

试验设计方差分析的算法整合与网络实现

张文军 齐艳红

(中山大学昆虫学研究所与生命科学院 广东 广州 510275)

摘 要 试验设计方差分析是统计学的重要分支,其中,单变量的方差分析应用最为广泛。本研究整合了 8 种单变量的试验设计方差分析,其中包括单因子设计 4 种:完全随机设计,随机区组设计,有亚样品的随机区组设计,巢式设计;两因子设计 2 种:双向分组设计,两因子试验 RCBD;多因子设计 2 种:裂区设计,正交设计。算法用 Java 网络化实现,图形用户界面友好,简单易用,驻留于远程服务器,可在兼容 Java 的网络浏览器上调用和运行。不仅可独立使用,也可作为网络软件包及网络数据库的嵌入程序。利用该网络算法,对无脊椎动物多样性数据进行了分析。

关键词 方差分析 算法整合 网络实现

ALGORITHMS INTEGRATION AND NETWORK IMPLEMENTATION OF EIGHT UNIVARIATE VARIANCE ANALYSES

Zhang Wenjun Qi Yanhong

(Research Institute of Entomology, School of Life Sciences, Zhongshan University, Guangzhou Guangdong 510275, China)

Abstract Eight univariate variance analyses were integrated and implemented as Java Applet which can be accessed and computed on Java-enabled web browser. The eight univariate variance analyses included are complete randomized design, randomized block design, randomized block design with subsamples, nested design, two way classification, two factor RCBD, split design, and orthogonal design.

Keywords Variance analysis Algorithms integration Network implementation

0 引 言

试验设计的方差分析理论始自英国数学家 Fisher,现已发展出众多方法。方差分析可用于:

- (1) 根据变异原因区分出各项方差;
- (2) 用合适的方差作为试验误差,以进行显著性检验^[1,2],利用方差分析,可以比较各处理、各因子等是否存在显著差异,发现处理或因子间的交互作用及原因。

试验设计的方差分析在工艺试验,农业试验,医学试验中具有重要的应用价值。

目前,在试验设计的方差分析中应用最多的是单变量的方差分析方法,包括单因子设计、两因子设计及多因子设计三类,每类中包含数种方法。迄今为止,已有不少方差分析的计算机程序。然而,它们或缺乏图形用户界面,或依赖于平台,或不能在网络环境下运行,或需要安装大型的软件包。在网络信息共享的意义上,需要做相应的改进。为了适应信息共享和网络化的迅猛发展,作者整合了 8 种主要的单变量试验设计方差分析,研制了 Java 网络软件,以此提供一种在线计算工具,并作为网络软件包及网络数据库的嵌入程序。

1 方差分析算法

为了正确地理解各种试验设计的方差分析过程,便于选择应用,有必要对本软件采用的算法作一简述。

1.1 单因子设计

1.1.1 完全随机设计

设试验无亚样品,数据矩阵为 (x_{ij}) , $i = 1, 2, \dots, n; j = 1, 2, \dots, m$, 其中, n 为处理数, m 为重复数。处理平方和 (SS_t) 和误差平方和 (SS_e) 分别为:

$$SS_t = \left(\sum_{i=1}^n \left(\sum_{j=1}^m x_{ij} \right)^2 \right) / m - C$$

$$SS_e = \sum_{i=1}^n \sum_{j=1}^m x_{ij}^2 - SS_t - C$$

其中, $C = \left(\sum_{i=1}^n \sum_{j=1}^m x_{ij} \right)^2 / (n * m)$ 。总自由度,处理自由度,以及误差自由度分别为 $df = n * m - 1$, $df_t = n - 1$, $df_e = n * (m - 1)$ 。F 值为: $F = (SS_t / df_t) / (SS_e / df_e)$, 若 $F > F(df_t, df_e)$, 则处理间存在显著性差异。用 LSD 法比较处理间均值差异,即计算 $S = (2 * (SS_e / df_e) / m)^{1/2}$, 则 $LSD_a = t_a(df_e) * S$, 此处 $a = 0.05, 0.01$, $t(df_e)$ 为自由度 df_e 的 t 分布值。计算 $x_{bi} - x_{bj}$, $i, j = 1, 2, \dots, n; i \neq j$, $x_{bi} = \sum_{j=1}^m x_{ij} / m$ 。若 $(x_{bi} - x_{bj}) \geq LSD_a$, 则处理 i 与 j 间差异显著。

