系统辨识方法总结2

驱动组 王彬 2016.11.11

# 1:辨识概述

辨识就是通过测量的输入和输出反推得到动态系统的数学模型。

辨识的步骤

1：获取输入输出数据

2：选择一个模型（根据经验）

3：应用一种估计方法来获取模型中的可调参数

4：评估估计的模型是否满足要求。

## 1.1：模型

动态模型定义：输出不仅和输入有关还和过去的输出有关系。

动态模型描述方法：差分方程，传递函数，状态空间表达，零极点增益模型

模型分为：连续模型和离散模型。

## 1.2：数据

数据很关键，模型估计的好坏直接和数据是否能反应系统特性有直接关系。

可使用时域数据或者是频域数据。

频域数据是时域数据进行傅里叶变换得到的数据。也代表了系统的频率响应。

数据需要有足够的精度和足够的宽度来保证获取到系统的动态特性。

1：需要有一个好的激励信号

2：测量足够长的时间

3：采样系统有足够好的信噪比

4：采样精度要满足要求。

在辨识之前需要将信号处理好。分析信号在系统带宽内是否有足够的power。

## 1.3：从数据获取模型

怎么选择模型？

1：尽可能选择简单的模型，然后通过测量数据进行验证，这种方法称为黑盒子建模。

2：根据先验信息知道模型的结构，但是不知道具体数值，那么可以先建模，然后估计出其中的变量，这种称为灰盒子建模。

估计模型的原则：最小化输出值和测量值。估计算法就是尽可能的使得这个值最小。

仿真响应：对给定输入，从估计模型得到输出

预测响应：对给定输入和过去的测量输出。得到的响应。

最常用的方法就是迭代。可以配置迭代算法和迭代次数等。标准也可以进行配置，例如关心的频率部分。

### 1.3.1：黑盒子建模

当你主要关注于匹配数据而不在意具体的数学模型的时候，使用该方法是很有用的。系统有线性和非线性之分。分别有不同的方法来处理。黑盒子模型主要考虑系统的阶数即可。

黑盒子建模一般是试错的过程。估计不同的结构然后比较其结果。一般从线性模型开始。假如简单的结构不能产生满意的结果，那么就需要更复杂的模型，

1：需要的阶数会更高。

2：建立噪声的模型。

3：使用非线性模型。

什么时候使用非线性模型？

1：线性模型不能满足要求。如果线性可以满足，那么就选择最简单的线性模型。

2：假如你发现对于输入等级和输入方向，响应不同，那么试着使用非线性模型。例如对于阶跃上升和阶跃下降，响应非常不同，那么就要考虑非线性模型。

3：如果系统有多个输入，那么输入和输出之间是非线性的，那么首先可以转化输入，例如将输入相乘，乘完的值和输出之间是否是线性关系。如果是，那么就不需要建立非线性的模型。如果不是，那么再使用Nonlinear ARX or Hammerstein-Wiener models。

### 1.3.2：灰盒子建模

有些时候你已经知道了系统的模型，只是不知道参数，这时候使用灰盒子就比较好。可以根据输入输出数据来决定未知参数。

一般步骤如下

1：创建一个模型

2：配置模型的初始值和限制值

3：使用估计方法来计算得到未知参数。

## 1.4：评估模型质量

假如你没有找到合适的模型，你需要不停的改变模型结果，改变估计算法，或者进行额外的数据处理，如果这些都没有用，那么你可能需要重新设计实验和数据获取过程。

几种评价方法：

1：对比模型输出和实际的测试输出。主要是计算出两条曲线的匹配度（自相关?）

2：残差分析。

3：分析模型不确定性。

### 1.4.1：残差分析

一阶预测输出和验证用测量输出之间的差值，称为残差，因此，残差表示验证数据在模型中没有被表示的部分。

残差分析有两种测试：1：白化测试。2：独立测试。

根据白化测试原则，好的模型的残差的自相关函数要在置信区间中，表明残差是不相关的。

根据独立测试原则，好的模型的残差和之前的输入不相关。相关表明了模型不能够描述哪一部分输出对应相应的输入。例如在k个滞后的峰值（超出置信区间）意味着对于输入U（t-k），输出y(t)不能够被模型很好的描述。

一般需要满足这两个原则：除非有以下情况

1：对于output-error OE模型和instrumental-variable(IV)模型，确定模型对于u和e是独立的。并且对于e的结果不在意。这种情况模型关注的是动态而不是扰动特性。

2：残差和输入之间是负lag，也就是超前，不能说这个模型就是不准确的模型。当现在的留数影响了将来的输入，那么可能是系统有反馈，当系统中有反馈的时候，在互相关验证中主要关注正的lag。

对于时域数据来说，验证两部分

1：每个输出残差的自相关。

2：输入和残差的互相关。

置信区间：就是对应一个区间，在这个区间中残差对于系统来说影响不大。Matlab使用模型参数的不确定性来计算置信区间并假设估计具有高斯分布。例如对于95%的置信区间，在零附近的区间代表残存值对于95%的区间都没有很大影响。你可以使用高斯分布的标准差（0-1）来确定置信区间。

### 1.4.2：参数不确定性

当你从数据估计模型参数的时候，你会在置信区间中获取标称值。这个区间大小由估计过程中参数不确定性来决定。不确定性的幅值大小是模型是否可靠的测量方法。大的参数不确定性来自几个方面，比如高阶模型，不适当的输入数据，或者是测量数据的不好的信噪比。

