

FastRWeb: Fast Interactive Web Framework for Data Mining Using R

Simon Urbanek
AT&T Labs, Statistics Research Department

Vortrag im Rahmen des Institutskolloquiums
20. Mai 2009, 16:15 Uhr
Seminarraum, Ludwigstraße 33 I

R is widely used and accepted as a very versatile tool for statistical computing and data analysis. It provides a plethora of cutting edge methods, tools and algorithms. However, typical data analytic work entails the use of command line that requires a good knowledge of the language. On the other hand the World Wide Web (the Web) infrastructure represents a technology for wide deployment and high accessibility. With current browsers and the dynamic technologies such as AJAX it is now possible to create highly interactive content. Our basic goal is to fully leverage R and yet to allow users to interact with the system without the need to resort to the R language. Other interpreted languages are routinely used on the Web. R has seen a slower adoption in this area mainly due to the lack of high-level web support and its high start-up costs. We propose a framework that addresses both issues and allows very fast responses. It also provides building blocks not only for reports, plots and analyses, but also fully interactive graphics. In addition it is highly modular, allowing a maintainable creation of complex user interfaces for reporting, monitoring and data analysis.

In this talk we will describe the various parts of the system ranging from the fast-response infrastructure, AJAX tools for on-demand data loading to R facilities for intuitive creation of web objects, plots and interactive graphics. We will illustrate the use of the framework on several examples, including our implementation of a real-world mining tool used in practice for exploratory data analysis and data mining in very large databases. The use of Web-based methods allows us to use one system to target both users without statistical knowledge and domain experts.