1.1.2 随机区组设计

设试验无亚样品,数据矩阵为 (x_{ij}) , $i = 1, 2, \dots, n; j = 1, 2, \dots, m$, 其中, n 为处理数, m 为区组数。

计算总平方和 (SS_T) , 区组平方和 (SS_b) , 处理平方和 (SS_t) 和误差平方和 (SS_e) 。

收稿日期:2004-02-27。国家自然科学基金(30170184)、教育科研基金(2000)。张文军,副教授,主研领域:计算生物学和生物应用数学。

$$\begin{aligned}
 SS_T &= \sum_{i=1}^n \left(\sum_{j=1}^m x_{ij}^2 \right) - C \\
 SS_b &= \sum_{j=1}^m \left(\sum_{i=1}^n x_{ij}^2 \right) / n - C \\
 SS_t &= \sum_{i=1}^n \left(\sum_{j=1}^m x_{ij}^2 \right) / m - C \\
 SS_e &= SS_T - SS_b - SS_t
 \end{aligned}$$

其中, $C = \left(\sum_{i=1}^n \sum_{j=1}^m x_{ij}^2 \right) / (n * m)$ 。区组自由度, 处理自由度, 以及误差自由度分别为 $df_b = m - 1$, $df_t = n - 1$, $df_e = (n - 1) * (m - 1)$ 。F 值为: $F = (SS_t / df_t) / (SS_e / df_e)$, 若 $F > F(df_t, df_e)$, 则处理间存在显著性差异。用 LSD 法比较处理间均值差异。

1.1.3 有亚样品的随机区组设计

设试验有亚样品, 数据矩阵为 (x_{ijk}) , $i = 1, 2, \dots, n; j = 1, 2, \dots, h; k = 1, 2, \dots, m$, 其中, n 为处理数, m 为区组数, h 为亚样品数。计算总平方和 (SS_T), 亚样品平方和 (SS_u), 区组平方和 (SS_b), 处理平方和 (SS_t), 误差平方和 (SS_e), 及抽样误差平方和 (SS_s)。

$$\begin{aligned}
 SS_T &= \sum_{i=1}^n \sum_{j=1}^h \sum_{k=1}^m x_{ijk}^2 - C \\
 SS_u &= \sum_{i=1}^n \sum_{k=1}^m \left(\sum_{j=1}^h x_{ijk}^2 \right) / h - C \\
 SS_b &= \sum_{k=1}^m \left(\sum_{i=1}^n \sum_{j=1}^h x_{ijk}^2 \right) / (n * h) - C \\
 SS_t &= \sum_{i=1}^n \left(\sum_{k=1}^m \sum_{j=1}^h x_{ijk}^2 \right) / (m * h) - C \\
 SS_e &= SS_u - SS_b - SS_t \quad SS_s = SS_T - SS_u
 \end{aligned}$$

其中, $C = \left(\sum_{i=1}^n \sum_{j=1}^h \sum_{k=1}^m x_{ijk}^2 \right) / (n * h * m)$ 。总自由度, 亚样品自由度, 区组自由度, 处理自由度, 误差自由度, 以及抽样误差自由度分别为 $df_T = n * h * m - 1$, $df_u = n * m - 1$, $df_b = m - 1$, $df_t = n - 1$, $df_e = (n - 1) * (m - 1)$, $df_s = n * m * (h - 1)$ 。计算 $F = (SS_t / df_t) / (SS_e / df_e)$, 若 $F > F(df_t, df_e)$, 则处理间存在显著性差异, 用 LSD 法比较处理间均值差异。

