

统计推断在数模转换系统中的应用

组号： 09 柏硕 5130309690, 盛爱国 5130309688

摘要：本报告以上海交通大学工程实践与科技创新 3A 的实验结果与研究数据，探究统计推断在数模转换系统中的应用。因为在工程实际中，若对每个样品都进行 51 组数值的完整测定，则既费时又高成本，若能减少需要观测的数值组数量，则可以提高工效。故尝试找出一个由少量点构成的解空间，通过该解空间元素的拟合曲线可以比较精确地描述整体的函数。报告以遗传算法为基本思路，用特征点选取确定初始解空间，采用三次样条插入法来求得满足评价函数的极优解。

关键词：统计推断，电压，占空比，曲线拟合，遗传算法

The Article For The Course of Statistic Application On The AD-DA Converting system

Abstract: This report by the Shanghai Jiao Tong University engineering practice and innovation of science and technology of 3A test results as the research data, inquiry statistical inference in digital to analog conversion applications. Because in the engineering practice, for each sample were 51 sets of numerical integrity determination, is both time consuming and high cost, if you can reduce the number of numerical groups need to be observed, it can improve the work efficiency. So try to find a small point of the solution space, by fitting the curve of the solution space elements can describe the overall function accurately. The report uses genetic algorithm as the basic train of thought, Determination of the initial solution space with feature point selection and three spline interpolation technique to obtain satisfaction evaluation function better solution.

Key words: statistical inference, voltage, duty cycle, genetic algorithm

1 引言

在电子元器件生产之后，检测其性能参数并在直角坐标系中将其拟合成直观的曲线图形式提供给用户，将能使用户更清晰地了解产品的性能。但是，在大批量生产线上，由于成本和时间的限制，最后的每一个器件的检测对于某个参数不可能取很多个不同值来绘制精确的曲线。一般的解决方案是通过在产品的参数曲线上找出少量的具有代表性的点进行测量，将这几个少量的测量值进行曲线拟合得出近似的产品参数曲线。如果取得点较多，曲线拟合得就越精确，但是成本就会上升。如果取得点较少，虽然成本得到了节约，但是曲线拟合的精确性就会下降，不便于用户使用。为了既兼顾成本和精确度，就必须限定取的点的数量并且在曲线最具有代表性的地方取这些有限的点。

2 数学模型和解决方案

2.1 初步方案

2.1.1 数学模型

这个问题抽象为一个数学问题就是在一个器件 D-U 曲线未知的前提下,求取 7 个电压值进行测量,得出 7 个 (U_n, D_n) 的点,通过这 7 个点拟合出 D-U 曲线。这 7 个电压值是根据抽样调查器件得到若干样本进行统计推断后来求取的。而不是事先已知一组 (U_n, D_n) 点,在这些点中寻找 7 个点来拟合曲线。

2.1.2 方案概述

第一步是 7 个特征点的选取。借助科创数据查看工具^[1],首先剔除一些受到怀疑的样本,得到含有 141 组的新的样本空间。然后得出在每个确定电压值 U_n 下新的样本空间中 PWM 信号占空比样本的分布。并且通过 PWM 信号占空比样本的分布,推断出该器件的 PWM 信号在各个电压值 U_n 下的概率分布情况。根据 PWM 信号的分布情况,得出每个确定电压值 U_n 下 PWM 信号的期望。通过每个确定电压值 U_n 下 PWM 信号的期望,绘出 PWM 信号的期望 $E(PWM)-U$ 的曲线关系。再对 PWM 信号的期望 $E(D)-U$ 的曲线关系进行分析,找出在该曲线上误差最小的 7 个电压值的大小 $\{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$ 。

第二步是函数关系的选取。主要可以采取的方法有如下两种:

一、插值

在特征点位置上,表达式与实验值无误差

二、拟合

在特征点位置上,表达式与实验值可以有误差(残差)

通过讲座我们了解到由于线性插值误差较大,而三次样条插值计算量较大,基于准确性和成本的考虑,应采用多项式拟合,由于不可能为线性,故从二次多项式开始考虑当阶数增加时,标准差降低,所得精度越来越高,但计算量也随之加大。除二阶拟合外,三阶四阶以至于更高阶相差不大。为了得到精度与成本的平衡,我们暂定选择三阶多项式拟合。

