The Predictive Enterprise: Creating Transparency and Making Decisions

Andreas S. Weigend, Ph.D.
www.weigend.com

27 September 2006
IIR Business Intelligence Forum 5

The Predictive Enterprise

• Pervasive Business Intelligence
  ◦ Metrics driving performance
• Data as the basis for Predictive Models
  ◦ From sales data to the attention stream
• Agility and a Culture of Experimentation
  ◦ The perfect product, or a permanent beta?
• Network Collective Intelligence
  ◦ What is the relevance of "Web2.0" to the enterprise?

• Consider
  ◦ Impact on structure of organization
  ◦ Communication, including mobile

1986 | 1996 | 2006
---|---|---
Read-only web | Read-write web | Pretest and validate
“Consumer”, “User” (increase barrier of exit) | “Producer” (Lower barrier of entry) | Launch and learn
Pretest and validate | Taxonomy (controlled) | “Tagsonomy”
Static Navigation spatial: link | Navigation temporal (time as ordering parameter) | Discovery based on hyperlinks (expression of author’s attention and control).
Broadcast email, newsletters | Discovery based on social relations (trust, reputation), and metadata by other humans. RSS. | Discovery based on social relations (trust, reputation), and metadata by other humans. RSS.
Pay for space | Pay for presentment | Pay for click, for action.
Pay for clicks | Pay for presentment | Pay for click, for action.

Examples

• Money economy → Intention economy → Attention economy
  ◦ Information economy → Recommendation economy, Platform economy
  ◦ Use attention stream: Attensa
  ◦ Use context: Cleverset
• Paying a few experts to create content → User generated content
  ◦ Implicit data → explicit data; Metadata
  ◦ Examples: Music, Tags
  ◦ Add social, psycho, demographics: Nugg.ad
  ◦ Currencies (itunes vs Spiralfrog)
• Algorithmic search → Social search
  ◦ People answer questions: Google Answers, MSN, Yahoo
  ◦ Use info in files on your computer to determine relevance: Illumio
• Discovery → Social discovery
  ◦ How → Long tail
  ◦ URLs: Del.icio.us
  ◦ Photos: Flickr
• Global → Local
  ◦ WLAN: Jambo.net
  ◦ Mobile: Socialight

Increasing communication

• 0. Business metrics and data collection
  ◦ An iterative process
• 1. Data Analysis
  ◦ Data mining: Description, prediction
• 2. Architectures of Experimentation
  ◦ A/B test, active learning, survey design...
• 3. Architectures of Participation
  ◦ Remember, share, discover
  ◦ Empower and incentivize people to contribute
  ◦ Self-expression
• 4. Architectures of Interaction

River Nile
**The Era of Faith**
- Massive investments into cathedrals etc.
- Unclear ROI (Return on Investment)
- No feedback, or long feedback cycle

**The Era of Data**
- Massive investments into measuring, networking, storing
- ROI measurable
- Short feedback cycle

**Characteristics of our Era**
- What do we do with data?
- Gather data
- Explore data
- Publish data
- Archive data
- Data too much of it...
- Gather data
- Explore data
- Publish data
- Archive data

**What does it mean?**
- how can I act on it...
- will it integrate with my systems?...
- No feedback, or long feedback cycle

**Turning behavior into data**
- Search
- Online dating
- Music

**Everything can and will become data**
- Additional sources of data about people
  - Movement
  - Brain activity
  - Identity of person
  - Mobile phones
  - Brain activity
  - DNA analysis

**Opportunities and challenges for marketers, publishers, agencies...**
- How can I act on it...
- will it integrate with my systems?...
- How can I act on it...
- will it integrate with my systems?...
RFIDs and e-business
电子标签技术和电子商务

- **Unique Identifiers for Physical Objects**
  - **Facts**
    - Barcodes
    - Price: 2 cent
    - Size: 2 mm
  - **Opportunities**
    - Supply chain
      - China: 1.5 billion savings expected per year by using RFIDs
      - Wal-Mart: $ 8 billion savings expected per year by using RFIDs
      - Need: Understand pros and cons and make conscious decisions

- **Opportunities**
  - Supply chain
    - China: 1.5 billion savings expected per year by using RFIDs
    - Wal-Mart: $ 8 billion savings expected per year by using RFIDs
    - Loss of privacy
    - Concentration
  - Reality
    - Wal-Mart: $ 8 billion savings expected per year by using RFIDs
  - Compare to web sites
    - Implicit: (Clicks etc.)
    - Explicit: (Surveys etc.)