1.1.4 巢式设计

设试验数据矩阵为 (x_{ijk}) , $i = 1, 2, \dots, n; j = 1, 2, \dots, h; k = 1, 2, \dots, m$, 其中, n 为处理数, m 为亚样品数, h 为区组数。计算总平方和 (SS_T), 处理平方和 (SS_t), 区组平方和 (SS_b), 误差平方和 (SS_e)。

$$\begin{aligned}
 SS_T &= \sum_{i=1}^n \sum_{j=1}^h \sum_{k=1}^m x_{ijk}^2 - C \\
 SS_t &= \sum_{i=1}^n \left(\sum_{k=1}^m \sum_{j=1}^h x_{ijk}^2 \right) / (m * h) - C \\
 SS_b &= \sum_{j=1}^h \left(\sum_{i=1}^n \sum_{k=1}^m x_{ijk}^2 \right) / m - SS_t - C \\
 SS_e &= SS_T - SS_t - SS_b
 \end{aligned}$$

其中, $C = \left(\sum_{i=1}^n \sum_{j=1}^h \sum_{k=1}^m x_{ijk}^2 \right) / (n * h * m)$ 。总自由度, 处理自由度, 区组自由度, 误差自由度, 分别为 $df_T = n * h * m - 1$, $df_t = n - 1$, $df_b = n * (h - 1)$, $df_e = n * h * (m - 1)$ 。计算 $F = (SS_t / df_t) / (SS_e / df_e)$, 若 $F > F(df_t, df_e)$, 则处理间存在显著性差异, 用 LSD 法比较处理间均值差异。

1.2 两因子设计

1.2.1 双向分组设计

设试验数据矩阵为 (x_{ijk}) , $i = 1, 2, \dots, n; j = 1, 2, \dots, m; k = 1, 2, \dots, r$, 其中, n 为 A 因子水平数, m 为 B 因子水平数, r 为重复数。计算总平方和 (SS_T), A 因子平方和 (SS_a), B 因子平方和 (SS_b), 处理组合平方和 (SS_t), A - B 因子交互平方和 (SS_{ab}), 及

误差平方和 (SS_e)。

$$\begin{aligned}
 SS_T &= \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^r x_{ijk}^2 - C \\
 SS_a &= \sum_{i=1}^n \left(\sum_{j=1}^m \sum_{k=1}^r x_{ijk}^2 \right) / (m * r) - C \\
 SS_b &= \sum_{j=1}^m \left(\sum_{i=1}^n \sum_{k=1}^r x_{ijk}^2 \right) / (n * r) - C \\
 SS_t &= \sum_{i=1}^n \sum_{j=1}^m \left(\sum_{k=1}^r x_{ijk}^2 \right) / r - C \\
 SS_{ab} &= SS_t - SS_a - SS_b \quad SS_e = SS_T - SS_t
 \end{aligned}$$

其中, $C = \left(\sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^r x_{ijk}^2 \right) / (n * m * r)$ 。总自由度, A 因子自由度, B 因子自由度, 处理组合自由度, A - B 因子交互自由度, 以及误差自由度分别为 $df_T = n * m * r - 1$, $df_a = n - 1$, $df_b = m - 1$, $df_t = n * m - 1$, $df_{ab} = (n - 1) * (m - 1)$, $df_e = (n * m * (r - 1))$ 。计算 A 因子、B 因子、A - B 因子交互的 F 值 $F_a = (SS_a / df_a) / (SS_e / df_e)$, $F_b = (SS_b / df_b) / (SS_e / df_e)$, $F_{ab} = (SS_{ab} / df_{ab}) / (SS_e / df_e)$ 。若 $F_a > F(df_a, df_e)$, $F_b > F(df_b, df_e)$, $F_{ab} > F(df_{ab}, df_e)$, 则 A 因子水平间、B 因子水平间、A - B 因子交互间存在显著性差异。用 LSD 法比较水平间均值差异, 计算 $S_a = (2 * (SS_e / df_e) / (m * r))^{1/2}$, $S_b = (2 * (SS_e / df_e) / (n * r))^{1/2}$, 则相应的 $LSD_a = t_a(df_e) * S_a$, $LSD_b = t_b(df_e) * S_b$, 此处 $a = 0.05, 0.01$ 。计算前述比较的均值差异, 若大于相应的 LSD_a , 则均值差异显著。

