Clickstream Analysis for Internet Marketing

Alan Montgomery
Associate Professor
Carnegie Mellon University
Graduate School of Industrial Administration

e-mail: alan.montgomery@cmu.edu

E-Marketing Success Drivers, 6 Mar 2001

1

Outline

- Principles of Internet Marketing
- Defining Clickstream Data
  - What does it tell us about an anonymous user?
- Analysis of Clickstream Data
  - User Profiling
  - Measuring Brand Equity: Choice Modeling
  - Cognitive Lock-In: Power Law of Practice
  - Forecasting Purchases using Clickstream Data

2

Are there any principles behind Internet Marketing?

Interactive Marketing/
1-to-1 Marketing/
Relationship Marketing/
Personalization/
Customer Relationship Management

3

Interactive Marketing

Blattberg and Deighton (Sloan Management Review 1991) discuss the differences between traditional mass marketing and interactive marketing:

- Use of actual behavior to identify customers/prospects
- Customized prices/product offers to the individual
- Customizing the advertising message via selected binding (split cable).
- Distribution: Direct links with the customer
- Sales force: Improved monitoring, less discretion, better data for the sales person

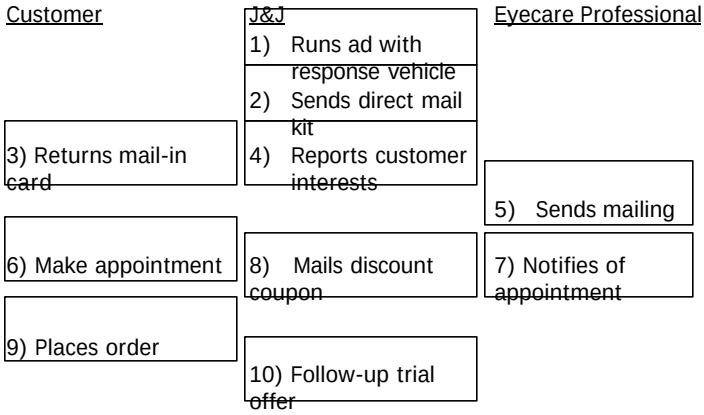
4

Comparing Mass and 1to1 Marketing

Product manager sells one product at a time to as many customers as possible	Sell as many products as possible to one customer at a time
Differentiate his products	Differentiate his customer
Acquire a constant stream of new customers	Tries to get a constant stream of business from current customers
Concentrates on economies of scales	Concentrates on economies of scope

Peppers and Rogers, **The One to One Future** 5

Acuvue Integrated Communications Program



6

The Web's Contribution

- Addressable marketing is not new
  - Mail and telephone have been around
  - For a long time the most valuable assets of companies like LL Bean, Fidelity Investments, American Express are their electronic customer transaction histories
- The difference between the web and these other technologies is that addressability and the collection of transaction histories is almost automatic

7

Interactive Marketing Requires...

- Ability to *identify* end-users
- Ability to *differentiate* customers based on their value and their needs
- Ability to *interact* with your customers
- Ability to *customize* your products and services based on knowledge about your customers

Peppers, Rogers, and Dorf (1999)

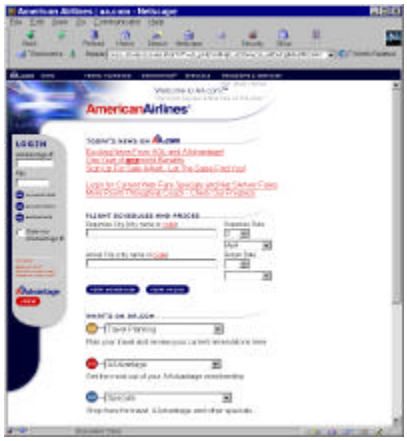
Information is key!

8

Example: American Airlines

- Updated in late 1998, has the capability to build custom pages for each of the airline's 2 million registered users
- Prior to the web, there was no cost-efficient way to tell millions of customers about a special fare available only this weekend, to a destination you personally will find attractive

Source



9

Definition of Clickstream Data

10

What is clickstream data?

