

文章编号:1671-8879(2007)01-0027-04

公路边坡病害的土工试验数据拟合及编程

南春丽^{1,2}, 申 康¹, 窦明健¹

(1. 长安大学 特殊地区公路工程教育部重点实验室, 陕西 西安 710064;

2. 长安大学 信息工程学院, 陕西 西安 710064)

摘 要:针对目前公路工程土工试验数据处理方法的不足,利用数据拟合原理和计算机可视化编程技术,对公路边坡病害研究中的土工试验数据进行分析,给出了多项式线性最小二乘拟合和三次 B 样条函数拟合两种方法及计算机实现算法。以含水量分布规律为例,对两种方法进行验证。结果表明,两种方法的拟合曲线均较好地逼近原始试验数据折线,三次 B 样条函数的偏差更小一些,平均绝对值误差比多项式线性最小二乘平均绝对值误差小 0.053 9%。

关键词:道路工程;公路边坡病害;土工试验;线性最小二乘拟合;三次 B 样条函数拟合;计算机实现算法

中图分类号:U416.14

文献标志码:A

Approximation and program of soil test data in highway slope diseases

NAN Chun-li^{1,2}, SHEN Kang¹, DOU Ming-jian¹

(1. Key Laboratory for Special Area Highway Engineering of Ministry of Education, Chang'an University, Xi'an 710064, Shaanxi, China; 2. School of Information Engineering, Chang'an University, Xi'an 710064, Shaanxi, China)

Abstract: In order to optimize the methods of current test data processing, this paper analyzed the soil test data by approximation and visual program in the study of the highway slope diseases. It presents the linear least square approximation, cubic B spline approximations and programming algorithms. In order to verify the methods above, the water percentage distribution regularity is analyzed. The results indicate that this method is good and the cubic B spline approximation is better. The mean absolute error of cubic B spline approximation is 0.053 9% smaller than of the linear least square approximation. 2 tabs, 2 figs, 11 refs.

Key words: road engineering; highway slope diseases; soil test; linear least square approximation; cubic B spline approximation; programming algorithms

0 引 言

新疆天山林木区公路边坡病害研究技术路线如图 1 所示^[1]。从图 1 可以看出,土工试验数据处理结果直接影响溜滑模拟试验和真实状态的逼近程

度。目前,工程技术人员常采用 Excel 应用软件的图表功能对土工试验数据进行回归分析,其主要缺点有:①处理比较简单,主观误差较大^[2-3],不能直接得到某些特殊指标;②数据组织缺少数据有效性规则检验,共享程度低,不利于程序自动实现。作为实

收稿日期:2005-10-20

基金项目:国家西部交通建设科技项目(200131800055)

作者简介:南春丽(1971-),女,陕西渭南人,讲师,博士研究生,E-mail:chlnan@chd.edu.cn。

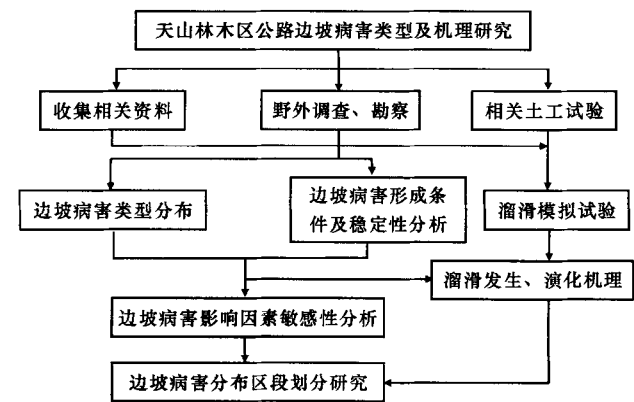


图1 新疆天山林木区公路边坡病害研究技术路线

例,土工试验中的含水量试验需要解决的问题主要有:①找出反映某断面含水量分布规律的数学模型;②根据数学模型求出任意点的含水量。为此,本文从数据拟合理论和有利于计算机技术实现的角度出发,给出了两种数据拟合算法及程序实现思路:多项式线性最小二乘拟合算法和 B 样条函数拟合算法。