1.2.2 两因子试验 RCBD

设试验数据矩阵为 (x_{ijk}) , $i = 1, 2, \dots, n; j = 1, 2, \dots, m; k = 1, 2, \dots, r$, 其中, n 为 A 因子水平数, m 为 B 因子水平数, r 为重复数。计算总平方和 (SS_T), A 因子平方和 (SS_a), B 因子平方和 (SS_b), 处理组合平方和 (SS_t), A - B 因子交互平方和 (SS_{ab}), 区组平方和 (SS_u), 及误差平方和 (SS_e)。

$$\begin{aligned}
 SS_T &= \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^r x_{ijk}^2 - C \\
 SS_a &= \sum_{i=1}^n \left(\sum_{j=1}^m \sum_{k=1}^r x_{ijk}^2 \right) / (m * r) - C \\
 SS_b &= \sum_{j=1}^m \left(\sum_{i=1}^n \sum_{k=1}^r x_{ijk}^2 \right) / (n * r) - C \\
 SS_t &= \sum_{i=1}^n \sum_{j=1}^m \left(\sum_{k=1}^r x_{ijk}^2 \right) / r - C \\
 SS_{ab} &= SS_t - SS_a - SS_b \\
 SS_u &= \sum_{k=1}^r \left(\sum_{i=1}^n \sum_{j=1}^m x_{ijk}^2 \right) / (n * m) - C \\
 SS_e &= SS_T - SS_t - SS_u
 \end{aligned}$$

其中, $C = \left(\sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^r x_{ijk}^2 \right) / (n * m * r)$ 。总自由度, A 因子自由度, B 因子自由度, 处理组合自由度, A - B 因子交互自由度, 区组自由度, 以及误差自由度分别为 $df_T = n * m * r - 1$, $df_a = n - 1$, $df_b = m - 1$, $df_t = n * m - 1$, $df_{ab} = (n - 1) * (m - 1)$, $df_u = r - 1$, $df_e = (n * m - 1) * (r - 1)$ 。计算 A 因子、B 因子、A - B 因子交互及区组的 F 值 $F_a = (SS_a / df_a) / (SS_e / df_e)$, $F_b = (SS_b / df_b) / (SS_e / df_e)$, $F_{ab} = (SS_{ab} / df_{ab}) / (SS_e / df_e)$, $F_u = (SS_u / df_u) / (SS_e / df_e)$ 。若 $F_a > F(df_a, df_e)$, $F_b > F(df_b, df_e)$, $F_{ab} > F(df_{ab}, df_e)$, $F_u > F(df_u, df_e)$, 则 A 因子水平间、B 因子水平间、A - B 因子交互间以及区组间存在显著性差异。用 LSD 法比较水平间均值差异, 计算 $S_a = (2 * (SS_e / df_e) / (m * r))^{1/2}$, $S_b = (2 * (SS_e / df_e) / (n * r))^{1/2}$, 则相应的 $LSD_a = t_a(df_e) * S_a$, $LSD_b = t_b(df_e) * S_b$, 此处 $a = 0.05, 0.01$ 。计算前述比较的均值差异, 若大于相应的 LSD_a , 则均值差异显著。