- A record of an individual's movement through time at a web site
- Contains information about:
  - Time
  - URL content
  - User's machine
  - Previous URL viewed
  - Browser type

11

A clickstream example

Household ID: Female born 12Jul42

Demographics: Philadelphia Area, Male and Female Married (husband born: 27Sep46), 3 members in household, income: $75,000-$99,999, Graduated College, employed 35 or more hours, 1 child age 13 to 17 (daughter born: 5Jul80), own single family home, white collar, own car & truck, microwave, three dogs, five cats

18JUL97:18:55:57	47	www.voicenet.com/
18JUL97:18:56:44	37	www.weather.com/
18JUL97:18:57:25	105	www.weather.com/weather/us/cities/MI_Traverse_City.html
18JUL97:19:03:00	7	www.weather.com/weather/us/cities/MI_Traverse_City.html
18JUL97:19:03:56	2	www.weather.com/weather/us/cities/MI_Traverse_City.html
18JUL97:19:03:58	6	www.weather.com/weather/us/cities/HI_Lahaina.html
18JUL97:19:04:58	2	www.weather.com/weather/us/cities/HI_Lahaina.html
18JUL97:19:05:00	1	www.weather.com/weather/us/cities/MI_Traverse_City.html
18JUL97:19:15:24	39	www.weather.com/weather/us/cities/MI_Traverse_City.html
18JUL97:19:17:00	7	www.weather.com/weather/us/cities/MI_Traverse_City.html
18JUL97:19:17:07	13	www.realastrology.com/
18JUL97:19:17:20	44	www.realastrology.com/libra.html

12

What does the ISP see?

www.voicenet.com would see the following requests:

18JUL97:18:55:57	www.voicenet.com/
18JUL97:18:56:44	www.weather.com/
18JUL97:18:57:25	www.weather.com/weather/us/cities/MI_Traverse_City.html
18JUL97:19:03:58	www.weather.com/weather/us/cities/HI_Lahaina.html
18JUL97:19:17:07	www.realastrology.com/
18JUL97:19:17:20	www.realastrology.com/libra.html

13

What does the server see?

www.weather.com would see the following requests:

18JUL97:18:56:44	www.weather.com/
18JUL97:18:56:44	www.weather.com/weather/us/cities/HI_Lahaina.html
18JUL97:18:57:25	www.weather.com/weather/us/cities/MI_Traverse_City.html

14

Sources of clickstream data

- Web Servers
  - Each hit is recorded in the web server log
- Media Service Providers
  - DoubleClick, Flycast
- ISP/Hosting Services
  - AOL, Juno, Bluelight.com
- Marketing Research Companies
  - Media Metrix, NetRatings, PCData

15

Processing clickstream data

- IS Departments
- Site Activity Tracking Suppliers
- Auditing and Verification Suppliers
- Marketing Research Suppliers

What do you do with your data?

16

What can we say about an 'anonymous' user?

17

The Raw Clickstream Data

All information sent by my web browser when requesting http://www.privacy.net/analyze:

Accept: image/gif, image/x-xbitmap, image/jpeg, image/pjpeg, image/png, */*
Accept-Language: en Connection: Keep-Alive
Host: www.privacy.net
Referer: http://www.privacy.net/
User-Agent:
Mozilla/4.75 [en] (WinNT; U)
Pragma: no-cache
Cookie: Date=10/18/00; Privacy.net=Privacy+Analysis
Accept-Encoding: gzip
Accept-Charset: iso-8859-1,*,utf-8

18

Using Domain lookups

- If we match my domain, cmu.edu, with its registered zip code, “15213”, we can think about geodemographic marketing
- The most likely visitor from the “15213” zip is the University USA segment
- What can we do with a ZIP Code?

ENDS/MicroVision Lifestyle Game

19

User Profiling

What does ‘where you go’ say about ‘who you are’?

20

"On the Internet, nobody knows you're a dog."

New Yorker, 5 July 1993, p. 61

Clickstream Example: Who is this web person?