1 多项式线性最小二乘拟合算法

1.1 线性最小二乘逼近原理

数值逼近是对给出的数据进行拟合及对复杂的函数进行逼近。数据拟合是数值逼近的一个重要分支,在试验过程中试验数据一般不可能精确获得,往往有一定程度的误差。为了尽可能减少误差,希望用一个拟合函数 $f(x)$,使拟合函数 $f(x)$ 在 x_i 处的函数值 $f(x_i)$ 与对应的试验实测数据值 $y_i(i=1,2,\dots,N)$ 形成误差平方和最小,即

$$\min(\sum_{i=1}^N (f(x_i) - y_i)^2)$$

这就是最小二乘数据拟合问题。最小二乘数据拟合问题最重要的是确定拟合函数的类型。取广义多项式为拟合函数的最小二乘拟合问题,称为线形最小二乘问题。

1.2 确定拟合函数

2004 年 2 月,在天山林木区伊昭公路沿山坡采取的草皮层土样试验数据见表 1。

表 1 草皮腐殖层含水量

测点距边坡边缘的距离/m	0	8	16	24	32	40
含水量/%	35.7	28.9	23.8	24.4	25.7	30.4

表 1 数据取自一个断面,各测点自边坡边缘沿地面斜坡以间距 8 m 向上延伸,取土深度约 20 cm。为了找出该断面含水量分布规律,记该组试验数据集为 $\{(x_i, y_i)\}_{i=1}^N$, x_i, y_i 分别表示一个测点距边坡边缘的距离和相应的含水量, y_i 是 x_i 的函数, x_i, y_i

之间的函数关系就是该断面含水量分布规律。利用程序生成试验数据折线图,发现数据点大体呈抛物线分布,于是确定拟合函数为二次多项式,即

$$f(x) = a + bx + cx^2 \tag{1}$$

式中: a, b, c 为二次多项式的系数。

根据最小二乘法原理, $f(x)$ 应满足以下条件,即

$$s = \sum_{i=1}^N (f(x_i) - y_i)^2 \tag{2}$$

取最小值。由于 s 可以看作是关于系数 a, b, c 的多元函数,故上述拟合函数的构造问题可归结为多元函数的极值问题。由多元函数极值的必要条件

$$\frac{\partial s}{\partial a} = 0, \frac{\partial s}{\partial b} = 0, \frac{\partial s}{\partial c} = 0$$

可得三元线性方程组

$$\begin{cases} aN + b \sum_{i=1}^N x_i + c \sum_{i=1}^N x_i^2 = \sum_{i=1}^N y_i \\ a \sum_{i=1}^N x_i + b \sum_{i=1}^N x_i^2 + c \sum_{i=1}^N x_i^3 = \sum_{i=1}^N x_i y_i \\ a \sum_{i=1}^N x_i^2 + b \sum_{i=1}^N x_i^3 + c \sum_{i=1}^N x_i^4 = \sum_{i=1}^N x_i^2 y_i \end{cases} \tag{3}$$

为了便于计算机编程,将式(3)线性方程组写成如下紧凑格式

$$AX = Y \tag{4}$$

其中 A, X 和 Y 分别定义为

$$A = \begin{bmatrix} N & \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 \\ \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 & \sum_{i=1}^N x_i^3 \\ \sum_{i=1}^N x_i^2 & \sum_{i=1}^N x_i^3 & \sum_{i=1}^N x_i^4 \end{bmatrix}$$
$$X = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, Y = \begin{bmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_i y_i \\ \sum_{i=1}^N x_i^2 y_i \end{bmatrix}$$

将表 1 中数据代入式(4),得到矩阵方程

$$\begin{bmatrix} 6 & 120 & 3\,520 \\ 120 & 3\,520 & 115\,200 \\ 3\,520 & 115\,200 & 4\,009\,984 \end{bmatrix} X = \begin{bmatrix} 168.9 \\ 3\,236 \\ 96\,954 \end{bmatrix} \tag{5}$$

可解得 $X = [35.632\,7, -1.054\,3, 0.023\,2]^T$, 即 $a = 35.632\,7, b = -1.054\,3, c = 0.023\,2$, 代入式(1),得拟合函数 $f(x)$ 的具体表达式为

$$f(x) = 35.632\,7 - 1.054\,3x + 0.023\,2x^2 \tag{6}$$

式(6)即为反映该断面含水量总体分布规律的数学模型。只要将任意点的 x 代入式(6),所得函数值即为该点处的含水量。对式(6)求导数,令其一阶导数等于 0,可得该断面最小含水量为 22.722%,说明该断面总体含水量偏大。

1.3 误差计算

根据线性最小二乘多项式误差式(2),计算出的 s 的值即为线性最小二乘误差。经计算 $s=0.1840$,误差较小,说明构造的拟合函数是合理的。

1.4 算法程序实现

Matlab 具有强大的数值分析功能^[4-8],但是在数据组织和数学模型处理的灵活性方面有一定的局限性;通用语言在处理数值分析时,一切必须从最底层做起,工作量特别大,但是它有极大的灵活性。所以,本文在选择开发平台时,充分考虑到各自的利弊,将两者的优势结合起来,基础平台选为通用语言 VC++,数值分析实验平台选为 Matlab,通过引擎实现 VC++ 和 Matlab 的混合编程。