1.3 多因子设计

1.3.1 裂区设计

设区组随机设计,主处理、亚处理固定。试验数据矩阵为 $(x_{ijk}), i=1,2,\dots,n; j=1,2,\dots,h; k=1,2,\dots,m$, 其中, n 为主处理数, h 为亚处理数, m 为区组数。计算总平方和 (SS_T) , 主区组平方和 (SS_m) , 区组平方和 (SS_r) , 主处理平方和 (SS_a) , 主区组误差平方和 (SS_{ea}) , 处理组合平方和 (SS_t) , 亚处理平方和 (SS_b) , 主处理-亚处理交互平方和 (SS_{ab}) , 及亚区组误差平方和 (SS_{eb}) 。

$$\begin{aligned} SS_T &= \sum_{i=1}^n \sum_{j=1}^h \sum_{k=1}^m x_{ijk}^2 - C \\ SS_m &= \sum_{i=1}^n \sum_{k=1}^m \left(\sum_{j=1}^h x_{ijk} \right)^2 / h - C \\ SS_r &= \sum_{k=1}^m \left(\sum_{i=1}^n \sum_{j=1}^h x_{ijk} \right)^2 / (n * h) - C \\ SS_a &= \sum_{i=1}^n \sum_{k=1}^m \left(\sum_{j=1}^h x_{ijk} \right)^2 / (m * h) - C \\ SS_{ea} &= SS_m - SS_r - SS_a \\ SS_t &= \sum_{i=1}^n \sum_{j=1}^h \left(\sum_{k=1}^m x_{ijk} \right)^2 / m - C \\ SS_b &= \sum_{j=1}^h \left(\sum_{i=1}^n \sum_{k=1}^m x_{ijk} \right)^2 / (n * m) - C \\ SS_{ab} &= SS_t - SS_a - SS_b \\ SS_{eb} &= SS_T - SS_r - SS_t - SS_{ea} \end{aligned}$$

其中, $C = \left(\sum_{i=1}^n \sum_{j=1}^h \sum_{k=1}^m x_{ijk} \right)^2 / (n * h * m)$ 。总自由度, 主区组自由度, 区组自由度, 亚处理自由度, 主区组误差自由度, 处理组合自由度, 亚处理自由度, 主处理-亚处理交互自由度, 以及亚区组误差自由度分别为 $df_T = n * h * m - 1, df_m = n * m - 1, df_r = m - 1, df_a = n - 1, df_{ea} = df_m - df_r - df_a, df_t = n * h - 1, df_b = h - 1, df_{ab} = (n - 1) * (h - 1), df_{eb} = n * (h - 1) * (m - 1)$ 。计算主处理、亚处理、主处理-亚处理互作的 F 值 $F_a = (SS_a / df_a) / (SS_{ea} / df_{ea}), F_b = (SS_b / df_b) / (SS_{eb} / df_{eb}), F_{ab} = (SS_{ab} / df_{ab}) / (SS_{eb} / df_{eb})$ 。若 $F_a > F(df_a, df_{ea}), F_b > F(df_b, df_{eb}), F_{ab} > F(df_{ab}, df_{eb})$, 则主处理间、亚处理间、主处理-亚处理交互间存在显著性差异。用 LSD 法比较处理间均值差异, 计算 $s_a = (2 * (SS_{ea} / df_{ea}) / (h * m))^{1/2}, s_b = (2 * (SS_{eb} / df_{eb}) / (n * m))^{1/2}, s_{ab} = (2 * (SS_{eb} / df_{eb} * (h - 1) + SS_{ea} / df_{ea}) / (h * m))^{1/2}$, 则相应的 $LSD_a = t_a(df_{ea}) * s_a, LSD_b = t_b(df_{eb}) * s_b, LSD_{ab} = s_{ab} * ((h - 1) * SS_{eb} / df_{eb} / df_{eb} * t_a(df_{eb}) + SS_{ea} / df_{ea} * t_a(df_{ea})) / ((h - 1) * SS_{eb} / df_{eb} + SS_{ea} / df_{ea})$, 此处 $a = 0.05, 0.01$ 。计算前述比较的均值差异, 若大于相应的 LSD_a , 则均值差异显著。