2.1.3 研究过程

我们暂且应用三次函数进行拟合,我们先用 51 个点进行拟合得出三次曲线函数拟合得到曲线 $y = -0.000005696558241x^3 + 0.000662326139424x^2 + 0.042418899879240x + 4.800113072120975$

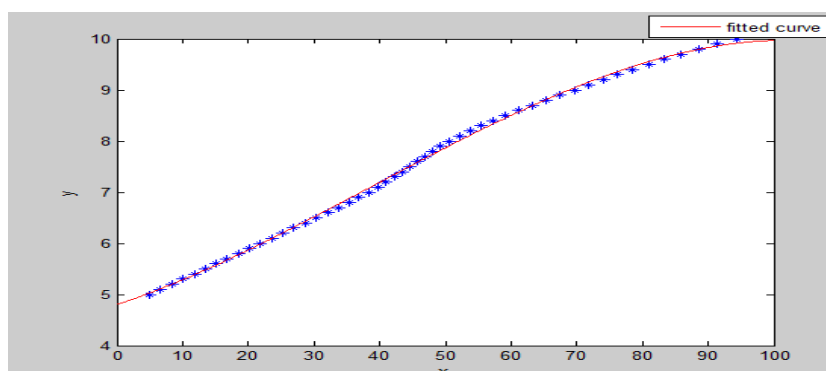


图 2-1-1 51 个点拟合的三次曲线

这个函数的代价函数是 $c=2.824237276549117e+02$

将 51 个点的 $E(D)$ 代入由 51 个点拟合出的三次多项式,然后将得出的电压值与理论值比较,找到 51 个点中的误差最小的 7 个点

$x = [25.2168 \ 8.3960 \ 44.5518 \ 88.5236 \ 65.3126 \ 23.5290 \ 63.2355];$

$y = [6.2 \ 5.2 \ 7.5 \ 9.8 \ 8.8 \ 6.1 \ 8.7];$

再用这 7 个点拟合三次曲线函数为

$y = -0.000005550996753x^3 + 0.000642752854526x^2 + 0.043121908025350x + 4.796588864565979$

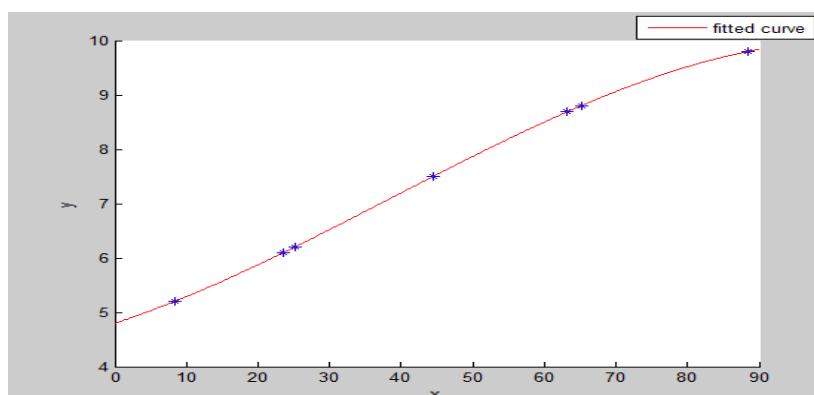


图 2-1-2 初步方案的三次曲线

这个函数的代价函数是 $c=2.821725532441016e+02$

我们又利用特征点选取中的方案二选得的特征点进行了四次拟合，得到 $c=2.809940480326293e+02$ ，四次略低于三次。

七个特征点

$x = [65.3126 \ 26.9295 \ 45.6698 \ 88.5236 \ 85.9337 \ 10.0845 \ 63.2355];$

$y = [8.8 \ 6.3 \ 7.6 \ 9.8 \ 9.7 \ 5.3 \ 8.7];$

2.2 改进方案

在和袁焱老师面谈后，了解到我们组的初步方案存在很多不足之处，首先代价值太高，促使我们不得不改变方案，另外利用期望值来进行一系列的操作就已经是不合理了。因此，我们这里采用了更科学的遗传算法，拟合过程使用三次样条插值法。

2.2.1 遗传算法^[2]