参数不确定可以帮助我们理解当用不同的数据输入估计的时候，模型的参数有多不同。

什么是模型协方差？

上述的模型不确定性就称为模型的协方差。当你估计一个模型的时候，估计参数的协方差矩阵和模型一起存储。

计算协方差矩阵是基于模型结构可以正确的表达系统的动态这一前提。对于模型中包含扰动模型H的系统，正确的不确定性估计是认为系统产生白噪声的残差。为了决定是否你能够相信估计模型的不确定性的值，你可以进行残差测试。假如可以通过残差测试，那么认为系统在置信区间内，并且任何参数不确定性来自输出的随机扰动。

## 1.5：系统辨识基本流程

系统辨识是迭代的过程。

1：准备数据，预处理，例如去平均，插值丢失的值，滤波来关注具体的频域段。或者重新采样（插值或者抽取）用不同的时间间隔。画图数据图形观察数据在时域和频域的特征。

2：识别模型，有很多种方式，例如传递函数，多项式，状态空间等

3：验证模型，当不满足结果的时候使用不同的算法和模型进行实验。有时候引入噪声模型可以提高模型的准确性。

4：后处理，例如连续和离散转化，不同模型之间转化，连接合并模型，线性化等

5：用辨识的模型仿真，预测，做控制设计

## 1.6：在线估计

在线估计算法是在系统操作过程中当有新的数据来的时候估计模型的参数。在线估计适用于已知工作点并且参数变化范围小的场合。在线估计典型的使用迭代的算法，为了获取某一时刻的值，需要当前的测量和之前估计的参数。因此，迭代算法在空间存储上是高效的。需要的计算量也小。适合于在线和嵌入式应用。

在线估计主要用于以下几种应用场合：

1：自适应控制，基于平台模型的改变来修改控制器。

2：误差检测，用在线模型和理想的模型进行比较。

3：无传感，基于估计的模型产生虚拟的测量值，用于反馈控制或者误差检测。

4：在离线估计之前验证实验数据的质量，在线估计可以通过几次迭代快速验证是否数据可以捕获到系统的动态，而离线估计通常比较耗时。

在matlab工具箱中只能够用命令行方式或者simulink方式实现在线估计。Simulink中有[Recursive Least Squares Estimator](file:///D:\matlab2016a\help\ident\ref\recursiveleastsquaresestimator.html) and [Recursive Polynomial Model Estimator](file:///D:\matlab2016a\help\ident\ref\recursivepolynomialmodelestimator.html) blocks。可以产生代码，命令行也几乎一样。

在线估计的模型必须是离散的线性或者是近似线性的。在估计过程中必须固定模型。

## 1.7：数据的处理

1：数据选择（部分频率或者是子集）（软件实现）

2：分析数据，做频域或者是阶跃响应看delay，激励等

3：预处理 detrend(去平均)，idfilt通带滤波，diff信号差分 misdata重建丢失的数据

nkshift移位，idresample重新采样通过抽取和插值的方式。

### 1.7.1：异常值

失效故障的时候会在测量数据中产生错误，称为异常值，异常值通常由信号的尖峰或者失效造成。假如不移除会影响估计的模型。为了识别这种情况发生，先画出数据的图形进行观测，模型估计完成后做残差分析看是否由很大的值。还可以通过将数据分割为很多段形成多组实验数据可以减小异常值的影响。但是数据组也不能太小。还可以通过滤波来去掉异常值，一般异常值都是突然的变化。

### 1.7.2：去均值

消除趋势是从时域采样输入-输出数据信号中删除、抵消线性趋势。这个数据处理操作可以帮助你估计更精确的线性模型,因为线性模型不能捕获任意输入和输出信号之间的差异水平。从去趋势数据估计的线性模型描述了输入信号和输出信号的变化之间的关系。

当趋势在实验过程中经常变化时，需要移除多段的趋势，，信号的漂移认为是一种低频的扰动，会导致不稳定的模型。

以下情况不需要去趋势

1：线性模型需要offset来描述系统的动态，例如系统中由积分的时候不需要去掉趋势

2：非线性黑盒子模型，也不需要。

3：非线性灰盒子模型。

### 1.7.3：采样（抽取或者插值）

通过一个抗混跌低通滤波器和通过抽取和插值来改变采样率，实现重新采样。

假如数据采样比需要的更快，可以通过抽取的方式，不会损失精度，假如数据采样比需要的还慢，可以会丢失高频的动态信息，尽管可以通过插值，但是测量点之间的值并不是新的测量信息，所以最好还是以需要的采样速度进行采样。

matlab抽取不会造成混跌，通常我们只希望通过抽取将高频的噪声去除掉，从而保存信号的能量，它是通过一个因子T来归一化频谱并在抽取后保存信号能量。假如你手动抽取，那么高频的能量会折叠到低频（称为混跌）因为总的信号能量保持不变，而不得不挤入一个小的频率范围。频谱的幅值会增加。

### 1.7.4：滤波

Matlab使用的是5阶带通bufferworth滤波器，滤波器可以去掉高频的噪声和低频的扰动。可以使得你的模型在你关心的频率区域。一般在带宽附近。对于线性系统，输入和输出的滤波不会改变线性系统的关系。但是确实影响噪声特性和估计的模型。为了得到一个可靠的噪声模型，最好不要进行滤波。如果用滤波的数据，那么只能够得到模型的动态，如果要获取噪声的模型，那么就用未滤波的数据。