Example

Session Number	Start Time	Seconds Actively Viewing	Seconds Elapsed between start and stop	URL
1	06.Jul.1999:20:49:31	4	100	http://www.merrysbeanies.com/merrysbeanies/pokadex.htm
1	06.Jul.1999:20:51:11	52	132	http://www.merrysbeanies.com/merrysbeanies/pokadex.htm
1	06.Jul.1999:20:53:23	51	86	http://www.aol.com/netfind/kids/home.html
1	06.Jul.1999:20:54:49	51	219	http://www.merrysbeanies.com/merrysbeanies/pokadex.htm
1	06.Jul.1999:20:58:28	50	11768	http://www.blue-planet.com/fun/packman/
1	07.Jul.1999:00:14:28	51	74	http://find.web.aol.com/channelfind/mainchannelfind?query=eric+mcc
1	07.Jul.1999:00:15:42	50	162635	http://www.lycos.com/network/
2	08.Jul.1999:21:26:17	4	113	http://www.avonorder.com/ssl/logon.asp
3	08.Jul.1999:21:28:10	19	0	http://www.avonorder.com/ssl/verification.asp
4	08.Jul.1999:21:28:29	4	43	http://find.web.aol.com/channelfind/mainchannelfind?query=stargate
4	08.Jul.1999:21:29:12	1	601	http://members.aol.com/anee/
4	08.Jul.1999:21:39:13	2	65	http://members.aol.com/anee/
4	08.Jul.1999:21:40:18	36	0	http://rds.simplenet.com/stargate/links.htm
4	08.Jul.1999:21:40:54	172	172	http://www.e-net.or.jp/user/ncc/1701/rda/
4	08.Jul.1999:21:43:46	1	20	http://www.e-net.or.jp/user/ncc/1701/rda/bbs/99sg1.html
4	08.Jul.1999:21:44:06	1	5	http://www.e-net.or.jp/user/ncc/1701/rda/
4	08.Jul.1999:21:44:11	1	15	http://www.e-net.or.jp/user/ncc/1701/rda/bbs/99.html
4	08.Jul.1999:21:44:26	6	6	http://www.e-net.or.jp/user/ncc/1701/rda/bbs/99.html
4	08.Jul.1999:21:44:32	1	15	http://www.e-net.or.jp/user/ncc/1701/rda/
4	08.Jul.1999:21:44:47	1	16	http://www.e-net.or.jp/user/ncc/1701/nwase/osmsis.html

22

How much have you learned about this person?

• Gender	• Female
• Age	• 34 years old
• Race	• White
• Marital Status	• Single
• Geographic Location	• East South Central
• City Size	• 250,000-499,999
• Household Size	• 2 household members
• Household Composition	• Female head living with others related
• Household Income	• $25,000-$29,999
• Rent or Own	• Own
• Education	• Graduated High School
• Age and presence of children	• No Children Under 18
• Car or truck ownership	• Two cars, no trucks
• Dog or cat ownership	• No dogs or cats

23

What is this user's gender?

Web sites visited during one month:

48%	aol.com	63%	libertynet.org
64%	astronet.com	39%	lycos.com
75%	avon.com	27%	netradio.net
52%	blue-planet.com	57%	nick.com
56%	cartoonnetwork.com	59%	onhealth.com
54%	cbs.com	83%	onlinepsych.com
76%	country-lane.com	44%	simplenet.com
47%	eplay.com	76%	thriveonline.com
41%	halcyon.com	59%	valupage.com
70%	homearts.com	71%	virtualgarden.com
66%	ivillage.com	66%	womenswire.com

Percentage of female viewers using PC Meter data

24

Bayesian Updating Formula

Application of Bayesian hypothesis updating.

$$\overline{\overline{p}} = \frac{p \cdot \overline{p}}{p \cdot \overline{p} + (1-p)(1-\overline{p})}$$