遗传算法（Genetic Algorithm）是一类借鉴生物界的进化规律（适者生存，优胜劣汰遗传机制）演化而来的随机化搜索方法。其主要特点是直接对结构对象进行操作，不存在求导和函数连续性的限定；具有内在的隐并行性和更好的全局寻优能力；采用概率化的寻优方法，能自动获取和指导优化的搜索空间，自适应地调整搜索方向，不需要确定的规则。遗传算法的这些性质，已被人们广泛地应用于组合优化、机器学习、信号处理、自适应控制和人工生命等领域。

遗传算法是解决搜索问题的一种通用算法，对于各种通用问题都可以使用。搜索算法的共同特征为：

- （1）首先组成一组候选解；
- （2）依据某些适应性条件测算这些候选解的适应度；
- （3）根据适应度保留某些候选解，放弃其他候选解；
- （4）对保留的候选解进行某些操作，生成新的候选解。

在遗传算法中，上述几个特征以一种特殊的方式组合在一起，：基于染色体群的并行搜索，带有猜测性质的选择操作、交换操作和突变操作。这种特殊的组合方式将遗传算法与其它搜索算法区别开来。

2.2.2 分段样条插值法

- （1）使用三次样条插值对确定一个可能解，即 7 个特征点组合 $S=\{s_1, s_2, \dots, s_7\}$ 进行拟合。

具体方法为：首先对于非两端点，以四个连续点确定一条三次曲线，但仅在中间两点之间用该三次曲线表示，以此类推，所有非两端点之间均有三次曲线。两端点由端点处三个点用二次曲线拟合。

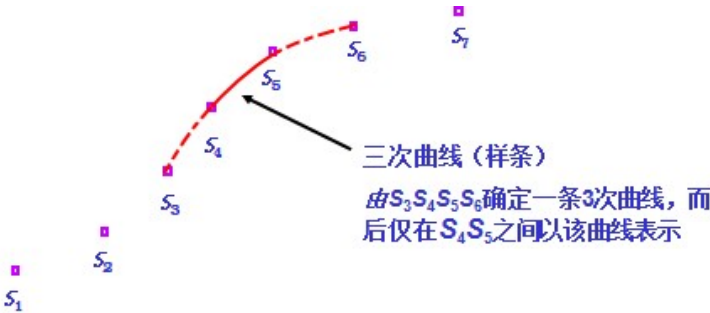


图 2-2-1 三次样条插值曲线拟合

(2) 评估该曲线的拟合度。将去掉 1 和 51 号点的其余 49 个点的占空比 D 代入得到三次样条插值后相应的 49 个电压值，通过评价函数得到评价分值。

(3) 现最优化过程，即不断的寻找存在的可能解，继续前两步骤。每次得到可能解，通过评价，做出一定概率的接受或者舍弃，实现当前解的优化，直至达到终止条件。

2.2.3 解决方法在 MATLAB 中的实现

(1) 确定初始条件，种群大小，繁殖次数，交叉互换概率，突变概率，并以七个点的标号作为每一个种群的个体，随机生成一百个个体作为初始种群。

(2) 用代价函数提供适应度，并使用三次样条差值法拟合曲线，从而计算每个个体的适应度，称之为 Q。

(3) 根据轮盘赌法，使 Q 值小的拥有更多的繁殖机会，通过交叉互换以及突变产生一个新种群。

(4) 记录每代中 Q 最小的个体，并继续进行下一代的繁殖，如此继续，直到达到设定的最大代数，此时记录中的最小个体即为我们得到的七个点选取的较优解。

2.2.4 程序运行结果

表 2-2-1 程序运行结果表

编号	运行结果	代价值
1	1 9 18 26 35 44 51	95.808102
2	1 10 20 28 35 44 51	95.487207
3	1 5 20 26 34 44 51	97.455224
4	1 10 21 28 37 46 51	97.037313
5	1 11 20 26 34 45 51	96.640725
6	1 9 20 26 33 44 51	95.311301
7	1 6 20 29 36 45 51	97.970149
8	1 7 18 24 33 44 51	97.140725
9	1 10 21 29 35 45 51	96.459488
10	1 11 21 29 36 46 51	97.325160

3 结果分析和结论

3.1 结果分析

2.1 的初步方案中四次略低于三次，但双方代价值都非常高，显然不可取。而 2.2 的改

进方案得出的代价值远远小于初步方案。另外有好几组评价相当接近，也有所波动，我们认为这是由于在所给数据中有部分数据本身的拟合性不好，导致代价有所上升，在实际解决问题时应该考虑这个因素的影响，尽可能地将实验错误的的数据排除，以减少实验误差。