2 B 样条函数拟合

在有些情况下,根据给定的数据,无法确定拟合函数的类型,这时用 B 样条函数作为拟合函数是解决这类问题的好办法。一方面 B 样条函数相对光滑,对原始数据有良好的近似;另一方面 B 样条函数的算法便于利用计算机来实现。工程实践中,应用最广泛的是二次或三次 B 样条,而其他的 B 样条几乎不用^[9-11]。

2.1 B 样条函数拟合原理

样条函数的实质是分段多项式光滑对接,如果以 B 样条函数为基底,且每段为三次 B 样条,即整个拟合函数是三次 B 样条基底的线性组合,则该样条函数称为三次 B 样条。每段 B 样条由 4 个控制点来决定(因为三次 B 样条曲线的特征多边形为四边形,顶点数为 4),假如有 N 个试验数据 $\{p_i\}_{i=1}^N$, $p_i=(x_i, y_i)$,取相邻的 4 个点为一组,总共分为 $N-3$ 组,分组情况如下:

第一组: p_1, p_2, p_3, p_4 ;

第二组: p_2, p_3, p_4, p_5 ;

第三组: p_3, p_4, p_5, p_6 ;

第 N 组: $p_{N-3}, p_{N-2}, p_{N-1}, p_N$ 。

现在的问题是三次 B 样条基底的构造。

2.2 B 样条函数的数学表达式

对于 n 次的 B 样条函数,有如下表达式

$$S_n(t) = \sum_{i=0}^n F_{i,n}(t) P_i \quad (0 \leq t \leq 1) \quad (7)$$

式中: P_i 为 n 次 B 样条函数曲线的特征多边形顶点; $F_{i,n}(t)$ 为 n 次 B 样条函数的基底函数,且

$$F_{i,n}(t) = \frac{1}{n!} \sum_{j=0}^n (-1)^j C_{n+1}^j (t+n-j)^n$$

其中 $C_{n+1}^j = \frac{(n+1)!}{j!(n+1-j)!}$

由式(7)可导出,当 $n=3$ 时,对应的基底函数分别为

$$F_{0,3}(t) = \frac{1}{3!} \sum_{j=0}^3 (-1)^j C_4^j (t+3-j)^3 = \frac{1}{6} (-t^3 + 3t^2 - 3t + 1)$$

$$F_{1,3}(t) = \frac{1}{3!} \sum_{j=0}^2 (-1)^j C_4^j (t+2-j)^3 = \frac{1}{6} (3t^3 - 6t^2 + 4)$$

$$F_{2,3}(t) = \frac{1}{3!} \sum_{j=0}^1 (-1)^j C_4^j (t+1-j)^3 = \frac{1}{6} (-3t^3 + 3t^2 + 3t + 1)$$

$$F_{3,3}(t) = \frac{1}{3!} \sum_{j=0}^0 (-1)^j C_4^j (t-j)^3 = \frac{1}{6} t^3$$

这就是三次 B 样条基底函数。

于是式(7)可写成参数表达式

$$\begin{cases} x(t) = F_{0,3}(t)x_0 + F_{1,3}(t)x_1 + \\ \quad F_{2,3}(t)x_2 + F_{3,3}(t)x_3 \\ y(t) = F_{0,3}(t)y_0 + F_{1,3}(t)y_1 + \\ \quad F_{2,3}(t)y_2 + F_{3,3}(t)y_3 \end{cases} \quad (8)$$

利用式(8),可对表 1 中数据进行三次 B 样条拟合,但曲线没有通过起点和终点。欲使拟合曲线满足特殊要求,需加边界条件,具体实现方法见文献^[6]⁵⁴。

2.3 任意点含水量的三次 B 样条函数值计算

由于三次 B 样条函数是分段函数,不同的自变量区间函数的具体表达式不同,所以要计算出任意点 x 对应的函数值 y (即含水量)必须先判断该点所在的区间。区间的划分方法与前述的分组方法相同,一个区间就是一个分组,每组的 4 个点控制一段三次 B 样条函数曲线。也就是说整个三次 B 样条函数曲线是由 n (n 为前述中的组数)段三次 B 样条函数曲线组成,每段三次 B 样条函数曲线又是由若干连续直线段逼近生成。步长越小,对原始数据近似性越好。所以判断出任意点 x 所在的区间后,给定合适的步长,根据步长计算出每个小直线段的端点坐标,于是肯定存在某直线段两端点 (x_i, y_i) , (x_{i+1}, y_{i+1}) ,使 $x_i \leq x \leq x_{i+1}$ 。以该两端点为结点构

造插值函数

$$y = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}(x - x_i) + y_i \tag{9}$$

将任意点的 x 代入式(9), 所得函数值 y 即为所求含水量. $x_i, x_{i+1}, y_i, y_{i+1}$ 根据式(8)计算. 实际中, 步长精度只要满足工程要求即可^{[7]35}, 本文取步长 $\Delta t=0.1$, t 的含义同前.