- **Fears**
  - Loss of privacy
  - Dictatorship

- **Reality**
  - Wal-Mart: $ 8 billion savings expected per year by using RFIDs

- **Dramatic drop in price**
  - Storage is free
  - Exponential increase in storage

- **Why now?**
  - Malthus's Law of Information:
    - New information content is doubling every year
    - Time spent on information consumption is constant

- **Large e-business company: Amount of data created per year**
  - Level
    - Customer
    - Orders
    - Session aggregates
    - Clicks
  - New data per year
    - 100 GB
    - 10 TB
    - 100 TB
  - Amount of data
  - Storage
  - Communication

- **Data collected implicitly:**
  - Dramatic growth over time
  - Communication
  - Storage
  - Cost
  - Time
The iterative process of modeling and decision making

1. Define Objectives
   • Objectives and benchmarks

2. Measure
   • Collect, store, manage the data

3. Describe
   • Exploratory data analysis

4. Predict and evaluate
   • Probabilistic models

5. Decide, act, and evaluate

Why is it hard?
• Even simple behavioral analysis requires significant infrastructure
• Reporting
• Behavioral analysis and predictive modeling
• Cost center
• Profit center

Increasing communication: 4 levels
• Data Analysis
  • Data mining: Description, prediction

• Architectures of Experimentation
  • A/B test, active learning, survey design...

• Architectures of Participation
  • Remember, share, discover
  • Empower and incentivize people to contribute
  • Self-expression

• Architectures of Interaction

Orders
Overall use of the site
Customer service contacts
Survey
Satisfaction
Intention / Goals / Modalities
Customer service response
Resolution
Number of items returned
Web page generation time
Actual delivery date
E-mail campaigns and responses
Also want to make recommendations! 添加推荐!

Customers who bought Kind of Blue also bought:
- Time Out — Dave Brubeck, et al
  Our Price: $6.99
  Sold for most from $3.79
- Calypso — John Coltrane
  Our Price: $13.99

- Complements (Buy in addition to)
  - Customers who bought X also bought Y

- Substitutes (Buy instead of)
  - Customers who shopped for X also shopped for Z

Result: Right vs Left
对比结果: 左还是右

- Metrics:
  - Conversion rate: Percentage of visits placing an order
  - Order size: Number of additional (from the second page) items put in cart

Result:
- "Your Shopping Cart" on right is about 1% better than on left
- "Your Shopping Cart" on right is about 1% better than on left

Some details:

<table>
<thead>
<tr>
<th></th>
<th>All customers</th>
<th>Existing customers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cart adds from 2nd page</td>
<td>+0.6%</td>
<td>+0.8%</td>
</tr>
<tr>
<td>DVD (s)</td>
<td>+1.1%</td>
<td></td>
</tr>
<tr>
<td>Wishlist adds</td>
<td>+1.4%</td>
<td></td>
</tr>
<tr>
<td>Cart adds from 2nd page</td>
<td>+0.8%</td>
<td></td>
</tr>
<tr>
<td>DVD (s)</td>
<td>+1.1%</td>
<td></td>
</tr>
<tr>
<td>Wishlist adds</td>
<td>+1.4%</td>
<td></td>
</tr>
</tbody>
</table>
Long-term Effects

Displaying Google’s “sponsored links” within Amazon.com’s site:

- Initial effect positive...
- ... but how to model whether this helps or hurts the company in the long run?

Offer free shipping:

- Competitors quickly adopt, how to model the overall game?

Pricing, Promotions...

- Role of price comparison engines (shopping.com, etc.)

Pricing, Promotions...

- Role of price comparison engines (shopping.com, etc.)

Survey

- "… Please let us know what you plan to do at Amazon.com today”
- "… Tell us what you are planning to do at the Amazon website today”

Free responses, hand-coded into non-exclusive categories:

- Multiple assignments possible
- Average: 1.2 categories per response
- Number of responses: 1023
- Response rate: 3.1%
- Date: February 11-12, 2003

Why do People Visit?