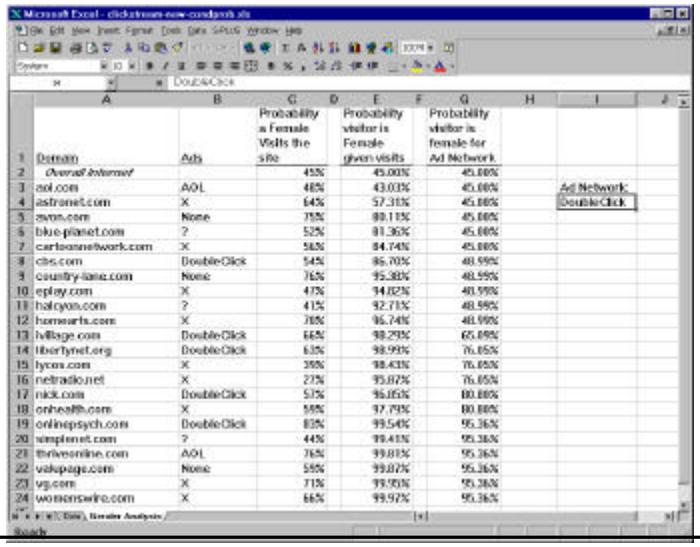
New probability

New information

Old probability

25

Statistical Analysis



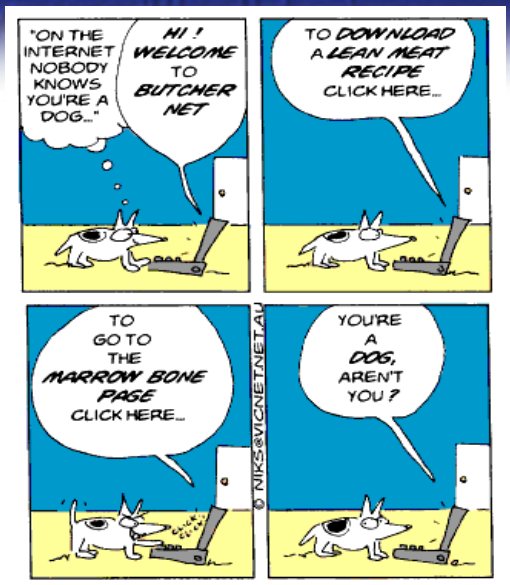
	Domain	Ads	Probability a Female Visits the site	Probability visitor is Female given visits	Probability visitor is female for Ad Network	I
1						
2	*Overall Internet*		45%	45.00%	45.00%	
3	aol.com	AOL	48%	43.03%	45.00%	Ad Network:
4	astranet.com	X	64%	57.31%	45.00%	DoubleClick
5	avon.com	None	75%	88.11%	45.00%	
6	blue-planet.com	?	52%	81.36%	45.00%	
7	cartoonnetwork.com	X	56%	84.74%	45.00%	
8	cbs.com	DoubleClick	54%	86.70%	48.99%	
9	country-lane.com	None	76%	95.38%	48.99%	
10	epiay.com	X	47%	94.02%	48.99%	
11	halcyon.com	?	41%	92.71%	48.99%	
12	homearts.com	X	78%	96.74%	48.99%	
13	hvillage.com	DoubleClick	66%	98.29%	65.09%	
14	iberlynet.org	DoubleClick	63%	98.99%	76.65%	
15	lycos.com	X	39%	98.43%	76.65%	
16	netradio.net	X	23%	95.87%	76.65%	
17	nick.com	DoubleClick	57%	96.05%	80.88%	
18	onhealth.com	X	59%	97.79%	80.88%	
19	onlinepsych.com	DoubleClick	83%	99.54%	95.36%	
20	simplenet.com	?	44%	99.41%	95.36%	
21	thriveonline.com	AOL	76%	99.81%	95.36%	
22	valupage.com	None	55%	99.87%	95.36%	
23	vg.com	X	71%	99.95%	95.36%	
24	womenswire.com	X	66%	99.97%	95.36%	

26

Results

- Analysis shows that there is a >99% probability this user is female.
  - Using only DoubleClick sites the probability is 95%.
- Using all user data for one month:
  - 90% of men are predicted with >80% confidence (81% accuracy)
  - 25% of women are predicted with >80% confidence (96% accuracy)

27



Http://www.moreinfo.com/au.cranlerma/fol2.htm

What is the value of profiling

Example

No targeting

Probability

Prediction \ Truth	Male	Female
Male	40%	10%
Female	25%	25%

Reward

Prediction \ Truth	Male	Female
Male	$ 0.50	$ 0.50
Female	$ 0.50	$ 0.50

Expected Value $ 0.50

Targeting

Probability

Prediction \ Truth	Male	Female
Male	40%	10%
Female	25%	25%

Reward

Prediction \ Truth	Male	Female
Male	$ 1.00	$ 0.20
Female	$ 0.20	$ 0.50

Expected Value $ 0.60

29

Choice Modeling

"The Great Equalizer? The Role of Shopbots in Electronic Markets" by Erik Brynjolfsson and Michael D. Smith (2000)

30

Which would you choose?

Amazon $43.90 **A** 5-10 days	Kingsbooks.com $32.06 **B** 16 days
1Bookstreet.com $35.96 **C** 6-21 days	Barnesandnoble.com $37.19 **D** 5-9 days