3.2 结论

通过实验分析，我们给出的最佳的七个特征点为：1 9 20 26 33 44 51，它的代价值为 95.311301。

4 参考资料

- [1] 袁焱老师. 统计推断课程讲座讲义. 上海: 上海交通大学 电子工程系
- [2] 百度百科. 网址: <http://baike.baidu.com/view/45853.htm> 百度公司。

5 附录 (MATLAB 运行代码)

5.1 初步方案代码

一. 三次拟合

```
format long
a=xlsread('20141010dataform_copy.csv');
b = mean (a)';c = var(a)';d = [5:0.1:10]';
b = [b,c,d];

x = b (:,1)';y = b(:,3)'
ft = fitttype('poly3');
yourLine = fit(x',y',ft);
hold on
coeffvalues(yourLine)
d = [5:0.1:10];
c = 0;
for n = 1:141
    c = c +sum (( yourLine (a(n,:))'-d).^2);
end
c

%拟合得到曲线 y = -0.000005696558241x^3 + 0.000662326139424x^2 +
0.042418899879240x^1 + 4.800113072120975
%用 51 个点拟合三次曲线可得到 c1 值为 2.824237276549117e+02

A=(yourLine (b (:,1))'-d).^2;
[B,id]=sort(A)
%得到 13 3 26 49 39 12 38 组为方差最小。
%找到对应的电压值和均值,
% 6.2 5.2 7.5 9.8 8.8 6.1 8.7
% 25.2168 8.3960 44.5518 88.5236 65.3126 23.5290 63.2355
x = [25.2168 8.3960 44.5518 88.5236 65.3126 23.5290 63.2355];
```

```

y = [6.2 5.2 7.5 9.8 8.8 6.1 8.7];
ft = fitttype('poly3');
yourLine = fit(x',y',ft);
%对比下拟合曲线和原来的点
plot(x,y,'*');
hold on
plot(yourLine)
coeffvalues(yourLine)
d = [5:0.1:10];
c = 0;
for n = 1:141
c = c +sum (( yourLine (a(n,:))'-d).^2);
end
c
%拟合得到曲线 y = -0.000005550996753x^3 + 0.000642752854526x^2 +
0.043121908025350x^1 + 4.796588864565979
%用这 7 个点拟合三次曲线可得到 c2 值 2.821725532441016e+02

```

二. 二次拟合

```

x = b(:,1)';y = b(:,3)'
ft = fitttype('poly4');
yourLine = fit(x',y',ft);
hold on
plot(yourLine)
coeffvalues(yourLine)
d = [5:0.1:10];
c = 0;
for n = 1:141
c = c +sum (( yourLine (a(n,:))'-d).^2);
end
c
%拟合得到曲线 y = 0.000000058258685x^4 + (-0.000017160288059)x^3 +
0.001407302647484x^2 + 0.024553959799427x^1 + 4.917853349522235
%用 51 个点拟合四次曲线可得到 c1 值为 2.810157113742719e+02
A=(yourLine (b(:,1))'-d).^2;
[B,id]=sort(A)
%得到 39 14 27 49 48 4 38 组为方差最小。
%找到对应的电压值和均值,
% 8.8 6.3 7.6 9.8 9.7 5.3 8.7
% 65.3126 26.9295 45.6698 88.5236 85.9337 10.0845 63.2355
x = [65.3126 26.9295 45.6698 88.5236 85.9337 10.0845 63.2355];
y = [8.8 6.3 7.6 9.8 9.7 5.3 8.7];
ft = fitttype('poly4');
yourLine = fit(x',y',ft);
%对比下拟合曲线和原来的点

```

```

plot(x,y, '*');
hold on
plot(yourLine)
coeffvalues(yourLine)
d = [5:0.1:10];
c = 0;
for n = 1:141
    c = c +sum (( yourLine (a(n,:))'-d).^2);
end
c
%拟合得到曲线 y = 0.000000076195043x^4 + (-0.000020608171535)x^3 +
0.001625265069711x^2 + 0.019389647463606x^1 + 4.959894264808884
%用这 7 个点拟合四次曲线可得到 c2 值 2.809940480326293e+02

