变得越来越重要.上述建模方法不需主观判断系统如何运行,也不需估计预测其状态导向,而是实事求是的从数据出

发,客观式建模并实现了离散与连续的深度融合.

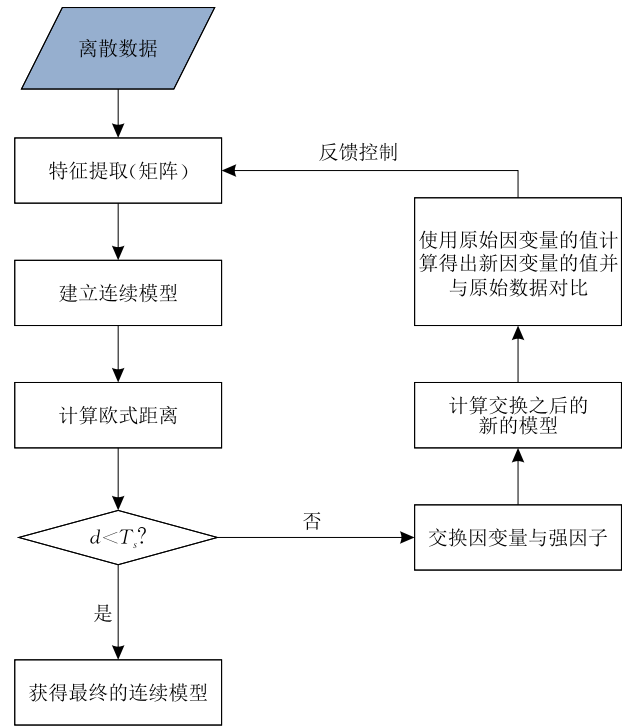


图 6 DCDM 整体流程

## 6 案例分析:方法的应用

我们提出的建模方法适用于处理大量的数据信息,实验选择一个包含 26 个属性因子的汽车实测数据集,它由加州大学欧文分校(UCI)在实际运行环境下测得,表 4 给出了数据集的一些相关属性.

表 4 实验数据集特征

数据集特征	实例数量	地区	数据类型	属性数量	网站点击次数	相关任务	有无丢失值
多因素分析	205	N/A	Categorical, integer, real	16	88847	建模	是

该数据集包含 3 种类型的实体:(1)在各种特征条件下汽车的自动规约值;(2)指定的风险等级;(3)与其他汽车相比其归一化的损耗值.其中第 2 个数据用来反应汽车的安全性能.在开始测试各数据之前会根据汽车价格为每一种类型初始化分配一个安全等级,如果该汽车测试得出其安全性能降低

(升高),则相应的增大(减小)其对应的风险等级,+3 表示汽车的安全性能最低,-3 则表示其安全系数最高.第 3 个实体则对应汽车每一年的平均损耗,这个值是根据汽车的不同类型(旅行/商务/私家车)做了归一化处理后的数据,表 5 给出了相关的属性信息.

表 5 实验数据集的属性<sup>①</sup>

symboling	Normalized-losses	Wheel-base	length	width	height
-3, -2, -1, 0, 1, 2, 3	from 65 to 256	From 86.6 to 120.9	From 141.1 to 208.1	From 60.3 to 72.3	From 47.8 to 59.8
Curb-weight	Engine-size	bore	stroke	Compression-ratio	horsepower
From 1488 to 4066	From 61 to 326	From 2.54 to 3.94	From 2.07 to 4.17	From 7 to 23	From 48 to 288
Peak-rpm	City-mpg	Highway-mpg	price		
From 4150 to 6600	From 13 to 49	From 16 to 54	From 5118 to 45400		

① <http://archive.ics.uci.edu/ml/datasets/Automobile>, 2003

实际上,原始数据集包含 26 个属性.但我们认为其中的 10 条对需要分析的对象(安全系数)影响因素较低可以忽略不计,因此被直接删除.在实验过程中,选取风险等级作为因变量,其他 15 个属性作为自变量构建系统的连续模型.在第 5 节建模方法中提到了经验值  $L$ ,如果  $L$  值设置过大,最终的模型中将包含较多的自变量,随着  $L$  的减小,最终模型中自变量的个数将减少,必须根据需求来设置正确的  $L$  值.为了不使问题变得复杂难以分析,算法规定在每一次迭代过程中  $L$  值的设定都使得模型不超过 4 对变量.图 7 给出了算法在每一次迭代过程中选择两对变量的 4 次迭代结果,其中  $u$  表示第 2 层迭代的参数.从图 7 可以看出第 1 次迭代后算法仅包含两对变量,所以接下来的一层只需用  $u_1$ 、 $u_2$  两个参数来分别表示第 1 对和第 2 对.以图 7 第 1 层

