People and Data: Understanding Customer Behavior

Andreas S. Weigend, Ph.D.
former chief scientist, amazon.com

weigend.com

Outline

• Making decisions based on experiments (A-B tests)
  以A-B实验为基础进行决策
  ▶ Three ingredients for innovation
    革新的三大因素
  ▶ Revealed vs stated preferences
    被揭示 vs 被说出的偏好

• The iterative process of modeling
  反复的建模过程
  ▶ Define 定义 → Measure 度量 → Describe 描述 → Predict 预测 → Act 行动

• Some insights into online customer behavior
  一些对在线客户行为的理解
  ▶ Levels of analysis and actionability
    ▶ Personalization vs Occasionalization
    ▶ Behavioral economics
Result: Right vs Left

- **Metrics**
  - **Conversion**: Percentage of visits placing an order
    - 变化：下订单者增加的比例
  - **Order size**: Additional items (from the second page) put in cart
    - 订单规模：（第2页开始）更多附加物品被放入购物车

- **Some details**
  - **Relative increase**: Blue band on right compared with blue band on left
    - 相对增加：左右两边的蓝条相比

<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>Cart-adds from 2nd page:</td>
</tr>
<tr>
<td>Wishlist-adds:</td>
<td>+1.4%</td>
<td>DVD Cart-adds:</td>
</tr>
<tr>
<td></td>
<td>+1.1%</td>
<td>DVD ($)</td>
</tr>
</tbody>
</table>

**Research and science:**
- **Human decision making**: 人们做出决策
  - Finance
    - 金融
  - E-Business
    - 电子商务
  - Dating
    - 约会
- **Statistics**
- **Machine learning**
- **Behavioral economics**
- **Computational marketing**
Why Now? 为什么是现在？

- Data collected implicitly: Dramatic growth over time
  隐式地采集数据：时刻保持急剧增长
- Data collected explicitly / tacitly: Constant over time
  显式地采集数据/无声的：时刻持续进行

Three Ingredients For Innovation 革新的三大因素

Data 数据
- Computer Science 计算机科学
- Database Research 数据库研究

Methodology 方法论
- Statistics 统计学
- Data Analysis 探索数据分析
- Machine Learning 机器学习

Domain expertise 知识领域
- Behavioral Sciences 行为科学
- Marketing 市场营销学
- Finance 金融学
Research Questions

- Characterize paths through website

- Understand and influence conversion

- Predict intention and modality of the visit

- Compute and apply customer network value
  - “Mining the Network Value of Customers” by Pedro Domingos and Matt Richardson, KDD-2001. ACM Press.

Example: Invest $10M to Improve Customer Satisfaction

- Base decision on analysis of behavioral data
  - Quantify
  - Model
  - Act

- Consider
  - Increase selection?
  - Increase availability?
  - Reduce clutter on web site?
  - Improve product search algorithms?

“So, as you can see, customer satisfaction is up considerably since phasing out the complaint forms.”

© 2004 Weigend Associates LLC | www.weigend.com
• Make decisions based on experiments (A-B tests)
  以A-B实验为基础进行决策
  • The need for a scientific framework
    我们需要一个科学的框架
  • Three ingredients for innovation
    革新的三大因素
  • Revealed vs stated preferences
    被揭示 vs 被说出的偏好

• The iterative process of modeling
  反复的建模过程

• Some insights into online customer behavior
  一些对在线客户行为的理解
  • Levels of analysis and actionability
    • Personalization vs Occasionalization
  • Behavioral economics
1. Define Objectives 定义目标

- Stock price 股票价格
- Profit 收益
- Number of items sold 销售数量
- Number of visits 访问量
- Rate of conversion 转化率
- Customer acquisition 客户获得
- Customer retention 客户保持
- Customer satisfaction 客户满意度
2. Measure 测量

- Orders 订单
- Overall use of the site 网站的综合利用
  - Buying vs selling 买者 vs 卖者
  - Searching vs browsing 搜索 vs 浏览
  - Writing reviews, lists, etc. 写评论, 列表等
- Customer service contacts 客户服务联系
  - E-mail, phone 电子邮件, 电话
- Surveys 调研
  - Intentions / Goals / Modalities 意图 / 目标 / 形式
  - Satisfaction 满意度
- Customer service response 客户服务回复
  - Resolution 结果
    - (Free replacement, Refund 免费更换, 退款)
- Delivery date vs promised date 交货日期 vs 承诺日期
- Page generation time 页面产生时间
- Search response 搜索回复
  - Number of search results 搜索结果数量
- E-mail campaigns and responses 电子邮件广告和回复

Amount of Data Created Per Day 每天产生的数据量

- Level 水平
  - New data per day 每日新数据
    - Comparison 比较
      - 1 mm
      - 10 cm
      - 1 m
      - 1 km
      - 10 km
      - 10 + km
  - 1 MB
  - 10... 100 MB
  - 1... 10 GB
  - 100 GB... 1 TB
  - 10 + TB

