Parallel computation in R

Robert Stojnic <rs550@cam.ac.uk>
Laurent Gatto <lg390@cam.ac.uk>

Cambridge System Biology Centre
University of Cambridge

June 23, 2013

Parallel execution

Introduction

- Applicable when repeating *independent* computations a certain number of times; results just need to be combined after parallel executions are done.
- A cluster of nodes: generate multiple workers listening to the master; these workers are new processes that can run on the current machine or a similar one with an identical R installation. Should work on all R plateforms (as in package *snow*).
- The R process is *forked* to create new R processes by taking a complete copy of the masters process, including workspace (pioneered by package *multicore*). Does not work on Windows.
- Grid computing.

Parallel execution

Packages

- Package *parallel*, first included in `R` 2.14.0 builds on CRAN packages *multicore* and *snow*.
- Package *foreach*, introducing a new looping construct supporting parallel execution. Natural choice to parallelise a `for` loop.

parallel example

Dropin replacement for `*apply` functions

- `mclapply(X, FUN, ...)` (adapted from *multicore*).
- `parLapply(cl, X, FUN, ...)` (adapted from *snow*).

(demonstration)

foreach example

```
library(doMC)
library(foreach)
registerDoMC(2)
foreach(i = ll) %dopar% f(i)
foreach(i = ll) %do% f(i) ## serial version

library(plyr)
llply(ll, f, .parallel=TRUE)
```

(demonstration)

In Bioconductor

BiocParallel

This package provides modified versions and novel implementation of functions for parallel evaluation, tailored to use with Bioconductor objects. Currently in Bioc devel (2.12) only.

```
> library("BiocParallel")
> ll <- replicate(8, matrix(rnorm(1e6),1000), simplify=FALSE)
> f <- function(x) mean(solve(x), trim=0.7)
> p <- SnowParam(4L)
> bplapply(ll, f, BPPARAM = p)
```

References

Further reading

- Parallel R, McCallum and Weston, O'Reilly (2011).
- *parallel* and *foreach* vignettes.
- *High Performance Computing* CRAN task view.