Using R to Model Click-Stream Data to Understand Users' Path To Conversion

Douglas Galagate[1*,2] , Professor Wolfgang Jank[1,3]

1. University of Maryland, College Park
2. Department of Mathematics
3. Robert H. Smith School of Business

*Contact author: galagate@math.umd.edu

Keywords: Click-stream data, data mining, advertising, internet

Advertisers spend a lot of time and effort directing consumers' attention to their client's website, enticing them to purchase a product or a service. While most models attribute a consumer's conversion only to the last site they visited ("last click model"), it is generally accepted that all of the information in a consumer's search path (i.e. in the sequence of all websites visited) play some role in the ultimate purchase decision. In this project, we use data mining techniques to characterize and model a consumer's "path to conversion." That is, we characterize the network of websites and sequence of clicks in order to gauge the impact of different types of online content, its order and its relationship to one another on the probability of a consumer's conversion. We use many *R* packages in the analysis.

References

Agresti, Alan (2002). Categorical Data Analysis

Wickham, Hadley (2009). ggplot2: Elegant Graphics for Data Analysis (Use R)

Liu, Bing (2010). Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data-Centric Systems and Applications)