*What was displayed, whether or not it was clicked on 无论点击与否，都会呈现
4. Building and Evaluating Predictive Models 预测模型的建立与评估

- **Tasks**: Predict, e.g., 目标：预测
  - Probability (buy in this visit without discount) vs Prob (buy in this visit with discount) 本次访问用优惠券购物与不用优惠券购物
  - Probability (current page is last page requested in this visit) 该页是本次访问的最后一页。

- **Use models from different model classes (different statistical assumptions)** 利用不同类别的模型（不同的统计假设）
  - Baseline, e.g., Poisson (independent, unconditional) 基线，例如，泊松（独立的，无条件的）
  - First order Markov 一阶马尔可夫过程
  - Beginning-of-visit information 访问开始的相关信息
    - HTTP-referrer HTTP转发
    - Search vs Browse 搜索与浏览
  - Aggregate visit so far (but time ordering ignored) 访问集合（但忽略定期订购）
Building More Complex Probabilistic Models  创建更复杂的概率模型

Joint work with Bruce D’Ambrosio, CleverSet Inc.  与Bruce D’Ambrosio, CleverSet Inc合作

- Add synthetic variables  增加综合变量
  - Combine observed variables (automatically generated)  与观察变量结合(自动产生)

- Add hidden variables  增加隐藏变量
  - Unobserved / hidden states  不可观察/隐藏状态

- Add relational structure  增加相关结构
  - E.g., use information from the products table, rather than only product ID  如产品表而非仅仅产品标识

- Evaluate out-of-sample accuracy  评估脱离例子分析的准确性
  - Standard: Area under ROC curve  ROC曲线下面的区域

Extended Relational Structure  扩展的关系结构

<table>
<thead>
<tr>
<th>Model</th>
<th>Performance (Area under ROC curve)</th>
</tr>
</thead>
<tbody>
<tr>
<td>First order Markov</td>
<td>.334</td>
</tr>
<tr>
<td>Hidden Markov</td>
<td>.513</td>
</tr>
<tr>
<td>Basic relational</td>
<td>.728</td>
</tr>
<tr>
<td>Extended relational</td>
<td>.777</td>
</tr>
</tbody>
</table>
Outline  

- Making decisions based on A-B experiments
- The need for a scientific framework

The iterative process of modeling

- Define 定义 → Measure 度量 → Describe 描述 → Predict 预测 → Act 行动

Some insights into online customer behavior

- Levels of analysis and actionability
  - Personalization vs Occasionalization
  - Behavioral economics

Insights 理解

- What works best for recommendations, and why?
  什么使推荐更有效，为什么？
  - Key distinction: individual-history free (current click only, same for everybody), vs history-dependent
    主要区别：个体历史无关 vs 历史相关

- Spectrum of information, generating different recommendations and actions
  不同类别的信息产生不同的建议和行动
  - Demographics (traditional marketing)
    人口统计 (传统市场营销)
  - Behavior (traditional collaborative filtering, customer signatures)
    行为 (传统的协同过滤，客户签名)
  - Current intentions and modalities (based only on information of current visit)
    目前的动机和形态 (仅基于本次访问的信息)
Using Social Networks for Customer Acquisition

- Social Networks
  - Research
    - STL proximity vs geographic/ZIP proximity (Mark Handcock)
    - Customer Lifetime Value: Intrinsic + Network (Pedro Domingos)
  - Applications
    - Customer acquisition
    - Compute Customer Network Value

- Data: “Share-the-Love”
  - From the Website: "Each time you place an order for books, music, DVDs, or videos with us, we'll offer you the chance to e-mail your friends and give them an additional 10% off the items you bought. (You select which items, of course.)" “If any of those people purchases one of those items within a week, you'll receive a credit to use the next time you shop with us!” “Your credit will equal the dollar amount of your friend's 10% discount.”

© 2004 by Weigend Associates LLC | www.weigend.com
Social Network of Books
based on Customers who Bought Also Bought (from Valdis Krebs)

Behavioral Economics

- Measure systematic deviations (heuristics, bounded rationality) in actual behavior in the real world
- vs traditional, neo-classical economics = 1st order / normative, theoretical

Hypotheses
- Choice set (Itamar Simonson)
- Sweepstakes (Dan Ariely)

<table>
<thead>
<tr>
<th>Choice:</th>
<th>Limited</th>
<th>Extended</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of jams</td>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>Customers stopping by</td>
<td>40%</td>
<td>60%</td>
</tr>
<tr>
<td>Conversion (buy rate)</td>
<td>30%</td>
<td>3%</td>
</tr>
</tbody>
</table>

Jam experiment

When Choice is Demotivating: Can One Desire Too Much of a Good Thing?
Contact Information

- Contact: Dr. Andreas S. Weigend
- Phone: +1 (917) 697-3800
- Web: www.weigend.com