35% Research: Response indicates having a target or topic for which customer wants to gather information in depth (”looking for books about Chile”)

31% Browse: Looks at items casually at a more general level than Research (”looking at music/books”, without further specification)

16% Buy: Response indicates intent to buy in this visit (”buy a book”)

10% Complain: Complains about some feature of the site

9% Post-buy: Mentions checking order status, other account activities

7% Community: Mentions usage of some feature of community, like reviews

7% Price: Mentions getting information about price for specific items

2.9% Goldbox: Mentions Goldbox activity

2.2% Gift: Indicates looking for a gift for someone else

2.0% Sell: Mentions a selling activity

1.4% Personalization: Mentions personalization feature (e.g., recommendations)

0.8% Used: Mentions finding or transacting a used item

Stated vs Revealed Preferences

- Obtain insights by combining individual survey response with click analysis:
- Combine individual survey response with click analysis:

Look at those who ended up buying something:

- Only about one-half of those making a purchase indicated that they wanted to buy something in this visit

Look at those who said they wanted to buy something:

- Only about one-third of those indicating intent to buy ended up making a purchase in that visit
Customer Satisfaction

- Relate satisfaction rating to:
  - Visit intent (from survey)
  - Actions (from clickstream)
- Intermediate goal
  - Predict satisfaction
- Ultimate goal
  - Increase satisfaction

Number of responses: 6117
Dates: March 15–24, 2003

Flickr

Money Economy → Intention Economy → Attention Economy

From intention to attention

From intention to attention

© 2006 Weigend Associates LLC
Information at www.weigend.com
Discovery

- Tags are distilled attention, a pure form of attention.
- You are what you tag.
- You are what you are tagged as / who you are tagged by.
- Follow a tag and discover a topic
- Follow a user and discover what he is interested in
- Discover other users who have tagged the page

What is happening now?

- Empower millions of users to contribute
- del.icio.us: tag web pages
- flickr.com: tag photos
- 43things.com: tag your goals

Levels of Analysis and Actionability

- D. Network of customers
  - Model: Apply social network research
  - Action: Discounts, better service
- C. Customer
  - Model: Demographics, behavior
  - Action: Personalization-based marketing
- B. Visit
  - Model: Intention, situation, mode
  - Action: Session-based marketing
- A. Page
  - Model: Content
  - Action: Show ad

Network of Books

Based on Customers who Bought Also Bought

Who is checking me out?
**Platform: Amazon reviews**

- **Ratings**: When a customer writes a review, they are asked to rate the item by giving it between 1 and 5 stars. Amazon.com has more than 5 million such ratings.
- **Fields**:
  - Rater ID (obfuscated)
  - Date (when rating was submitted)
  - Item
  - Product ID (ASIN)
  - Our Price, List Price
  - Product Group
  - Rating of item
  - Helpfulness of review

**Distribution of ratings**

<table>
<thead>
<tr>
<th>Rating</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Star</td>
<td>7%</td>
</tr>
<tr>
<td>2 Star</td>
<td>5%</td>
</tr>
<tr>
<td>3 Star</td>
<td>8%</td>
</tr>
<tr>
<td>4 Star</td>
<td>60%</td>
</tr>
<tr>
<td>5 Star</td>
<td>20%</td>
</tr>
</tbody>
</table>

**Guess?**

- What is the Shape of the Distribution of Number of Stars?

**Make customer feedback trivially easy**

- Capture context automatically
- Suggestion Box
  - Your comments can help make our site better for everyone. If you've found something incorrect, broken, or frustrating on this page, let us know so that we can improve it. Please note that we are unable to respond directly to suggestions made via this form.
  - If you need help with an order, please contact Customer Service.
  - Please mark any of the following boxes that apply:
    - [ ] Product information is missing important details
    - [ ] Product information is incorrect
    - [ ] The page contains typographical errors
    - [ ] The page takes too long to load
    - [ ] The page has a software bug in it
    - [ ] Content violates Amazon.com's policy on offensive language
    - [ ] Product offered violates Amazon.com's policy on items that can be listed for sale

**Comments or Examples:**

E.g., missing information such as dimensions and model number, typing errors, etc. I went to the bathroom and came back, and the page was still loading!!
Platforms

- Enabling others
  - Example: Amazon’s “Mechanical Turk” (www.mturk.com)
  - Application: CastingWords.com

Increasing communication: 4 levels

- 1. Data Analysis
  - Data mining: Description, prediction

- 2. Architectures of Experimentation
  - A/B test, active learning, survey design...

- 3. Architectures of Participation
  - Remember, share, discover
  - Empower and incentivize people to contribute
  - Self-expression

- 4. Architectures of Interaction

The Attention Economy: Universal sets the exchange rate: 90 secs = 99 cents

To learn more...

www.web2.0-kongress.de

- 10th - 12th October 2006
- Offenbach
- Talk to me or Sabine Schuetze here!

To contact me:

- Andreas S. Weigend, Ph.D.
  - aweigend@stanford.edu
  - DE: +49 179 660 0 660
  - US: +1 (650) 906-5906
  - www.weigend.com