

Others Also Use: A Robust Recommender System for Scientific Libraries

Andreas Geyer-Schulz, Andreas Neumann, Anke Thede

Information Services and Electronic Markets,
Institute for Information Engineering and Management,
Universität Karlsruhe (TH), D-76128 Karlsruhe, Germany
{geyer-schulz, neumann, thede}@em.uni-karlsruhe.de

7th European Conference on
Research and Advanced Technology for Digital Libraries
August 18, 2002, Trondheim, Norway

Table of Contents

- Introduction and Motivation
- An Architecture for Distributed Recommender Services
- Adapting a Stochastic Model from Repeat-Buying Theory
- The Recommender System at the Library of the Universität Karlsruhe (TH)
- Evaluation
- Further Research

Usage of Electronic Scientific Information in Academic Education

Klatt, R. et al. <http://www.stefi.de/> (2001).

Survey by order of the Federal Ministry of Education and Research.

- Students:
 - ★ Electronic literature research is very important: 76,4 %
 - ★ Their information expertise is fair to poor
 - ★ Problems:
 - * Supply complexity: 35,9 %
 - * Evaluation of the quality: 32,8 %
 - * Information overload: 26,4 %
 - ★ Typical literature research path: Asking classmates (60,2 %)

- Scientists:
 - ★ Usage of the Computer for literature research:
 - * E-mail consultation of students and colleagues: 91,4 %
 - * Free search via internet search engines: 66,4 %

Recommender Systems for Scientific Libraries

- Scientists and students are more and more incapable of efficiently finding relevant literature in conventional database oriented catalog systems.
 - More sophisticated access paths to prevent information overload are needed.
- Many universities are teaching a growing number of students with a more or less fixed staff size.
 - Students, university teachers and researchers could free some of their valuable time that is needed now to recommend each other e. g. standard literature of their fields which could be done easily by behavior-based expert advice services.
- Usage of the self-selection constraint to identify homogeneous library user segments

A Motivational Example I

Suchergebnis - Mozilla {Build ID: 2002060511}

Back Forward Reload Stop <http://www.ubka.uni-karlsruhe.c> Search Print

Universitätsbibliothek Karlsruhe

Suchergebnis

Katalog: UB Karlsruhe
Suchanfrage: find nd=10205141

[Neue Suche](#) [Trefferliste](#) [MAB Format](#) [Sitzung Ende](#)

Lazinger, Susan S.
Digital preservation and metadata : history, theory, practice / Susan S. Lazinger. With an annotated list of electronic social science data archives and cultural heritage digitization projects and centers by Helen R. Tibbo. – Englewood, Colo. : Libraries Unlimited, 2001. – XXII, 359 S.; (engl.) ISBN 1-563-08777-4
Includes bibliographical references and index

UB Karlsruhe:
Signatur: 2002 A 183
Notation: buch 4.2

- ◆ [Band bestellen](#)
- ◆ [Verwandte Literatur](#)
 - ◇ [1. Autor](#)
 - ◇ [Notation](#)
 - ◇ [Empfehlungen Neu \(Others also use...\)](#)
- ◆ [Mehr zu diesem Titel](#)

Document: Done (1.02 secs)

Detailed View of Documents (Books, Journals, Multi-Media,...)

A Motivational Example II

UB Karlsruhe - Empfehlungen - Mozilla {Build ID: 2002060511}!

http://ubrec.em.uni-karlsruhe.de/cgi-bin/get_recom

Universitätsbibliothek Karlsruhe

Empfehlungen für

Digital preservation and metadata / Lazinger, Susan S. (2001)

Dokument: nd=10205141 ([link](#))
Katalog: UB Karlsruhe

Dieser Service zeigt eine Liste von Dokumenten, die andere Benutzer zusammen mit dem obigen Dokument benutzt haben. Die gezeigten Dokumente können also zur Benutzung gemeinsam mit dem gewählten empfohlen werden ("Others also use..."). Die Liste ist nach der Güte der Empfehlungen sortiert (Anzahl der gemeinsamen Benutzungen in Klammern).

