Optimizing Web Sites for Customer Retention

("Analyzing Repeat Usage of Web-based Information")

2005 International Workshop on Customer Relationship Management:
Data Mining Meets Marketing
November 18th & 19th, 2005 - New York City - USA

Michael Hahsler
Department of Information Systems and Operations
Vienna University of Economics and Business Administration

Motivation

- Company Web sites are important channels which can be used as part of a CRM strategy to attract and retain customers.

- Optimizing Web sites (e.g., in terms of customer retention) requires analyzing how customers browse through the Web site and what information within the site they repeatedly view:

  - Data mining/Web mining techniques
  - **Existing models developed in marketing research for traditional channels**

- In this paper we propose an application of **NBD-type models for repeat buying/usage of Web-based information.**

Overview

1. Repeat buying theory and Web-based information
2. The basic NBD model
3. Develop the LSD/OTB model
4. Application and results

Repeat Buying Theory: The NBD Model

- Chatfield, Ehrenberg and Goodhardt (1966) introduced a simple but very successful model of stationary purchasing behavior, the so-called negative binomial distribution (NBD) model.

- Ehrenberg (1988) used the model to describe repeat-buying behavior for consumer products.

  Model assumptions and simplifications:
  - Only one brand (product) is analyzed at a time.
  - The aggregated purchasing behavior is stable (stationary) during the analyzed period of time.
  - The model uses the concepts of purchasing occasions and purchasing incidents.

- The model was extensively tested for **consumer products** using **panel data.**

Differences between using consumer panels and Web usage data

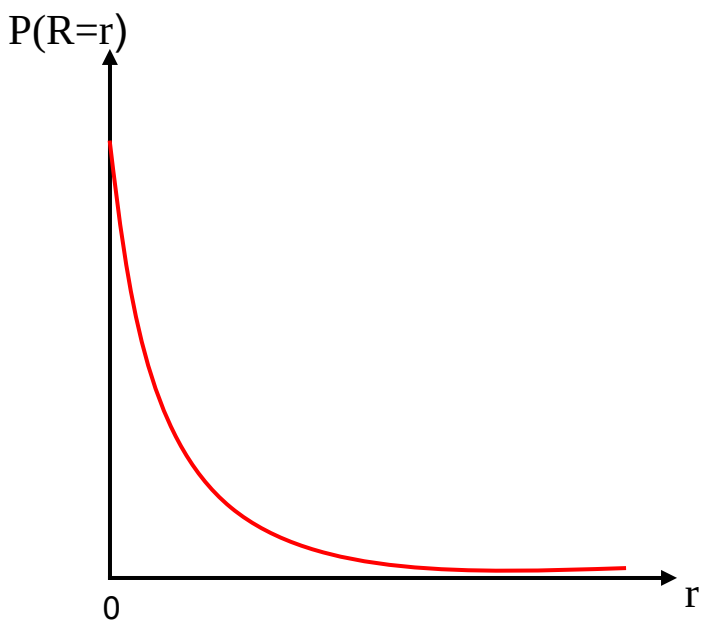
	Consumer Panel	Web Usage Data
Items	consumer products	information goods (Web sites, collections of Web pages, dynamic content, multimedia content, etc.)
Unit of Analysis	market baskets	browser sessions
Analyzed Behavior	purchasing incidents	following links to Web sites or documents, downloads, etc.
Ignored Behavior	quantity bought, package size	repeat usage per session, number of pages browsed within an information good, time spent on a page
Identity of Customers	known	varies between anonymous user sessions and complete personalization
Purchase History	known	unknown or known
Non-buyers	known	unknown
Incorrect Recording	omission, over-reporting	dynamic IP addresses, caching, proxy servers, Web robots

The Basic NBD Model

1. One product is analyzed for a specified time interval t.
2. Purchases for each individual customer c_i are modeled by a homogeneous Poisson Process with parameter λ_i
3. Customer heterogeneity is accounted for by different parameters λ_i; the distribution of the long-run λ_i over all customers is assumed to follow a truncated Γ-distribution.

The distribution of the number of customers with r purchases is a special case of a Poisson mixture distribution, namely the negative binomial distribution (NBD).