31

Modeling the Choice Decision with a Multinomial Logit

$\exp\{Value_A\} = .5 - .2\ Price - .02\ Delivery = .0002$

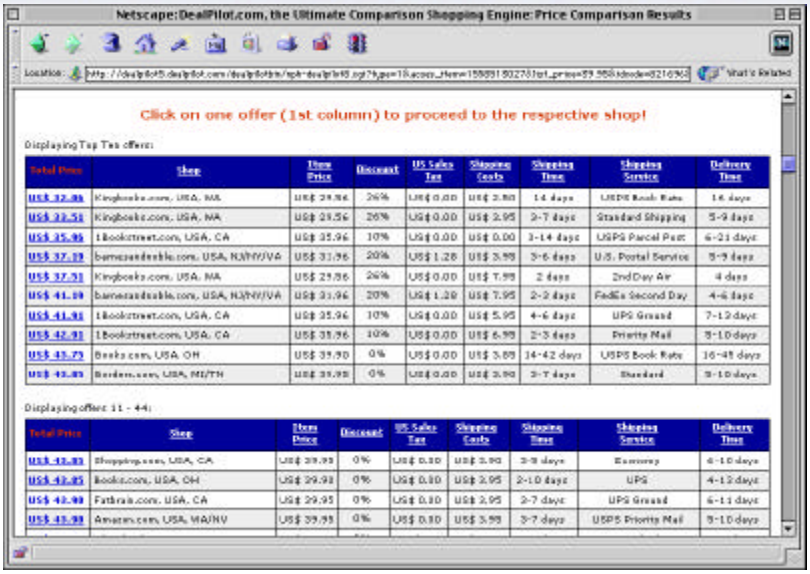
$\exp\{Value_B\} = 0 - .2\ Price - .02\ Delivery = .0010$

$\exp\{Value_C\} = 0 - .2\ Price - .02\ Delivery = .0008$

$\exp\{Value_D\} = .2 - .2\ Price - .02\ Delivery = .0003$

$$\text{Probability of Choosing A} = \frac{e^{.0002}}{e^{.0002} + e^{.001} + e^{.0008} + e^{.0003}} = 7\%$$

32



Data

Offer Data:	Total Price, Item Price, Shipping Cost, US Sales Tax, Shipping Time, Delivery Time, Delivery "n/a"
Session Data:	Date/Time, ISBN, IP Number, Sort Key
Customer Data:	Cookie Number, Cookies On (97.1%), *Prior Visits, Prior Choices*
Choice Data:	Click-Through, Last Click-Through

Data Statistics

- 69 Days (August 25 - November 1, 1999)
- 1,513,439 Offers
- 39,654 Sessions
- 20,227 Unique Visitors (7,478 Repeat Visitors)

34

Shopbot Choice Model

Parameter	Estimate	
Total Price		
Item Price	-.19	($1.00)
Shipping Price	-.37	($1.95)
U.S. Tax	-.43	($2.26)
Delivery Average	-.02	($.10)
Delivery "n/a"	-.37	($1.94)
"Big 3"		
Amazon	.48	($2.52)
BarnesandNoble	.17	($.89)
Borders	.27	($1.42)

35

Implications

- Customers approximately twice as sensitive to changes in shipping charges and sales tax as item price
- Price and delivery time are important
  - "Big 3" holds $1.13 (.284/.252) advantage over unbranded offers
  - Amazon.com holds $1.85 advantage over unbranded offers
  - BN/Borders hold $0.72 advantage over unbranded offers
- Shopbot customers highly price sensitive, how generalizable are these results?

36

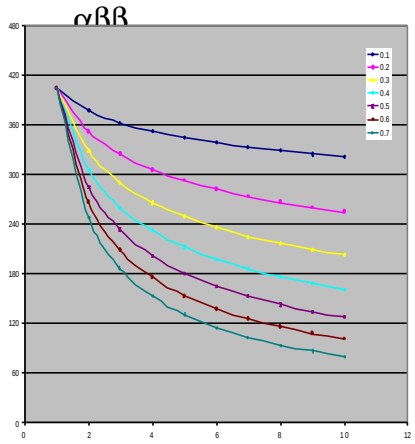
The Power Law of Practice

"Cognitive Lock In" by Eric Johnson, Jerry Lohse, and Steve Bellman (2000)

37

Power Law of Practice: $T=\beta N^{\alpha}$

- Macro 'empirical regularity'
- Describes increases in performance in many domains.
  - Text editing and other HCI tasks. (Newell and Rosenbloom, 1981
  - Learning curve for firms(Zangwill & Cantor 1998)



38

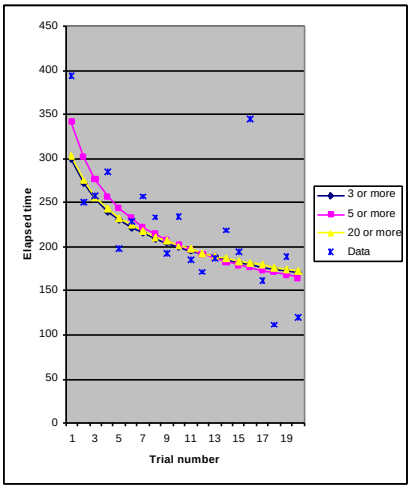
The argument

- By using a site, I learn, lowering the cost of further use.
  - Analogy to learning a supermarket, software,
  - Raises relative switching costs.
- The site (designer) and I coproduce solutions.
- Lower costs → More learning → More loyalty, purchases, etc.