迭代作为例子,其连续模型可以表示为

$$y = -0.08176 + 0.86285u_1 + 0.40608u_2 + 0.08404u_1^2 + 0.03254u_2^2 - 0.20932u_1u_2,$$

其中

$$\begin{aligned} u_1 &= 58.029 - 0.01918x_1 - 1.0302x_2 + 0.00043x_1^2 + 0.00433x_2^2 - 3.63e^{-5}x_1x_2, \\ u_2 &= 83.647 - 1.274x_2 - 0.90374x_{13} + 0.00592x_2^2 + 0.00478x_{13}^2 - 0.00489x_2x_{13}. \end{aligned}$$

算术表达式中, $y$  代表汽车电子的“风险等级”, $x_1$ 、 $x_2$ 、 $x_3$  分别表示其“归一化损耗”、“轮距”、“车速”,算法在下一次的迭代过程中选择将当前分析因子影响较大的变量作为因变量,注意到  $x_2$  在多项式回归方程中出现的次数比  $x_1$ 、 $x_3$  要多,那么轮距将作为下次迭代需调整的变量继续其反馈控制.

error	0.570 87					
parameters( $u_1, u_2$ )	-0.081 76	0.862 85	0.406 08	0.084 04	0.032 54	-0.209 32
parameters( $x_1, x_2$ )	58.029	-0.019 18	-1.0302	0.000 43	0.004 33	-3.63E-05
parameters( $x_2, x_{13}$ )	83.647	-1.274	-0.903 74	0.005 92	0.004 78	0.004 89

↓

error	0.525 03					
parameters( $u_1, u_2$ )	0.054 11	1.1345	0.002 93	0.467 61	-0.386 12	-0.183 74
parameters( $x_1, x_3$ )	39.953	0.168 57	-0.893 29	-0.001 49	0.004 72	2.59E-005
parameters( $x_1, x_4$ )	3.0612	0.160 57	-0.1119	-0.000 63	-0.000 41	-8.18E-05

↓

error	0.520 68					
parameters( $u_1, u_2$ )	0.008 49	0.801 32	0.218 63	-0.485 92	0.184 63	0.280 76
parameters( $x_1, x_2$ )	49.425	0.048 14	-0.959 25	-0.000 46	0.004 52	0.000 11
parameters( $x_1, x_3$ )	11.415	0.046 95	-0.136 46	-0.000 14	0.000 32	5.27E-05

↓

error	0.519 13					
parameters( $u_1, u_2$ )	0.009 32	0.863 98	0.166 96	0.0743	-0.1345	0.034 75
parameters( $x_1, x_2$ )	47.614	0.062	-0.935 49	-0.000 61	0.004 48	0.000 11
parameters( $x_1, x_3$ )	7.8615	0.065 58	-0.106 41	-0.000 25	0.000 27	5.02E-05

图 7 两对参数的 4 次迭代结果

为了篇幅上的方便,图 8 仅给出 3 对参数时出现的两次迭代结果,其计算连续模型的方式与两对参数时类似.由于第 1 层包含了 3 对参数: $x_1$  和  $x_2$ ,  $x_{13}$  和  $x_2$ ,  $x_{14}$  和  $x_2$ , 因此第 2 层也需  $C_3^2 = 3$  个参数: $u_1$ 、 $u_2$ 、 $u_3$  来表示该层.由于首次迭代时,算法选择了相同的变量,如图 8 中的  $x_1$  和  $x_2$ ,  $x_{13}$  和  $x_2$  与图 7 中的选择相同,故第一次迭代的系数结果相同,其起始误差也与图 7 的相同,但是随着迭代的深入,通过反馈控制修改参数,属性的系数及产生的误差将变得

差距很大.