1.3.2 正交设计

设区组随机设计。试验数据矩阵为 $(x_{ij}), i=1,2,\dots,n; j=1,2,\dots,w+(m-w)+r$, 其中, n 为处理数, w 为因子数, m 为列数, r 为重复数。 x_{ij} 为因子水平, $i=1,2,\dots,n; j=1,2,\dots,w+(m-w)+r$; x_{ij} 为试验结果, $i=1,2,\dots,n; j=w+(m-w)+1,\dots,w+(m-w)+r$ 。设 q 为因子水平数。首先计算:

$$\begin{aligned} s_i &= \sum_{j=m+1}^{m+r} x_{ij} \quad (i=1,2,\dots,n) \\ kk_{ij} &= \sum_{k=1}^n s_k \quad (i=1,2,\dots,q; j=1,2,\dots,m) \quad x_{ij} = i \\ p_{ij} &= kk_{ij} / (w * r) \quad r_j = \max p_{ij} - \min p_{ij} \quad (i=1,2,\dots,q; j=1,2,\dots,w) \end{aligned}$$

计算总平方和 (SS_T) , 区组平方和 (SS_r) , 各因子平方和 (SS_j) , 列平方和 (SS_c) , 及误差平方和 (SS_e) 。

$$SS_T = \sum_{i=1}^n \sum_{j=m+1}^{m+r} x_{ij}^2 - C$$

$$\begin{aligned} SS_r &= \sum_{j=m+1}^{m+r} \sum_{i=1}^n \left(\sum_{j=1}^h x_{ij} \right)^2 / n - C \\ SS_j &= \sum_{i=1}^q kk_{ij}^2 / (r * q) - C \quad (j=1,2,\dots,w) \\ SS_c &= \sum_{j=w+1}^q \sum_{i=1}^q kk_{ij}^2 / (r * q) - C \\ SS_e &= SS_T - SS_r - SS_c - \sum_{j=1}^w SS_j \end{aligned}$$

其中, $C = \left(\sum_{i=1}^n s_i \right)^2 / (n * r)$ 。总自由度, 区组自由度, 因子自由度, 列自由度。误差自由度分别为 $df_T = n * r - 1, df_r = r - 1, df_j = q - 1, df_c = (m - w) * (q - 1), df_e = df_T - df_r - n + 1$ 。若 $F = (SS_r / df_r) / (SS_e / df_e)$ 不显著, 则令 $SS_e = SS_e + SS_r, df_e = df_e + df_r$ 。计算区组、因子、空列的 F 值 $F_r = (SS_r / df_r) / (SS_e / df_e), F_j = (SS_j / df_j) / (SS_e / df_e), F_c = (SS_c / df_c) / (SS_e / df_e)$ 。若 $F_r > F(df_r, df_e), F_j > F(df_j, df_e), F_c > F(df_c, df_e)$, 则区组间、因子间、空列间存在显著性差异。用 LSD 法比较因子水平间均值差异, 计算 $s_j = (2 * (SS_e / df_e) / (r * q))^{1/2}, j=1,2,\dots,w$; 则相应的 $LSD_a = t_a(df_e) * s_j$, 此处 $a = 0.05, 0.01$ 。计算前述比较的均值差异, 若大于相应的 LSD_a , 则均值差异显著。

2 算法实现

方差分析整合算法用 Java Applet 实现, 共 5 个类和 1 个 HTML 文件:

® VarianceAnalysis 类。方差分析的计算在该类中进行。该类同时调用其它类共同完成全部任务。Applet 被载入浏览器后, 显示算法选择窗口, 可任选上述 8 种算法, 选择后弹出 ParamInput 类窗口。最后, 打开原始数据文件。