The 2 parameters of the NBD can be estimated for empirical data using the method of moments, maximum likelihood estimation and some other estimation methods (Tripathi, 1985). Parameter estimation is cumbersome if zero class at $r=0$ cannot be observed.



A Simplified Model: LSD Model

- A zero-truncated NBD tends toward the **logarithmic series distribution (LSD)** with only one parameter (q), if the proportion of customers who "buy" the item in the population is less then 20%. This requirement is easily met for most parts of a large Web site.

$$P_{LSD}(R = r) = \frac{-q^r}{r \ln(1-q)} \; for \; r = 1, 2, ...$$

- The average number of purchases per customer ω is a function of q. q can be efficiently obtained from observed ω by standard numeric approximation.

$$\omega = \frac{-q}{(1-q)ln(1-q)}$$

The LSD/OTB Model

- Real-world data sets contain a **much higher proportion of observations at *r=1*** than expected.

- Since the model only describes repeat-usage behavior, the discrepancy at $r=1$ can be explained by users who only use a certain information product once and then never again. Fader and Hardie (2002) incorporated such “one-time buyers” (OTB) into the NBD model.

- In the same way we can expand the LSD model to account for **one-time users**. π is the proportion of one-time buyers and $\delta_{r,1}$ is the Kronecker delta with $\delta_{r,1}=1$ if $r=1$ and 0 otherwise.

$$P_{LSD/OTB}(R=r) = \pi\delta_{r,1} + (1-\pi)P_{LSD}(R=r) \;\; for \;\; r = 1, 2, \ldots$$

The LSD/OTB Model (cont'd)

- The parameters of the LSD/OTB model can be estimated using the method of maximum likelihood. The log-likelihood function is given by:

$$LL(q,\pi) = \sum_{r=1}^{max(r)} f_r ln(P_{LSD/OTB}(R = r|q,\pi))$$

f_r represents the observed number of users with r repeat-buys.

Application

- We use one year (2001) of transaction data from the application level log of an educational portal, the Virtual University information system at the Vienna University of Economics and Business Administration (see http://vu.wu-wien.ac.at)

- An item is defined as a collection of Web pages (including also other media like word processor files) that present a logical unit of information on a specific topic, e.g., "*Course material on International Marketing*."

- The system requires no login. We use heuristics to identify different users (filter Web robots, analyze IP addresses and use permanent cookies stored in the users' Web browsers)

Results

In the data set we observed usage of 8340 different items from 31214 different IP addresses. The data contained 94032 sessions. We selected items with more than 10 observations. This reduced the items to 1627. For these items we estimated the parameters of the LSD and the LSD/OTB model and used the Chi-Square Goodness-of-Fit-Test to evaluate the fit of the model.

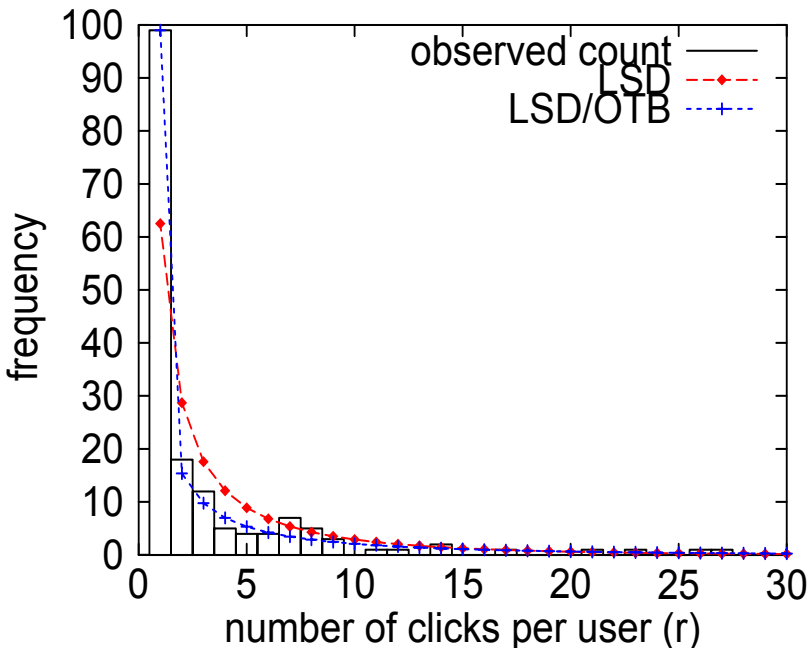
model	items	No q	no χ2 test	no significant* differences	Significant* differences	% of fitting models
LSD	1627	452	962	57	156	**26.68%**
LSD/OTB	1627	1008	398	**184**	37	**83.26%**