图 9 描述了每一次迭代过程中选择不同数量的参数对数对反馈控制带来的影响.由于迭代开始时每一次都是从未修改的原始数据出发,其初始误差是一致的.随着选择变量对数的增加,其误差减少速度越明显,这是由于选择更多的变量而更容易分散每一个变量所造成的影响,通过修改强因子的值而更能减少连续模型与离散数据之间的误差.另外从图也可以看出,前面几次迭代时,误差的缩小速度更



快,这是由于前几次迭代所选择的参数往往是更强的影响因子,而后面选择的修改参数对原始分析属性的影响越来越弱的缘故.注意到有时也存在特殊情况:后一次的迭代使得误差减少的速度快于前一

次,这是由于在一次迭代过程中,多项式可能包含多个参数,当两个参数出现次数相同而无法决定哪一个作为下一次迭代的修改参数时,算法采用的是一种随机选取机制,这就可能造成以上影响.

error	0.570 87					
parameters( $u_1, u_2$ )	-0.081 76	0.862 85	0.406 08	0.084 04	0.032 54	-0.209 32
parameters( $u_2, u_3$ )	-0.069 82	0.940 59	0.2685	0.009 84	0.107 86	-0.191 12
parameters( $x_1, x_2$ )	-0.031 59	2.241	-1.0998	2.0314	-0.6784	-1.3929
parameters( $x_1, x_2$ )	58.029	-0.019 18	-1.0302	0.000 43	0.004 33	-3.63E-005
parameters( $x_2, x_{13}$ )	83.647	-1.274	-0.903 74	0.005 92	0.004 78	0.004 89
parameters( $x_2, x_{14}$ )	98.612	-1.4763	-1.0452	0.006 97	0.004 66	0.005 57

error	0.305 62					
parameters( $u_1, u_2$ )	-0.077 12	1.1132	0.4245	0.7866	0.067 82	-0.0776
parameters( $u_1, u_3$ )	-0.054 37	1.0789	-0.8762	0.006 43	-0.7764	-1.2285
parameters( $u_2, u_3$ )	-0.030 18	1.9776	-1.1774	1.9986	-0.3347	-0.5731
parameters( $x_1, x_2$ )	57.067	-0.1091	-1.0873	0.000 32	0.001 72	-2.78E-04
parameters( $x_{13}, x_{14}$ )	67.891	-1.0065	-0.890 01	0.004 55	0.003 37	0.000 996
parameters( $x_{11}, x_{13}$ )	72.337	-1.0896	-1.0443	0.0009 75	0.004 69	0.003 94

图 8 3 对参数的两次迭代结果

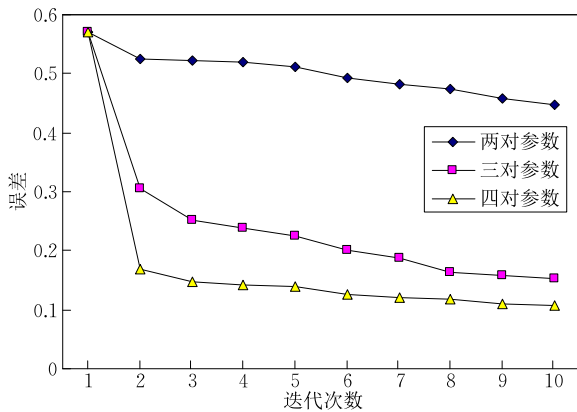


图 9 不同数量参数对的迭代结果

在相同的硬件环境下,运行我们的 DCDM 算法与几种经典的从离散到连续的建模算法<sup>[33]</sup>时间比较如表 6 所示.其中  $a, b, c, d$  代表采集的数据集个数分别为 189, 125, 95, 63.从表 6 可以看出当数据集较大时 SPD 与 DCDM 算法优势明显,其运行时间差不多只有 CAS 和 SP 算法的 1/10,数据集越大

表 6 算法时间(MS)比较

	$a$	$b$	$c$	$d$
CAS	616	459	451	431
SP	580	41	32	24
SPD	67	25	19	13
DCDM	46	27	17	11

优势越明显.随着数据集的减少,SP 算法逐渐与后两种算法运行时间接近.从总体上来说,我们的 DCDM 算法运行时间优于以上 3 种算法,且数据集越大优势越明显.