® ParamInput 类。弹出该类的窗口后: (1) 若选择完全随机设计, 输入处理数, 重复数; (2) 若选择随机区组设计, 输入处理数, 区组数; (3) 若选择有亚样品的随机区组设计和巢式设计, 输入处理数, 区组数, 亚样品数; (4) 若选择双向分组设计和两因子试验 RCBD, 输入 A 因子水平数, B 因子水平数, 重复数; (5) 若选择裂区设计, 输入主处理数, 亚处理数, 区组数; (6) 若选择正交设计, 输入处理数, 空列数, 重复数, 因子数, 因子水平数。

® ResultShow 类, Hint 类以及 WarningShow 类, 见参考文献 [3~6] 所述。

® VarianceAnalysis.html 文件。该文件将 VarianceAnalysis 类载入网络浏览器, 并传入窗口大小及组件大小参数。

原始数据文件为普通的 MS-DOS 文本文件 (.txt)。

Applet 运行后输出方差分析的详尽结果。VarianceAnalysis 用功能强大的网络程序设计工具 Java 开发, 网络分布式计算, 用户界面友好, 不依赖于平台, 可即时更新 [3~6]。该软件占用内存少, 全部文件只有 118K 大小, 易于嵌入网络软件包及网络数据库。

3 实例分析

于 2003 年 6 月 12 日在清新县三坑镇水稻田取样调查无脊椎动物多样性, 用吸虫器收集稻田取样点内的无脊椎动物, 取 10 个样点。带回鉴定计数各样点的无脊椎动物种类和数量。最后, 合并为捕食性无脊椎动物, 寄生性无脊椎动物, 植食性无脊椎动物, 及中性无脊椎动物在各样点的数量, 得到原始数据文件。用完全随机设计的分析方法 (处理数 = 4, 重复数 = 10)。用 VarianceAnalysis 计算, 结果表明, 无脊椎动物类群均值间存在极显著差异 ($F = 33.451, a = 0.01$)。捕食性无脊椎动物、寄生性

无脊椎动物、植食性无脊椎动物间无显著差异。但中性无脊椎动物数量与前三个类群差异极显著。

参 考 文 献

- [1] 范福仁,生物统计学(修订本),江苏科学技术出版社,1980,164~248.
- [2] 中山大学数学力学系《概率论及数理统计》编写小组,概率论及数理统计(下册),高等教育出版社,1984,262~287.
- [3] 齐艳红、张文军,“有害生物侵扰在多样化环境中的一种随机扩散过程及网络计算软件”,《现代计算机》,2002,133:16~19.
- [4] 齐艳红,“图书期刊评价分析的混合序图及网络计算软件研究”,《现代计算机》,2002,151:14~16,56.
- [5] 齐艳红、张文军,“CorreDetector:一种用于信息资料相关性分析的网络共享软件”,《情报学报》,2003,22(Suppl.):266~268.
- [6] 张文军、齐艳红,“生物多样性和均匀度显著性的统计检验及网络计算软件”,《现代计算机》,2002,145:6~9.

(上接第70页)

找 Servant,并将 Servant 指派给 CORBA 对象。当一个对象被指派到一个 Servant 时,该抽象对象即为活动对象。Server 类与对象实现过程的关键步骤如下:

```
public class Server{
    .....
    public static void main(String[] args){
        try{
            .....
            初始化 ORB
            org.omg.CORBA.ORB orb = org.omg.CORBA.ORB.init(args,
            null);

            取根 POA 的一个引用
            POA rootPOA = POAHelper.narrow(orb.resolve_initial_references(
            ("RootPOA")));

            创建持久 POA,用户默认 Servant 请求处理和对象 ID 策略
            org.omg.CORBA.Policy[] BookOperatePolicies = {
            rootPOA.create_lifespan_policy(LifespanPolicyValue.PERSISTENT),
            rootPOA.create_request_processing_policy(RequestProcessingPolicyValue.USE_DEFAULT_SERVANT),
            rootPOA.create_id_uniqueness_policy(IdUniquenessPolicyValue.MULTIPLE_ID)};

            String name = "BookOperate";

            获取对象 ID
            byte[] BookOperateId = name.getBytes();

            创建 poaBookOperate POA
            POA poaBookOperate = rootPOA.create_POA(name + "_poa",
            rootPOA.the_POAManager(),BookOperatePolicies);

            创建 servant
            BookOperateImpl BookServant = new BookOperateImpl();

            显示激活对象
            poaBookOperate.activate_object_with_id(BookOperateId,BookServant);

            激活 POA 管理器
            rootPOA.the_POAManager().activate();

            创建对象引用
            org.omg.CORBA.Object book = poaBookOperate.servant_to_reference(BookServant);

            .....
        }catch(Exception ex){
            System.err.println(ex);
        }
    }
}
```

```
System.err.println(ex);
}
```

