Scientific Computing :: Drop In

Arne Pommerening

Session: 7
Date: 16 November 2018
Activity: Extending R with C++

Extending R

In some data management or simulation situations, loops are unavoidable and a good strategy is to marry the strengths of R with the strengths of a higher programming language, e.g. C++ or Fortran. The key strategy is to outsource those parts of the code where we know that R has difficulties to process them or where we can expect R to be slow. Complex loops for example run much faster in C++ or Fortran and considerable time savings can be achieved with this strategy.

The Rcpp package facilitates the linking of R and C++. When working with Mac OS Xcode needs to be installed from the Apple Developer site (`https://developer.apple.com/xcode/`). On MS Windows you need to download and install Rtools (`http://cran.r-project.org/bin/windows/Rtools/`.

To get started, we re-program the calculation of the arithmetic mean in C++ (Rcpp). You can do this directly in RStudio by opening a new C++ file and type the following code:

```
#include <Rcpp.h>

using namespace Rcpp;

// [[Rcpp::export]]
```

```
double meanVector(NumericVector a) {

int n = a.size();
double xsum = 0;

for (int i = 0; i < n; i++)
xsum += a[i];

return xsum / n;
}
```

The C++ function `meanVector` is now ready for use in R. The Rcpp library is loaded in line 3. The code in line 5 is crucial, as the C++ code is loaded and compiled “on the fly”. This may take a few seconds. Potential C++ programming errors can now be viewed in the R console.

```
1  > rm(list = ls())
2  >
3  > library(Rcpp)
4  >
5  > sourceCpp("/Users/arnepommerening/Dropbox/Rcpp/
6  + meanVector.cpp")
7  > v <- c(1.5, 2.9, 3.2, 4.1, 5.5)
8  > mean(v)
9  [1] 3.44
10 > meanVector(v)
11 [1] 3.44
```

If the compilation was successful it is now possible to use the C++ function `meanVector` from within R. For this purpose we first construct a vector `v` in line 7. Then we use the R function `mean` to compute the arithmetic mean of the vector values. In a final step, the self-made function `meanVector` is applied to calculate the same mean of vector values in line 10 and the results in lines 9 and 11 coincide.

Reflect on this code, try to apply it to different data and finally extend the C++ code to do more complex tasks.