图 10 总结对比了不同算法误差值与迭代次数的关系.其中 CAS 算法起始误差值较大,随着迭代次数的增加误差呈现减小的趋势. DCDM 算法起始误差值较 SPD 大,这是因为我们的算法在第一次迭代时未调整相应的参数造成的,随着迭代次数的增加,DCDM 误差值下降明显且呈现较以上两种算法更优的误差值.

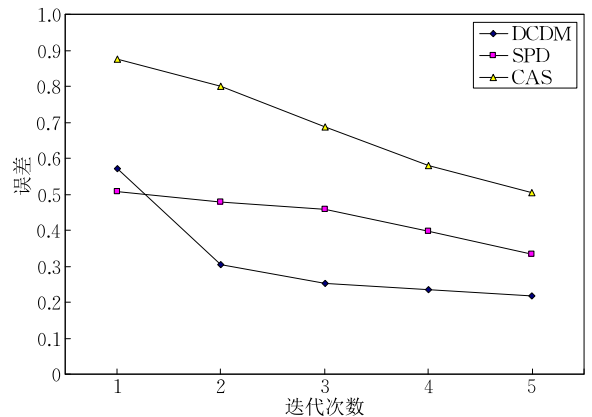


图 10 不同算法误差对比

## 7 总 结

为了研究 CPS 的建模问题,许多研究学者将其工作集中于系统的状态转变及分析.然而,在真实的物理环境中,系统的状态都是难以捕获且实时追踪的.因此,我们提出了一种新的建模机制,在这种机制下,不需估计系统的运行环境及所处状态,只需通过传感器捕获 CPS 系统的实时数据,从这一系列所获得的离散数据出发建立系统的连续模型.为了深度融合信息与物理系统,算法同时提出了一种反馈机制,实现离散数据与连续系统之间的相互控制.实验表明,我们的算法缩小了两者之间的误差,并建立了相关模型.另外,离散-连续-离散的建模机制使得物理事件的处理最终回归到计算科学可以处理的离散数据上来,实现了两者之间的深度融合.

