

Modeling Consumer Behavior in Online Environments

Alan L. MONTGOMERY, *Tepper School of Business, Carnegie Mellon University, USA*, E-mail: alan.montgomery@cmu.edu

KEY WORDS: Clickstream Data, Hierarchical Bayesian Models, Multinomial Probit Model, Hidden Markov Chain Models, Vector Autoregressive Models

MATHEMATICAL SUBJECT CLASSIFICATION: 91B99

Abstract: Marketers are often interested in customizing offers and content based upon a consumer's interests and goals. A potentially rich source of information that can be used to make inferences about the consumer is clickstream data. Clickstream data provides information about the sequence of pages or the path viewed by users as they navigate a web site. In this talk I illustrate two potential applications of statistical modeling to make inferences about a consumer using clickstream data. The first focuses upon how path information can be used to make predictions about a web browser's purchase intention at a major online bookseller. Web browsing is modeled by categorizing pages and then predicting category choice using a dynamic multinomial probit model. My results show that the memory component of the model is crucial in accurately predicting a path. In comparison traditional multinomial probit and first-order markov models predict paths poorly. These results suggest that paths may reflect a user's goals, which could be helpful in predicting future movements at a web site. One potential application of this model is to predict purchase conversion. I find that after only six viewings purchasers can be predicted with more than 40% accuracy, which is much better than the benchmark 7% purchase conversion prediction rate made without path information. This technique could be used to personalize web designs and product offerings based upon a user's path. A second application of clickstream data is known as user profiling. The problem is to predict a consumer's characteristics using web site visitations. Consumers are likely to choose web sites that reflect their interests. For example, a user's gender could be predicted with a small sampling of the web sites visited over the course of a month. These applications illustrate that clickstream data contains a wealth of data about consumers. However this data poses many challenging statistical problems since the information is largely textual and graphical in nature. This talk is partially based upon joint work with Shibo Li, Kannan Srinivasan, and John C. Liechty.