

Pengolahan Data Statistik Secara Simultan Menggunakan R-Package

Bambang Gonggo M. dan Mohammad Chozin
Fakultas Pertanian Universitas Bengkulu
bgonggo@yahoo.com

Pendahuluan

R-Package merupakan programming environment untuk analisis statistik yang bersifat interaktif. Dibanding dengan software sejenis berlisensi public lainnya, R-package merupakan software yang paling lengkap untuk menangani berbagai metoda analisis data secara statistik. Adanya sifat interaktif ini, pengguna dapat membuat sendiri berbagai fungsi atau rutin sesuai dengan kebutuhan. Fleksibilitas dari software ini juga tercermin dari kemampuannya untuk dijalankan pada komputer dengan prosesor tunggal (PC) maupun pada prosesor ganda melalui jaringan. Demikian juga, untuk data berukuran sangat besar, R-package telah menyediakan interface dengan server database sehingga data yang tersedia dapat diakses secara acak.

Secara default, proses pengolahan data dalam R-package dilakukan secara serial, yaitu setiap tahap pengolahan hanya mencakup satu metode analisis untuk satu variabel, kecuali untuk analisis multivariabel. Jika dalam analisis data melibatkan banyak metode dan banyak variabel, maka proses analisis menjadi tidak praktis. Oleh karena itu tujuan dari penelitian ini adalah untuk mengembangkan routine yang dapat digunakan untuk mengolah data secara simultan.

Metode

Dalam penelitian ini kami menggunakan 4 (empat) buah PC, yang terdiri dari 1 PC dengan prosesor Pentium 4 dan 3 PC dengan prosesor Pentium III yang saling dihubungkan sehingga membentuk sebuah klaster dengan topologi bintang. PC dengan prosesor Pentium 4 digunakan sebagai master dalam klaster sekaligus digunakan untuk menjalankan pengolahan data secara stand alone. R-Package yang digunakan adalah versi 1.9.1 (The R Foundation for Statistical Computing, 2004) dan dijalankan pada platform Mandrake Linux Versi 9.2. Pengujian routine yang dikembangkan dilakukan dengan menggunakan data simulasi.

Routine yang dikembangkan diuji untuk mengolah data secara simultan baik yang dilakukan dengan PC stand alone dan klaster PC secara paralel. Metoda statistik

yang dicoba adalah anova, manova, analisis korelasi dan analisis kovarian dan data yang digunakan terdiri dari 4- 8 variabel.

Hasil

1. Pengolahan secara simultan pada PC stand alone
 - a. Fungsi untuk analisis varians (anova) dan manava

```
manava <- function()
{
  mvdata <- read.csv("data2.csv", header=FALSE)
  mvrespon <- cbind(mvdata[,1],mvdata[,2],mvdata[,3], ...
)
  mvtreat <- gl(640,5)
  print("MANOVA")
  print(summary(manova(mvrespon ~ mvtreat), test="Wilks"))
}
```

- b. Fungsi untuk analisis korelasi dan kovarians

```
korelasi <- function()
{
  mydata <- read.csv("data2.csv", header=FALSE)
  matkorelasi <- cor(mydata)
  matkovarian <- cov(mydata)
  myresult <- data.frame(matkorelasi, matkovarian)
  print(myresult)
}
```

2. Pengolahan secara simultan pada kluster PC

```
olahpararel <- function()
{
  # mulai pewaktu
  ptm <- proc.time()

  # jalankan job() di slave
  mpi.bcast.cmd(source("jobs.R"), comm = 1)
  print(mpi.remote.exec(job(), comm = 1))

  # selisih waktu
  proc.time() - ptm
}
```

Routine untuk aktivasi anggota kluster seperti terlihat di atas. Sedangkan untuk pengolahan datanya disesuaikan dengan metode analisis yang akan dijalankan. Misalnya untuk anava dan manava routine yang digunakan sebagaimana pada pengolahan stand alone seperti berikut :

```
"jobs.R" # nama file yang berisi fungsi anava dan manava

manava <- function()
{
  mvdata <- read.csv("data2.csv", header=FALSE)
  mvrespon <- cbind(mvdata[,1],mvdata[,2],mvdata[,3],...)
  mvtreat <- gl(640,5)
  print("MANOVA")
  print(summary(manova(mvrespon ~ mvtreat), test="Wilks"))
}
```

Penutup

Dari routine yang ditampilkan di atas terlihat bahwa perbedaan pengolahan secara stand alone dan kluster PC terletak pada routine yang mengaktivasi anggota kluster. Sedangkan routine untuk pengolahan data secara statistik pada dasarnya sama

Ucapan Terima Kasih

Terima kasih kepada Ditjen Dikti Depdiknas atas bantuan dananya melalui Hibah Pekerti 2004/2005. Kepada Dr.Ir.Hermawan KD, MEE Departemen Teknik Fisika atas transfer teknologi Paralel Kluster PC-nya.

Pustaka

R Development Core Team. 2004. R Installation and Administration. The R Foundation for Statistical Computing Version 1.9.1. <http://cran.r-project.org/doc/manuals/R-admin.pdf>. Aug, 3 2003.