Choice and the Internet

From Clickstream to Research Stream

Choice Symposium June 2001

The Clickstream Dream Team: Asim Ansari, David Bell, Randy Bucklin, Eloise Coupey, Sunil Gupta, Jim Lattin, John Little, Carl Mela, Alan Montgomery, Joel Steckel

Overview

• Focus: Choice behavior on the Internet
  ‣ How is this behavior different?
  ‣ New source of data: clickstream data
• How to think about Internet choice
• How to think about research questions
• Promising research questions
• Methods
• Data limitations
Internet Choice Behavior

- Every session is characterized by repeated request and response
  - Requests by the individual (for information, choice alternatives, etc.)
  - Response by the marketer (with information, new options, different presentation format, etc.)
- Each individual request constitutes a choice
  - Choice context depends upon previous series of requests and responses
  - Individual participates in constructing (or influencing) choice context

How is this different?

<table>
<thead>
<tr>
<th>Internet choice …</th>
<th>Contrast to scanner …</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intent is not clear</td>
<td>You go into a grocery store to buy</td>
</tr>
</tbody>
</table>
  - Search? Buy? | Passive |
  - Active | Fixed |
  - Addressable | Choice context is same for all |
    - Choice context is personalizable | Static |
  - Dynamic | Marketers can intervene at low cost |
Clickstream Data

- Measures of interactivity in electronic media
  - Record of movement in a site or across sites
- Currently takes the form of
  - Server log files of activity within specific web sites
  - Panel records of URLs visited (e.g., Media Metrix)
- Little academic research to date
  - Delays in obtaining access to data
  - Difficulties in cleaning, processing, filtering
  - Questions about what to study and to model (focus of our group discussion)

How to Think About Internet Choice

- What is the objective of the individual?
  - Search (information)
  - Purchase (transaction)
- What is the scope of the session?
  - Within site
  - Across sites
- Who is the client?
  - Individual
  - Marketer (web site owner)
How to Think About Research Issues

Does it require a new paradigm? What is new?
- Substantive
- Methodological
- Conceptual

Internet Choice Applications

<table>
<thead>
<tr>
<th>Search</th>
<th>Within</th>
<th>Across</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Search Engine</td>
</tr>
<tr>
<td>Purchase</td>
<td>E-commerce site</td>
<td>Shopbot</td>
</tr>
</tbody>
</table>
Promising Research Questions

- How do we infer an individual’s objective?
  - Is this a browser or a purchaser?
- How can I predict where an individual is going to click next?
  - Within site versus across site
- How do the active, constructive requirements of the Internet choice process change behavior?

More Research Questions

- What factors influence the Internet choice behavior of the individual?
  - What makes a site “sticky?” (Bucklin and Sismeiro 2001)
- Can we identify the “anti-persuadable” segment?
- What is the role of the Internet in the creation of individual consideration sets?
More Research Questions

• How do we personalize the interaction with the individual to increase response?
  ▪ Recommendation systems (Ansari et al 2000)
  ▪ E-customization (Ansari and Mela 2001)

• To what extent can we automate the process?
  ▪ Learn through adaptive control

More Research Questions

• Is customer lifetime value an appropriate metric to use to guide the design of a website?
  ▪ Impact on acquisition, retention, cross-selling

• How do we design a Shopbot?
  ▪ Montgomery et al (2001)
  ▪ Marketer as client versus individual as client

• How do we understand market structure on the Internet?
Methods

- What should be our research objectives in answering these questions?
  - Prediction versus explanation versus optimization
  - Scalability?
- Is there some potential in combining the strengths of data mining techniques with existing empirical approaches?
  - Collaborative filtering, hierarchical Bayes
  - Which should come first?

Data Limitations

Within site:
- Limited to single site interaction
- How do we know individual has achieved objective from session?
  - Is long or short session a sign of success?
- How do we treat rich media?
  - RealAudio not recognized as page view

Across site:
- What did the individual see on the page?
  - Truncated URL
- Lacks within-site interaction details
Data Limitations

- Clickstream is only a fraction of relevant behavior
  - Single source data does not yet exist

Conclusions

- To get started, we need to find ways to keep it simple
  - Can we start with Internet choice behavior that is predominantly within site with purchase as the objective?
- Need to draw on learning from other disciplines, domains
  - Literature on machine learning
  - Practitioners, data mining methods