根据上述思路, 可实现可视化编程.

3 拟合方法对比

两种拟合算法程序运行结果如图 2 所示. 偏差分析见表 2.

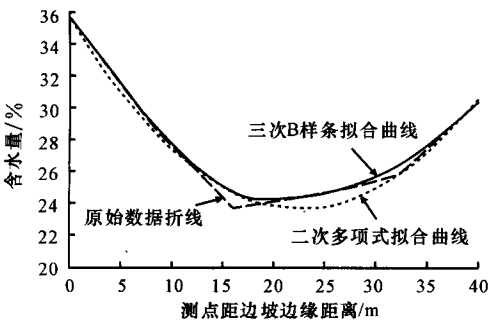


图 2 三次 B 样条拟合与二次多项式拟合曲线对比

表 2 偏差分析 %

原始试验数据	35.7	28.9	23.8	24.4	25.7	30.4
二次多项式拟合函数值	35.632 7	28.683 1	24.703 1	23.692 7	25.651 9	30.580 7
偏差	0.067 3	0.216 9	-0.903 1	0.707 3	0.048 1	-0.180 7
平均绝对误差	0.353 9					
三次 B 样条拟合函数值	35.700 0	28.583 3	24.600 0	24.516 7	26.266 7	30.400 0
偏差	0	0.316 7	-0.800 0	-0.116 7	-0.566 7	0
平均绝对误差	0.300 0					

从图 2 和表 2 可以看出, 三次 B 样条函数和二次多项式拟合对原始试验数据均有较好的逼近, 前者的近似性更好一些, 平均绝对误差要小 0.053 9%.

4 结 语

(1) 试验数据分布规律比较明显时, 采用多项式线性最小二乘法拟合; 根据折线图分布规律找不到合适拟合函数时, 可采用三次 B 样条函数拟合.

(2) 通用可视化语言与数值分析软件 Matlab 相结合进行编程, 既有利于数据合理组织, 提高灵活性, 又可以减少复杂运算带来的工作量, 处理结果以直观化的图形方式显示.

(3) 对新疆伊昭公路沿山坡所取草皮层土样含水量进行分析, 结果与实际相符, 证明拟合方法是可

行的.

参考文献:

References:

[1] 申 康. 天山林木区公路边坡病害类型及机理研究[D]. 西安: 长安大学, 2005.

[2] 冯忠居, 谢永利. 标准击实试验最佳含水量和最大干密度的理论计算[J]. 长安大学学报: 自然科学版, 2002, 22(2): 10-13.

FENG Zhong-ju, XIE Yong-li. Numerical method to result analysis of indoor standard compaction test[J]. Journal of Chang'an University: Natural Science Edition, 2002, 22(2): 10-13.

[3] 南春丽. 计算机辅助土工试验数据处理与分析[J]. 交通与计算机, 1999, 17(3): 47-49.

NAN Chun-li. Soil test data processing and analyzing with program[J]. Traffic and Computer, 1999, 17(3): 47-49.

[4] Shoichiro N. Numerical analysis and graphic visualization with MATLAB[M]. Beijing: Electronics Industry Press, 2002.

[5] 赵祥模, 南春丽, 马 建, 等. 汽车制动性能检测中制动力数据拟合与优化方法研究[J]. 中国公路学报, 2003, 16(3): 100-104.

ZHAO Xiang-mo, NAN Chun-li, MA Jian, et al. Study of braking data quasi-harmony and optimization method in auto brake function test[J]. China Journal of Highway and Transport, 2003, 16(3): 100-104.

[6] 崔洪斌. 计算机辅助设计基础及应用[M]. 北京: 清华大学出版社, 2004.

[7] JTJ 051-93, 公路土工试验规程[S].

[8] 杨高波, 元 波. 精通 MATLAB 7.0 混合编程[M]. 北京: 电子工业出版社, 2006.

[9] 赵祥模, 南春丽, 施维颖. 交通工程 CAE 软件系统的设计与实现方法[J]. 交通运输工程学报, 2003, 3(3): 96-100.

ZHAO Xiang-mo, NAN Chun-li, SHI Wei-ying. Design and realizing method of traffic engineering CAE system[J]. Journal of Traffic and Transportation Engineering, 2003, 3(3): 96-100.

[10] 程正兴, 李水根. 数值逼近与常微分方程数值分析[M]. 西安: 西安交通大学出版社, 2000.

[11] 吴刘平, 侯艳斌. 高填路堤下沉原因分析及防治措施[J]. 筑路机械与施工机械化, 2006, 23(4): 30-31.

WU Liu-ping, HOU Yan-bin. Analysis and pretratment of high-filled embankent sinking[J]. Road Machinery & Construction Mechanization, 2006, 23(4): 30-31.