**致 谢** 本文在湖南大学完成,感谢湖南大学嵌入式与网络计算省重点实验室,向本文的审稿专家致以深深的谢意!

## 参 考 文 献

- [1] Li Ren-Fa, Xie Yong, Li Rui, et al. Survey of cyber-physical systems. *Journal of Computer Research and Development*, 2012, 49(6): 1149-1161(in Chinese)  
(李仁发, 谢勇, 李蕊等. 信息-物理融合系统若干关键问题综述. *计算机研究与发展*, 2012, 49(6): 1149-1161)
- [2] Lee E A. CPS foundations//*Proceedings of the 47th Design Automation Conference*. New York, USA, 2010: 737-742
- [3] Hei X, Chang L, Ma W, et al. A safety framework and alarming model for train operation environment based on CPS//*Proceedings of the 2011 7th International Conference on Computational Intelligence and Security (CIS)*. Hainan, China, 2011: 1418-1421
- [4] Yang Zhi-Bin, Pi Lei, Hu Kai, et al. AADL: An architecture design and analysis language for complex embedded real-time systems. *Journal of Software*, 2010, 21(5): 899-915 (in Chinese)  
(杨志斌, 皮磊, 胡凯等. 复杂嵌入式实时系统体系结构设计与分析语言: AADL. *软件学报*, 2010, 21(5): 899-915)
- [5] Zhou Xing-She, Yang Ya-Lei, Yang Gang. Modeling methods for dynamic behaviors of cyber-physical system. *Chinese Journal of Computers*, 2014, 37(6): 1411-1423(in Chinese)  
(周兴社, 杨亚磊, 杨刚. 信息-物理融合系统动态行为模型构建方法. *计算机学报*, 2014, 37(6): 1411-1423)
- [6] Park J, Kang S, Chun I, et al. Goal modeling method for autonomic control of cyber-physical systems//*Proceedings of the 2014 16th International Conference on Advanced Communication Technology (ICACT)*. Pyeongchang, Korea, 2014: 295-299
- [7] Tang Jun-Jie, Zhao Jian-Jun, Ding Jian-Wan, et al. Cyber-physical systems modeling method based on modelica//*Proceedings of the 2012 6th International Conference on Software Security and Reliability Companion (SERE-C)*. Gaithersburg, USA, 2012: 188-191
- [8] Li-na C, Hong-bin H, Su D. Research on CPS spatio-temporal event model based on the state//*Proceedings of the 2011 6th International Conference on Computer Science & Education (ICCSE)*. Singapore, 2011: 195-198
- [9] Lee E A. Cyber physical systems: Design challenges//*Proceedings of the 11th IEEE International Symposium on Object Oriented Real-Time Distributed Computing (ISORC)*. Orlando, USA, 2008: 363-369
- [10] Derler P, Lee E A, Vincentelli A S. Modeling cyber-physical systems. *Proceedings of the IEEE*, 2012, 100(1): 13-28
- [11] Ma Zhi-Qiang, Fu Xiao, Yu Zhen-Hua. Object-oriented Petri nets based formal modeling for high-confidence cyber-physical systems//*Proceedings of the 2012 8th International Conference on Wireless Communications, Networking and Mobile Computing*. Shanghai, China, 2012: 1-4
- [12] Mitchell R, Chen I R. Effect of intrusion detection and response on reliability of cyber physical systems. *IEEE Transactions on Reliability*, 2013, 62(1): 199-210
- [13] Magureanu G, Gavrilescu M, Pescaru D, et al. Towards UML modeling of cyber-physical systems: A case study for gas distribution//*Proceedings of the 2010 8th International Symposium on Intelligent Systems and Informatics (SISY)*. Subotica, Yugoslavia, 2010: 471-476
- [14] Wan K, Alagar V. Resource modeling for cyber physical systems//*Proceedings of the 2012 International Conference on Systems and Informatics (ICSAD)*. Yantai, China, 2012: 2541-2546
- [15] Sun Z, Zhou X. Extending and recompiling AADL for CPS modeling//*Proceedings of the 2013 IEEE International Conference on Cyber, Physical and Social Computing*. Beijing, China, 2013: 1225-1230
- [16] Zhu Y, Dong Y, Ma C, et al. A methodology of model-based testing for AADL flow latency in CPS//*Proceedings of the 2011 5th International Conference on Secure Software Integration & Reliability Improvement Companion (SSIRI-C)*. Jeju Island, Korea, 2011: 99-105
- [17] Li Ye-Jing, Chen Ming-Cai, Zhang Guang-Quan, et al. A model for vehicular cyber-physical system based on extended hybrid automaton//*Proceedings of the 2013 8th International Conference on Computer Science & Education (ICCSE)*. Colombo, Sri Lanka, 2013: 1305-1308
- [18] Yalei Y, Xingshe Z. Cyber-physical systems modeling based on extended hybrid automata//*Proceedings of the 2013 5th International Conference on Computational and Information Sciences (ICCIS)*. Shiyang, China, 2013: 1871-1874