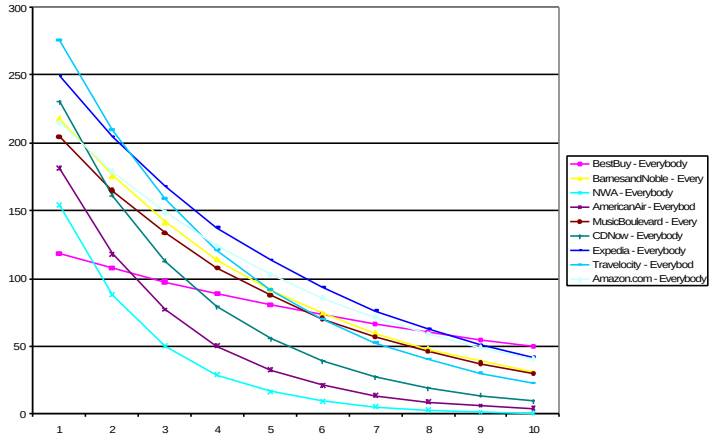
39

Illustrative Example: Visits to Amazon

- MediaMetrix: 6 months, all visits.
- Power law provides a good fit $r^2 = .45$
- Note decrease between 1st and 5th visit is about 200 seconds.
- Time savings = $1.44



Power Law Learning Curves



41

Implications

- Rapid site redesign is bad.
  - Refresh content, not navigation.
  - Simpler navigation is asset.
- Raising the cost of price shopping
- Brand (interface) extension may be a viable strategy.

42

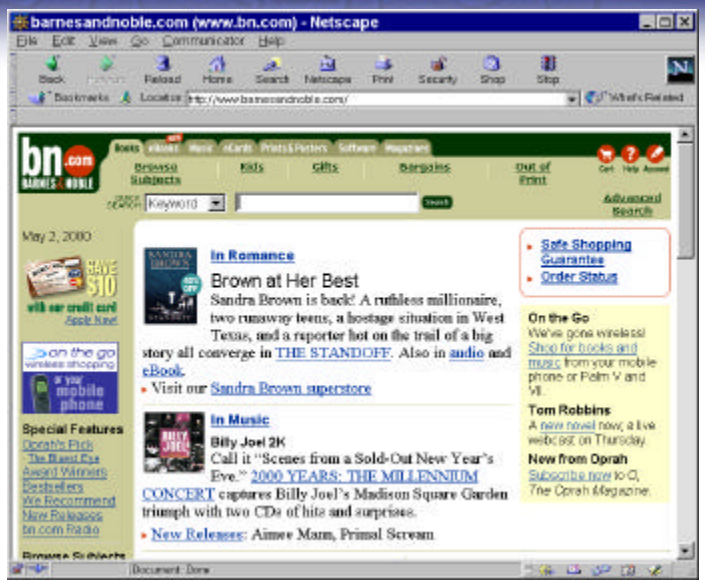
Using the clickstream to predict purchases

43

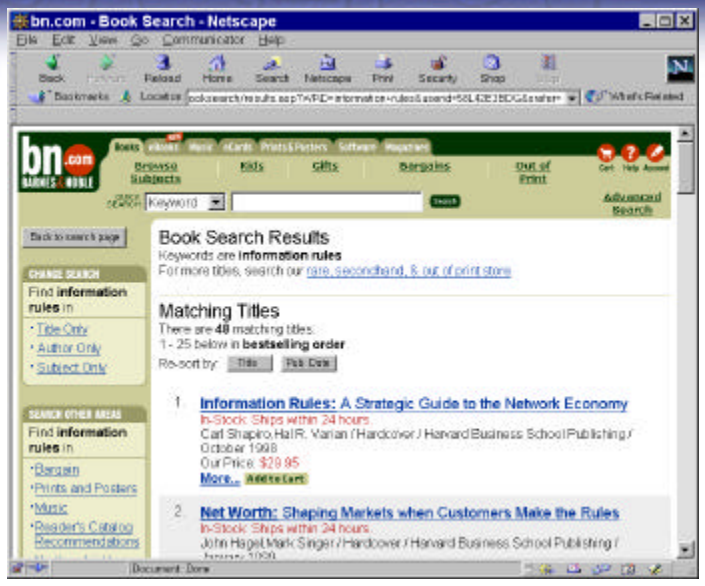
Problem

- How do we predict (or influence) a purchase decision?
- How much information is there in the path that an individual takes about purchase? Price sensitivity?
- What is an appropriate form of the model?

