

决策系统中统计模型建模方法的研究

沈沧海 (中国科学院计算技术研究所 100080)

摘要:本文提出一个具有实用价值的统计模型的快速建模方法,该方法通过已知的统计数据作了曲线与相关的一些曲线方程进行拟合,求出它们之间拟合误差最小的曲线方程作为该统计数据的统计模型。

关键词:决策系统 模型库 统计模型

一、引言

一个系统中存在一批数据,而要通过它描绘一个系统时,往往用一个解析数学式,并且这个解析式要有解,但由于许多系统的信息是不连续的,常常是些离散的点,例如市场调查的结果,试验数据等,为了找到解析表达式,需要对原始数据作统计分析,回归分析和曲线拟合等工作来建立数学模型,而本文采用曲线拟合的方法来建模。

现实生活中某个应用领域中存在一批常用的数学模型,我们选用一些与已知统计数据相匹配的模型放入模型库,用最小二乘法求其拟合误差最小的曲线方程作为该统计数据的数学模型,它以模型库为基础,通过对误差的计算和分析,可以最优地拟合一组数据所逼近的方程。

二、统计模型建模方法

1. 建立模型库

把一些与已知的统计数据的曲线相近的曲线方程放入模型库,在构造该模型库过程中充分考虑如下几个因素:

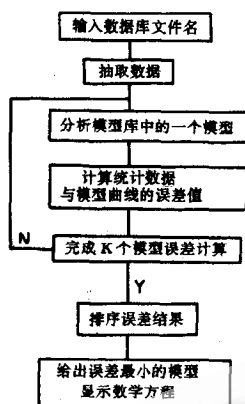
- 统计的数据是单调递增还是单调递减,是有一个还是有几个极值,极值是稳定的还是周期变化的。
- 统计数据的极大值和极小值是否有限制。
- 统计数据是否有拐点。
- 统计数据是否有对称性。

根据以上四个准则,选用一批相似的模型放入模型库,从该模型库中选择拟合误差最小的模型。

2. 统计模型建立过程

(1)系统的设计思想。建模对象的统计数据放入数

据库文件,从该指定数据库文件中抽取数据与模型库中一个模型曲线计算其误差,对模型库中 k 个模型逐个计算误差,对结果误差进行排序求出误差最小的模型,其程序流程如图所示。



(2) 程序的数据结构

```

struct mul-dis{
    int i;
    double para [n];
    double err,
        {mu-disp[k]
  
```

其中, i 模型库中的模型号
 $para$ 模型库中的参数
 n 模型库中最多的模型参数个数
 err 计算误差
 $mu-disp$ 存放模型的模型库共有 k 个模型。

(3) 用最小二乘法计算误差。实际例子(略)

三、统计模型建模的实现

将要建模的统计数据保存在数据库文件中,用户运行建模系统的程序时,只须输入数据文件名,该系统的程序从数据库文件中抽取数据,根据数据点自动生成坐标,用数据文件中数据和模型库中曲线方程求出误差最小的一种模型,画出其曲线并自动给出方程式及其拟合误差。

程序中建立的主要函数:

```

* coordination(n, x, y)
    int n;
  
```

```

double [], y[];
* draw(n, x, y)
    int n;
    double x[], y[];
  
```

由 x, y 数组的值,画出实际的数据点

```

* win1()
  
```

将方程形式和参数装入 $buf1[]$ 中。

```

* win2()
  
```

将数据点个数和数据点装入 $buf2[]$ 中

```

* win3()
  
```

将文件中的输入数据存于 $mu-disp[]$ 中,并根据方程及参数计算误差。

```

* win4()
  
```

循环调用 $win1(), win2(), win3()$ 逐个对模型库中各组模型求其拟合误差。

```

* disew(equation, err, k)
  
```

```

    char * equation;
  
```

```

    char * err;
  
```

```

    int k;
  
```

对 k 个方程的误差进行排序,选用最小误差。

```

* diserr(equation, err)
  
```

```

    char * equation;
  
```

```

    char * err;
  
```

显示误差最小的曲线方程及误差。

```

* opnization(numdata, x, y)
  
```

```

    int numdata;
  
```

```

    double x[], y[];
  
```

根据 x, y 数组的数据,用各种曲线方程去逼近,找出误差最小的一种,画出曲线,显示其数学方程和误差。

四、结束语

本文简要地介绍了统计模型的设计思想和实现方法,它适用于专门领域中的统计数据快速建立相对应的统计模型。本系统建立在统计数据采样正确并且存在一批与它相匹配的统计模型的基础上。如果不满足上述两个前提,则会导致建模失败。

(来稿时间:1997年1月)