** α=0.05*

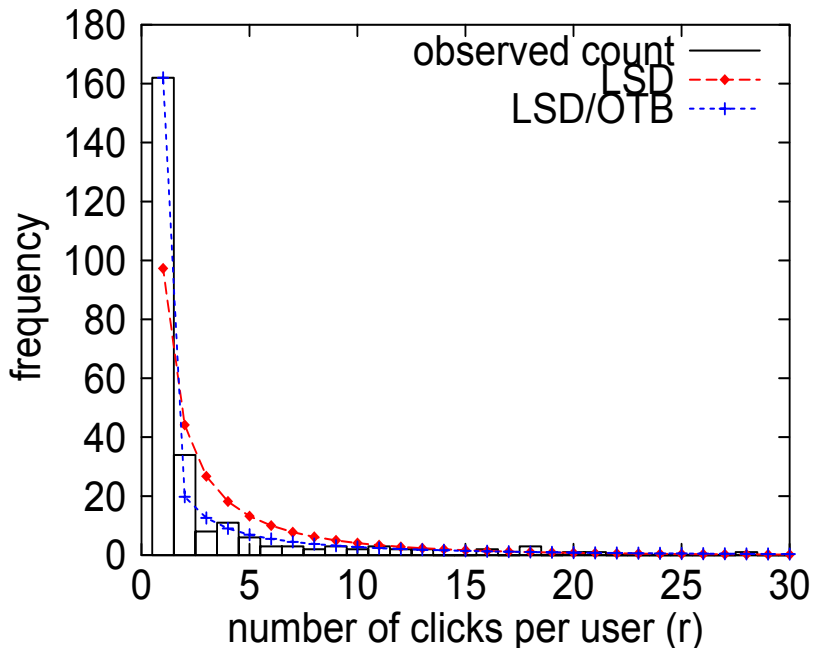
Results (cont'd)

Two example items (for better visibility of the differences both plots are truncated at *r=30)*.

C++ Standard Library



Freshmeat



Results (cont'd)

