Checking gene expression signatures against random and known signatures with *SigCheck*

Rory Stark

Edited: March 23, 2015; Compiled: October 14, 2015

Contents

1 Introduction

COMING SOON... [1]

2 Technical notes

2.1 Use of BiocParallel and parallel

This note shows how to control the parallel processing in *SigCheck*.

There are two different aspects of *SigCheck* that are able to exploit parallel processing. The primary one is when multiple signatures or datasets are being evaluated independently. This include the `nIterations` random signatures in `sigCheckRandom`, the database of known signatures in `sigCheckKnown`, and the `nIterations` permuted datasets in `sigCheckPermuted`. In this case, the *BiocParallel* package is used to carry out these comparisons in parallel. By default, *BiocParallel* uses *parallel* to run in multicore mode, but it can also be configured to schedule a compute task across multiple computers. In the default multicore mode, it will use all of the cores on your system, which can result in a heavy load (especially if there is inadequate memory). You can manually set the number of cores to use as follows:

```
> CoresToUse <- 6
> library(BiocParallel)
> mcp <- MulticoreParam(workers=CoresToUse)
> register(mcp, default=TRUE)
```

which limits the number of cores in use at any one time to six. If you want to use only one core (serial mode), you can set `CoresToUse <- 1`, or `register(SerialParam())`.

The other aspect of processing that can use multiple processor cores is when performing leave-one-out cross-validation (LOO-XV). In this case, the underlying *MLInterfaces* package takes care of the parallelization using the *parallel* package. You can set the number of cores that will be used for this as follows:

```
> options(mc.cores=CoresToUse)
```

Note that in the LOO-XV case, as every random or known signature, or permuted dataset, requires parallel evaluation of cross-validated classifiers, the parallelization at the level of `nIterations` is disabled automatically.

3 Acknowledgements

We would like to acknowledge everyone in the Bioinformatics Core at Cancer Research UK's Cambridge Institute at the University of Cambridge, as well as members of the Ponder group (particularly Kerstin Meyer), for their support and contributions.

4 Session Info

```
> toLatex(sessionInfo())
```

- R version 3.2.2 Patched (2015-10-08 r69496), `x86_64-apple-darwin10.8.0`
- Locale: `en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8`
- Base packages: base, datasets, graphics, grDevices, methods, stats, utils
- Other packages: BiocParallel 1.4.0
- Loaded via a namespace (and not attached): BiocStyle 1.8.0, futile.logger 1.4.1, futile.options 1.0.0, lambda.r 1.1.7, parallel 3.2.2, tools 3.2.2

References

[1] David Venet, Jacques E Dumont, and Vincent Detours. Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS computational biology*, 7(10):e1002240, 2011.