```

5.2 改进方案代码

```

function out = select(gene,cost,pop)
% 自然选择,out(1)为父代最优予以保留
out=zeros(pop,51);
cost0=max(cost)-cost;
s0=sum(cost0);
s=zeros(pop+1);
s(1)=0;
s(pop+1)=1;
s(2:pop)=sum(cost0(1:pop-1))/s0;
for i=2:pop
    t=rand();
    j=search(t,s,1,pop+1);
    out(i,:)=gene(j,:);
end
sort0=[[1:pop]',cost];
sort0=sortrows(sort0,2);
out(1,:)=gene(sort0(1,1),:);
end

```

```

function [out] = search(in,s,l,r)
% 二分法查找
mid=floor((l+r)/2);
if in<=s(mid)
    if in>s(mid-1)
        out=mid-1;
    else
        out=search(in,s,l,mid);
    end
else

```

```

        if in<=s(mid+1)
            out=mid;
        else
            out=search(in,s,mid,r);
        end
    end
end
end

function out = mutate(gene,pop,pm)
% 变异
out=gene;
for i=2:pop;
    for j=2:50;
        t=rand();
        if t<=pm
            out(i,j)=~out(i,j);
        end
    end
end
end
end

function out = generate(gene,pop,pc)
% 交叉,保留 gene(1)
% i 与 pop-i+2 配对
for i=2:floor(pop/2+1)
    out=gene;
    mid=floor(rand()*50)+1;
    t=rand();
    if t<=pc
        out(i,1:mid)=gene(pop-i+2,1:mid);
        out(pop-i+2,1:mid)=gene(i,1:mid);
        out(i,mid+1:51)=gene(pop-i+2,mid+1:51);
        out(pop-i+2,mid+1:51)=gene(i,mid+1:51);
    end
end
end
end

function out = geneinit(pop)
% 随机产生初始种群
out=round(rand(pop,51)-0.2);
out(:,1)=1;
out(:,51)=1;
end

```



```

function [out] = errorcost(dy)
% 单个个体平均误差成本函数
t=abs(dy);
t0=sum(sum(t<=0.5));
t1=sum(sum(t<=1))-t0;
t2=sum(sum(t<=2))-t0-t1;
t3=sum(sum(t<=3))-t0-t1-t2;
t4=sum(sum(t<=5))-t0-t1-t2-t3;
t5=sum(sum(t>5));
out=0.5*t1+1.5*t2+6*t3+12*t4+25*t5;
end

function [out] = assess(in,y)
% 计算成本
out=length(in)*12;
x=5:0.1:10;
xx=5+(in-1)*0.1;
yy=y(:,in); % 测试点 y 值矩阵
f=spline(xx,yy);
dy=ppval(f,x)-y; % 理论值与实际值的差
out=out+errorcost(dy)/469;
fid=fopen('answer.txt','a');
fprintf(fid,'position: [ ');
fprintf(fid,'%2d ',in);
fprintf(fid,']    mean_cost: %7f\n\n',out);
fclose(fid);
end

```

```

function out = adapt(gene,y,pop)
% 计算每个个体平均成本
out=zeros(pop,1);
x=5:0.1:10;
for i=1:pop
    c=sum(gene(i,:)==1); % 测试点数量
    pos=find(gene(i,:)==1); % 测试点位置
    xx=5+(pos-1)*0.1; % 测试点 x 值
    yy=y(:,pos); % 测试点 y 值矩阵
    f=spline(xx,yy);
    dy=ppval(f,x)-y; % 理论值与实际值的差
    out(i)=12*c+errorcost(dy)/469; % 单个个体平均成本
end
min(out)
mean(out)
end

```

```

% main()主函数
data=csvread('20141010dataform.csv');
pop=100; % 种群数量
pc=0.9; % 交叉概率
pm=0.01; % 突变概率
n=100; % 进化代数
y=zeros(469,51);
y(1:469,:)=data(2:2:938,:);
gene=geneinit(pop);
for g=1:n
    display(g);
    cost=adapt(gene,y,pop);
    gene=select(gene,cost,pop);
    gene=generate(gene,pop,pc);
    gene=mutate(gene,pop,pm);
    display(find(gene(1,:)==1));

display('-----')
end
xx=find(gene(1,:)==1);
assess(xx,y); % 计算测试点组合的平均成本

```