

An automated R tool for identifying individuals with difficulties in a large pool of raters

Pete Meyer and Shaun Lysen
Google, Inc.
Santa Monica, California, USA

R is used extensively by the analysts at Google for analyzing everything from very small to very large datasets, from one-off analyses to regular production runs. In this talk we describe the use of R in flagging raters involved in the assessment of ad quality, who appear to be having difficulty performing their rating tasks. The use of this R script has resulted in an increase in system efficiency, improved timeliness of responding to rater needs, and decreased burden on those managing the raters.

The package RMySQL allows R to seamlessly integrate with MySQL databases, enabling data access directly to the production databases containing rater scores. Likewise, the R2HTML package provides output in a browser supported format, enabling report generation that can display web content and which enables movement between summary tables and supporting documentation using hyperlinks. Leveraging these features of R, we describe generating flags for three warning signs of rater difficulty:

1. excessive run lengths of repeated values,
2. the repetitive use of identical values for two distinct measures, and
3. identifying sequences of scores that appear to be assigned randomly rather than specific to the ads involved.

These tests could not be done by eye, either because of the large number of tasks involved or because they depend upon comparisons to reference distributions that are not visually apparent. However, those managing the raters easily grasp the conceptual basis for the tests and the summary tables contain hyperlinks to documentation that enables them to quickly find, cut and paste constructive feedback to the raters into emails in a simple and efficient manner. While these flags would be difficult to program in SQL, they are straightforward in R.