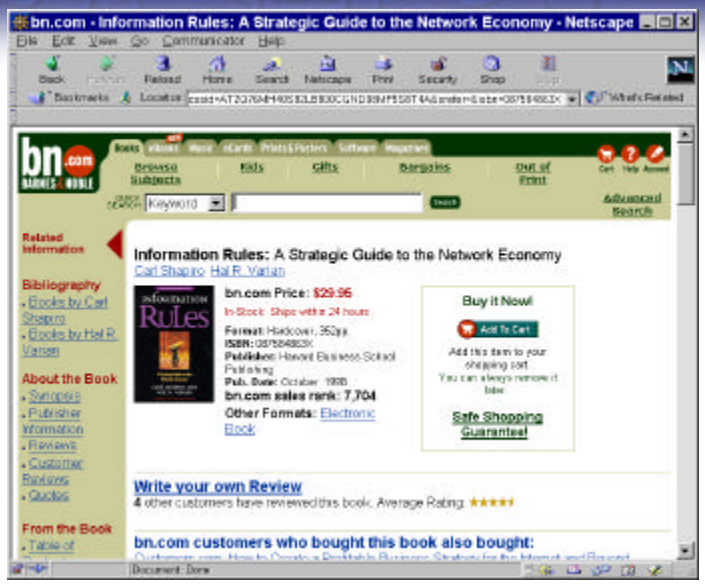
44


barnesandnoble.com (www.bn.com) - Netscape
bn.com
Browse Subjects
Kids
Gifts
Bargains
Out of Print
Keyword
Advanced Search
May 2, 2000
In Romance
Brown at Her Best
Sandra Brown is back! A ruthless millionaire, two runaway teens, a hostage situation in West Texas, and a reporter hot on the trail of a big story all converge in THE STANDOFF. Also in audio and eBook.
Visit our Sandra Brown superstore
In Music
Billy Joel 2K
Call it "Scenes from a Sold-Out New Year's Eve." 2000 YEARS: THE MILLENNIUM CONCERT captures Billy Joel's Madison Square Garden triumph with two CDs of hits and surprises.
New Releases: Aimee Mann, Primal Scream
Safe Shopping Guarantee
Order Status
On the Go
We've gone wireless! Shop for books and music from your mobile phone or Palm V and VII.
Tom Robbins
A new novel now; a live webcast on Thursday.
New from Oprah
Subscribe now to O, The Oprah Magazine.
Special Features
Oprah's Pick
Award Winners
Bestsellers
We Recommend
New Releases
bn.com Radio
Document: Done