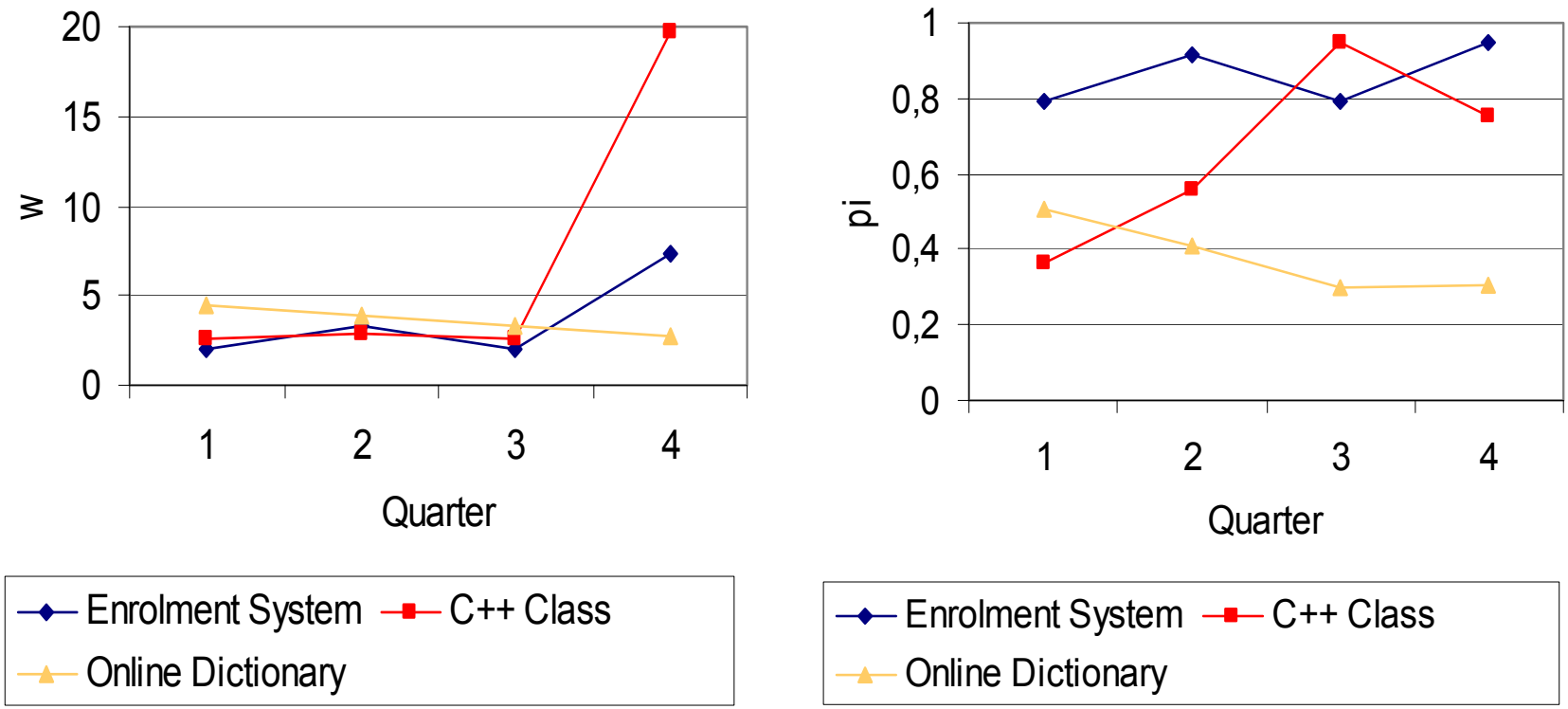
The parameters of the model provide important diagnostics for the usage of different items.

- ***ω:*** The average number of repeat-usages per user. A measure for the intensity of use.

- ***π:*** A proportion of one-time users which is considerably higher for some items than for the rest in a Web site indicates that the items do not provide the customers with the expected information or are not **fulfilling their expectations,** thus do not contribute to the aim of retaining customers.

Results (cont'd)

By considering more than one time-period with the LSD/OTB model, the development of repeat-usage behavior can be analyzed



Conclusion

- Customer retention is a major goal of CRM. For the online channel, we suggested analyzing repeat usage behavior of commercial Web sites.

- Instead of applying standard data mining technology, we applied and modified a NBD-type model from marketing research.
  - In order to be able to deal with large data volumes, we applied the LSD model, the most simple and easy to estimate version of the NBD-type models.
  - To account for one-time users (OTB), we expanded the model and developed the LSD/OTB model. Using an example data set, we show that the LSD/OTB model gives a good fit for the items.

- The parameters of the extended model are directly usable for the manager of a Web site for
  - finding problems with the item quality or the navigational structure and
  - monitoring changes of the users' repeat-usage behavior over time.

- This type of analysis can also be performed on a collection of usage data gathered from various channels (Web site, call center, written correspondence, …)

References

- C. Chatfield, A. S. C. Ehrenberg, and G .J. Goodhardt, “Progress on simplified model of stationary purchasing behaviour,” *Journal of the Royal Statistical Society, Series A*, 129(3):317-367, 1966.

- A. S. C. Ehrenberg, *Repeat-Buying: Facts, Theory and Applicatio*n. London: Charles Griffin & Company Ltd., 1988.

- R.C. Tripathi, Negative binomial distribution. In Samuel Kotz and Norman L. Johnson, editors, *Encyclopedia of Statistical Science*, volume 6, pages 169-177. John Wiley & Sons, New York, 1985.

- P. S. Fader and B.G.S. Hardie, “A note on an integrated model of customer buying behavior,” *European Journal of Operational Research*, 139(3):682-687, 2002.