3.4 客户调用服务的过程

服务器实现了接口后,客户端就可以调用接口以达到检索的目的。使用 JSP 页面作为客户端的统一检索界面,完成相应的交互处理、结果显示和检索信息传送工作。在客户端定义了两个类 Connect 和 Bean,用户通过使用包含 Bean 类的 JSP 页面,生成 Connect 对象来完成对服务器接口的调用。在 Bean 类中主要完成将用户的检索条件及选择范围等信息经过格式化,传送给 Connect 类进行检索,并将检索结果返回给用户。

Connect 类中主要完成与服务器的连接以及接口的调用。客户端必须包含编译 IDL 模块生成的客户端的 IDL STUB。在初始化时,客户端同服务器端一样也要启动 ORB,然后通过使用客户端的 IDL STUB,根据 URL 查找对象,得到一个 BookOperate 对象实例,根据客户的检索要求调用不同的检索接口。在客户程序中不需要绑定到 URL 命名解析器,只需要在调用 bind 方法时简单地指定 URL 即可。客户端的用户使用证书验证方式实现安全机制,程序实现方式与过程同服务器端相似。图3描述了客户调用服务的过程。

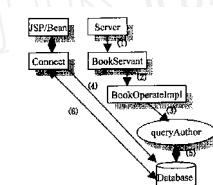


图3 客户调用服务的过程

首先由服务端启动对象和 ORB 总线,等待用户请求。当 Server 主类启动后,即创建了一个 BookServant 对象,获得对象的引用,并将服务器的 URL 绑定到该对象引用上,准备接收用户请求。客户端通过由 JSP 和 JavaBeans 构建成的 Web 页面,调用 Connect 类,通过绑定服务器对象的 URL 和端口获得服务器对象的引用以及安全证书的验证。调用对象提供的接口函数,连接数据库并进行数据检索,同时将检索结果返回给客户端 Web 页面显示。至此完成一项完整的客户端调用服务的过程。

4 小 结

目前数字图书馆中跨库检索问题已经成为数字图书馆界研究的热门之一。本文详细探讨了数字图书馆中分布式异构数据库的信息检索系统模型,并对其中的四个关键部分-IDL 映射、数据字段统一映射、服务器端服务对象和客户端调用作了技术实现。该模型对于目前各数字图书馆系统的集成提供了一个基于现有资源的集成方案,可以充分满足各数字图书馆自行管理的要求。此外利用此系统模型方案与现有各数字图书馆数据库提供的检索接口相结合,能够开发出更多功能的统一检索。使未来数字图书馆更具有人性化、更加易用。

参 考 文 献

- [1] Object Management Group, "The Common Object Request Broker: Architecture and Specification (Revision 3.0)", <http://www.omg.org/corba>.
- [2] Robert Orfali, Dan Harkey, "Client/Server Programming with Java and CORBA", 1997.
- [3] K. Fleming, S. Aslam Mir, J. Damstra, M. Vilicich, "Distributed Transactions using CORBA".
- [4] "Distributed Objects & Components: CORBA ORBs and other software", http://www.huihoo.com/corba/oo_object_request_brokers.html.