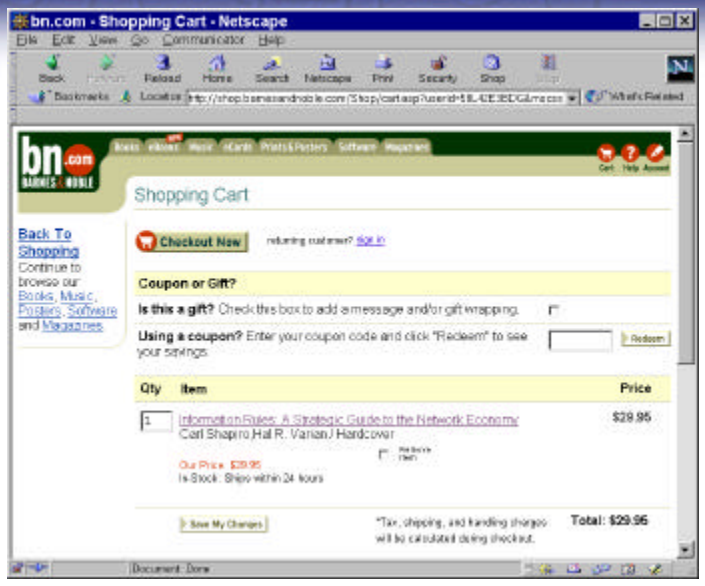
45


bn.com - Book Search - Netscape
bn.com
Browse Subjects
Kids
Gifts
Bargains
Out of Print
Keyword
Advanced Search
Back to search page
Book Search Results
Keywords are information rules
For more titles, search our rare, secondhand, & out of print store
Matching Titles
There are 48 matching titles
1 - 25 below in bestselling order
Re-sort by:
1. Information Rules: A Strategic Guide to the Network Economy
In-Stock: Ships within 24 hours
Carl Shapiro,Hal R. Varian / Hardcover / Harvard Business School Publishing / October 1998
Our Price: $29.95
More... Add to Cart
2. Net Worth: Shaping Markets when Customers Make the Rules
In-Stock: Ships within 24 hours
John Hagel,Mark Singer / Hardcover / Harvard Business School Publishing /
CHANGE SEARCH
Find information rules in
Title Only
Author Only
Subject Only
SEARCH OTHER AREAS
Find information rules in
Bargain
Prints and Posters
Music
Reader's Catalog Recommendations
Document: Done

46


bn.com - Information Rules: A Strategic Guide to the Network Economy - Netscape
Information Rules: A Strategic Guide to the Network Economy
Carl Shapiro Hal R. Varian
bn.com Price: $29.95
Buy it Now!
Write your own Review
bn.com customers who bought this book also bought:

47


bn.com - Shopping Cart - Netscape
Shopping Cart
Back To Shopping
Checkout Now
Coupon or Gift?
Qty
Item
Price
Total: $29.95

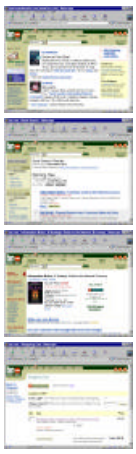
48

Motivation

- Assume that our independent variable is a latent measure of interest (e.g., utility) instead of time web used
- Again relate current interest to past values of interest and unpredictable changes
- Also incorporate measures of information and content about a web page
- Finally, assume the coefficients follow a switching model for 'surfing' behavior and 'goal' oriented behavior

49

Illustrating the Model



Interest or Value

$V_1 = -.7$

Previous Value *Effect of Page*

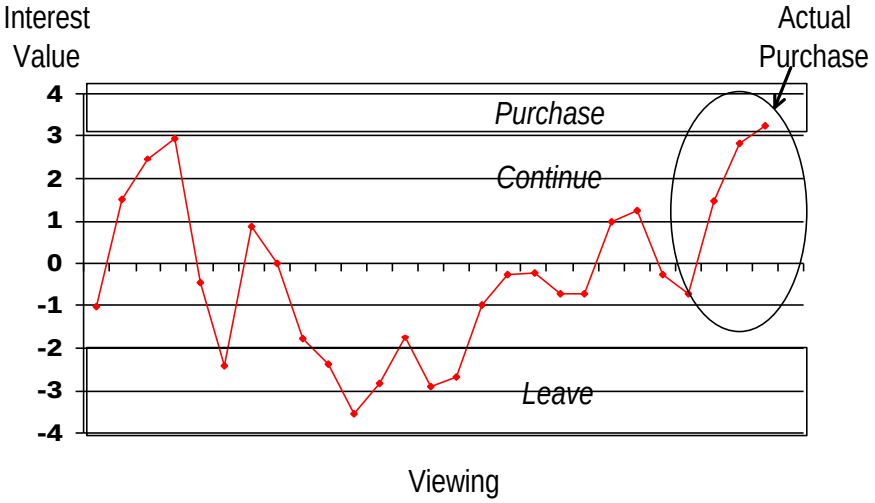
$V_2 = V_1 + 2.2 = 1.5$

$V_3 = V_2 + 1.3 = 2.8$

$V_4 = V_3 + .4 = 3.2$

50

How the model works



51

Implementation

- Statistical models can describe modeling very well
- The biggest challenge is extracting textual information and navigation information from the web page

52

Review

53

Summary

- A foundation for Internet Marketing can be found in One-to-One Marketing
- Information is critical, the most valuable information is purchase history
- New and better measures of the information content of clickstream data and customer acquisition costs are needed

54

Interactive Marketing Requires...

- Identify
  - Clickstream data can be used to identify your visitors
- Differentiate
  - User Profiling can be used to differentiate them
- Interact
  - Choice modeling can predict what they want
- Customize
  - The next step is responding to this information, will explore this in the area of web advertising

55