A stateful R web-interface, leveraging XML and XSLT

Richard P. Waterman

Analytic Business Services

And Department of Statistics, The Wharton School

University of Pennsylvania

Thomas A. Hurley

Analytic Business Services [*]

February 15, 2004

We will present work in progress that integrates R as the engine behind a stateful web friendly analytics package.

The work is funded by an NIH/SBIR grant to create user friendly software that will guide users through propensity score analysis for causal modeling (2). As such the likely users will come from a variety of backgrounds, for example epidemiology and marketing, where a command line driven interface will not work for most people. Further, providing a web browser based front end will reduce the ownership costs, both technologically and materially to many users.

One of the key design decisions behind the project is to only use open standards, and the project makes full use of the XML functionality now available in R (1) through the StatDataML representation of R objects.

In order to provide for a session memory a lightweight server termed the clientBrain (implemented in JAVA) has been created. The role of this server is to accept requests from the client (via a TomCat servlet engine) and pass these requests onto R, by way of an additional servlet termed the R-server.

Communication through to R is by socket connections, and we make full use of R's various connections; pipes and sockets as well as simple file reading and writing.

The presentation layer is governed by extensive use of XSLT, the Extensible Stylesheet Language Transformations. XSLT is used in two places – first the StatDataML representations are transformed into a canonical representation, and then this canonical representation is transformed into a browser friendly format, for example HTML or SVG.

These transformed XML documents are kept in memory on the clientBrain, so that additional requests from the browser for transformation based requests, for example sorting, are accomplished via dynamic XSLT edits, rather than resubmitting requests to R.

The rationale behind the two stage transform, first from StatDataML to a canonical XML form is that R is potentially one of a number of plug-in components to the architecture. Another natural one are user driven SQL queries to a database, the responses to which would not be in the StatDataML format.

We will share some of the difficulties that we have encountered and describe further possibilities for this architecture, in particular with regard to our business, which is an internet based quantitative market research company.

In summary, the philosophy of this implementation is to only use R for it's analytic capabilities, with all representation and presentation decisions relying on the XML/XSLT paradigm.

References

[1] D. MEYER, F. LEISCH, T. HOTHORN, AND K. HORNIK, *StatDataML: An XML format for statistical data*, Computational Statistics, (2004, Forthcoming).

[2] P. ROSENBAUM AND D. B. RUBIN, *The central role of the propensity score in observational studies for causal effects*, Biometrika, 70 (1983), pp. 41–55.