

Experimental Features

Benilton Carvalho

April, 2009

1 Experimental Features

Starting on series 1.12.x, the `oligo` package offers (experimental) high performance computing features:

- **Support to larger datasets; and**
- **Support to parallel computing.**

These features are initially available for RMA methods on Expression/Gene/Exon arrays and will be implemented in other methods as necessity arrives.

The use of such features is as simple as loading the required packages (and setting a cluster option, if parallel computing is desired). The methods themselves are able to detect if these experimental features are enabled and use them if possible, without any modification of the method call.

2 Support to large datasets

The `oligo` package uses the features implemented by the `ff` package (preferably version $\geq 2.1-4$) to provide a better support to large datasets.

If the user prefers not to use the `ff` package, then regular R objects are used and the usual memory restrictions apply.

The support to large datasets is enabled by simply loading the `ff` package. Once that is done, `oligo` saves `ff` files to the directory pointed by `ldPath()`.

```
> library(oligo)
> library(ff)
> ldPath()
```

```
[1] "/nexsan/bst2/microarray/madman/Rpacks/oligo/inst/scripts"
```

Methods (`rma`) uses batches to process data. When possible (eg., background correction), it uses at most `ocSamples()` samples simultaneously at processing. For procedures that process probes (like summarization), a maximum of `ocProbesets()` are used simultaneously. Therefore, the user should tune these parameters for a better performance.

```

> ocSamples()

[1] 100

> ocSamples(50)
> ocProbesets()

[1] 1000

> ocProbesets(100)

> library(oligo)
> library(ff)
> rawData <- read.celfiles(list.celfiles())
> rmaRes <- rma(rawData)
> exprs(rmaRes)[1:10, ]

```

3 Parallel computing

The `oligo` package can make use of a parallel environment (with `rma` in the meantime) set via `snow` package, as long as the user:

- enables support to large datasets (load `ff`);
- loads the `snow` package;
- sets the cluster variable with `snow::makeCluster()`.

A simple example is shown below:

```

> library(ff)
> library(snow)
> options(cluster = makeCluster(2, "SOCK"))
> library(oligo)

> rawData <- read.celfiles(list.celfiles())
> rmaRes <- rma(rawData)
> rma

```

4 Details

This document was written using:

```

> sessionInfo()

```

R version 2.11.0 Under development (unstable) (2009-11-22 r50541)
x86_64-unknown-linux-gnu

locale:

[1]	LC_CTYPE=en_US.iso885915	LC_NUMERIC=C
[3]	LC_TIME=en_US.iso885915	LC_COLLATE=en_US.iso885915
[5]	LC_MONETARY=C	LC_MESSAGES=en_US.iso885915
[7]	LC_PAPER=en_US.iso885915	LC_NAME=C
[9]	LC_ADDRESS=C	LC_TELEPHONE=C
[11]	LC_MEASUREMENT=en_US.iso885915	LC_IDENTIFICATION=C

attached base packages:

[1]	tools	stats	graphics	grDevices	utils	datasets	methods
[8]	base						

other attached packages:

[1]	snow_0.3-3	ff_2.1-4	bit_1.1-4
[4]	oligo_1.11.39	oligoClasses_1.9.55	Biobase_2.7.5

loaded via a namespace (and not attached):

[1]	affxparser_1.19.6	affyio_1.15.2	Biostrings_2.15.25
[4]	DBI_0.2-4	IRanges_1.5.74	preprocessCore_1.9.0
[7]	splines_2.11.0		