- [19] Li T, Tan F, Wang Q, et al. From offline toward real-time; A hybrid systems model checking and CPS co-design approach for Medical Device Plug-and-Play (MDPnP). *IEEE Transactions on Parallel and Distributed Systems*, 2014, 25(3): 642-652
- [20] Tan F, Wang Y, Wang Q, et al. Guaranteeing proper-temporal-embedding safety rules in wireless CPS: A hybrid formal modeling approach//Proceedings of the 2013 43rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). Budapest, Hungary, 2013; 1-12
- [21] Wang B, Baras J S. HybridSim: A modeling and co-simulation toolchain for cyber-physical systems//Proceedings of the 2013 IEEE/ACM 17th International Symposium on Distributed Simulation and Real Time Applications (DS-RT). Delft, Holland, 2013; 33-40
- [22] Bogdan P, Marculescu R. Towards a science of cyber-physical systems design//Proceedings of the 2nd International Conference on Cyber-Physical Systems. Chicago, USA, 2011; 99-108
- [23] Bogdan P, Marculescu R. Cyber-physical systems: Workload modeling and design optimization. *IEEE Design & Test of Computers*, 2011, 28(4): 78-87
- [24] Zhang Zhen-Kai, Eyisi E, Karsai G, et al. Co-simulation framework for design of time-triggered cyber physical systems//Proceedings of the 2013 ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS). Philadelphia, USA, 2013; 119-128
- [25] Moreno J, Damm M, Haase J, et al. Unified and comprehensive electronic system level, network and physics simulation for wirelessly networked cyber physical systems//Proceedings of the 2012 Forum on Specification and Design Languages (FDL). Vienna, Austria, 2012; 68-74
- [26] Mueller W, Becker M, Elfeky A, et al. Virtual prototyping of cyber-physical systems//Proceedings of the 2012 17th Asia and South Pacific Design Automation Conference (ASP-DAC). Sydney, Australia, 2012; 219-226
- [27] Liu J, Liu X, Koo T K J, et al. A hierarchical hybrid system model and its simulation//Proceedings of the 38th IEEE Conference on Decision and Control. Phoenix, USA, 1999, 4; 3508-3513
- [28] Banerjee A, Gupta S K S. Spatio-temporal hybrid automata for safe cyber-physical systems; A medical case study//Proceedings of the 2013 ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS). Philadelphia, USA, 2013; 71-80
- [29] Liu J, Lee E A. Component-based hierarchical modeling of systems with continuous and discrete dynamics//Proceedings of the International Symposium on Computer-Aided Control System Design. Anchorage, USA, 2000; 95-100
- [30] Kondo T, Ueno J, Takao S. Hybrid feedback GMDH-type neural network self-selecting various neurons and its application to medical image diagnosis of lung cancer//Proceedings of the 2012 Joint 6 International Conference on Advanced Intelligent Systems (ISIS). Kobe, Japan, 2012; 1925-1930
- [31] Kondo T, Pandya A S. GMDH-type neural networks with a feedback loop and their application to the identification of large-spatial air pollution patterns//Proceedings of the 39th SICE Annual Conference. Lizuka, Japan, 2000; 19-24
- [32] Xiao Jin, He Chang-Zheng. Dynamic classifier ensemble selection based on GMDH//Proceedings of the 2009 International Joint Conference on Computational Science and Optimization. Perugia, Italy, 2009; 731-734
- [33] Mansourifar H, Bastanfard A. Fast conic spline data fitting of noise-free data points//Proceedings of the 2011 8th International Conference Computer Graphics, Image and Visualization. Innsbruck, Austria, 2011; 59-64



**YANG Fan**, born in 1985, Ph. D. candidate. His main research interests include embedded system and computer architecture.

**LIU Yan**, born in 1979, Ph. D. His main research interests focus on embedded system.

**LI Ren-Fa**, born in 1956, Ph. D., professor, Ph. D.

supervisor. His main research interests include computer architecture, embedded system and network.

**DUAN Meng-Qin**, born in 1989, M. S. candidate. Her main research interests focus on embedded system.

**XIE Guo-Qi**, born in 1983, Ph. D. His main research interests include computer architecture and fault-tolerant computing.

**HUANG Jing**, born in 1986, Ph. D. candidate. His main research interest is embedded system.

## Background

Cyber-Physical System (CPS) has attracted attention of the researchers all over the world. Cyber contains the information consisting of computing system and network

which are involved in discrete computing process, logic communication and feedback control, Cyber system focuses on the problem of computer science and information profession,

and its core concept is discrete. On the other hand, physical system has to face the real existed whole world, the mountain, the water, weather, sunshine, chemical reactions, biological genetic, environmental change and so on. It represents the process, objects and things of the world, all of which change over time. The main problem of physical system is how to deal the continuity. So how to design a timed, reliable and secure CPS must become the main problem in this field. The design method of traditional embedded system uses code to realize and analyze the system directly, which has to modify all the process during the development when an error occurs, it has to waste manpower and material resources, system design though code development method has been out-of-date based its weakness.

In order to solve the modeling problem of CPS, many researchers focused on the states evaluating, but in the real

situation, with the complexity of physical environment, the states are difficulty to master. In this paper, we proposed a new modeling method, which didn't need to grasp the running situation of system and but began from the discrete data which is easy to be get through sensor, network and other tools, in order to build a continuous modeling. Then, it must exist error between the discrete data and continuous equation, through feedback control, our experiment results showed that our method decreased the relative error effectively, furthermore, our algorithm realized the connection between the discrete system and continuous system, which made the mutual control between them become possible.

Under the support of National Natural Science Foundation of China under Grant No. 61173036. The authors are doing relevant works on the modeling method and verification of CPS.