Neue Suche Ich finde den Empfehlungsdienst 👍 super 👉 gut 👎 benutzbar 👏 verbesserungsbedürftig 👎 sinnlos
allgemein

1. [Proceedings / Schmidt, Ralph \(2002\)](#) (18)
2. [Bibliotheks-Management / Paul, Gerd \(2000\)](#) (16)
3. [Encyclopedia of library and information science, 72](#) (12)
4. [Der Börsenverein des Deutschen Buchhandels 1825 – 2000 / Füssel, Stephan \(2000\)](#) (12)
5. [Encyclopedia of library and information science, 71](#) (12)
6. [Zeitschrift für Bibliothekswesen und Bibliographie / Sonderhefte / hrsg. von Jürgen Hering / Tröger, Beate \(2000\)](#) (12)
7. [Politik für Bibliotheken / Ruppelt, Georg \(2000\)](#) (11)
8. [Bibliothek in der Wissensgesellschaft / Blum, Askan \(2001\)](#) (11)
9. [World guide to libraries](#) (6)
10. [Bibliotheken führen und entwickeln / Bürger, Thomas \(2002\)](#) (6)
11. [World guide to libraries](#) (6)

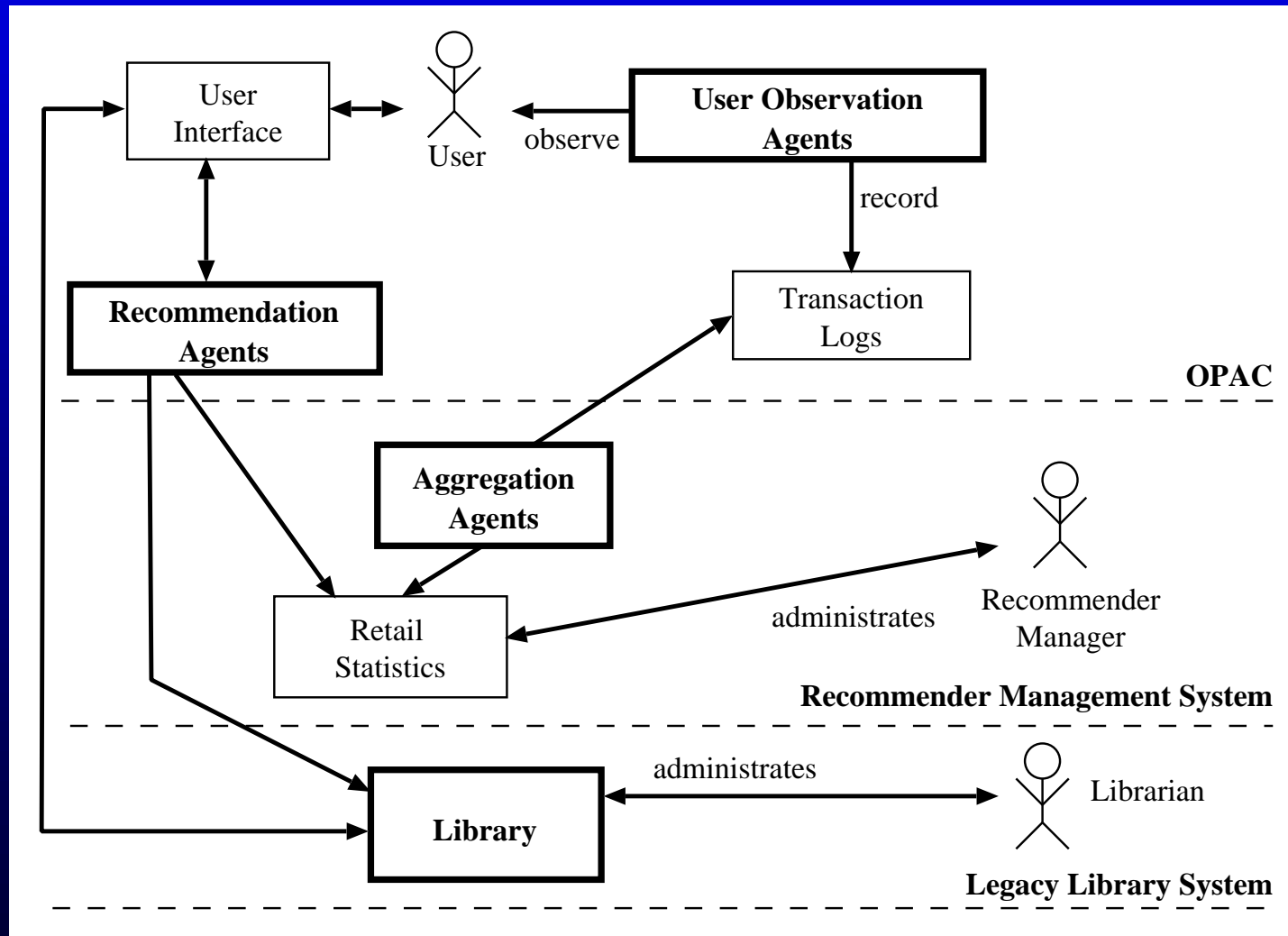
developed by
Schroff-Stiftungslehrstuhl für
Informationsdienste und elektronische Märkte

Gefördert von der
Deutschen Forschungsgemeinschaft

Document: Done (0.776 secs)

List of Recommendations

The Architecture of the Recommender System I

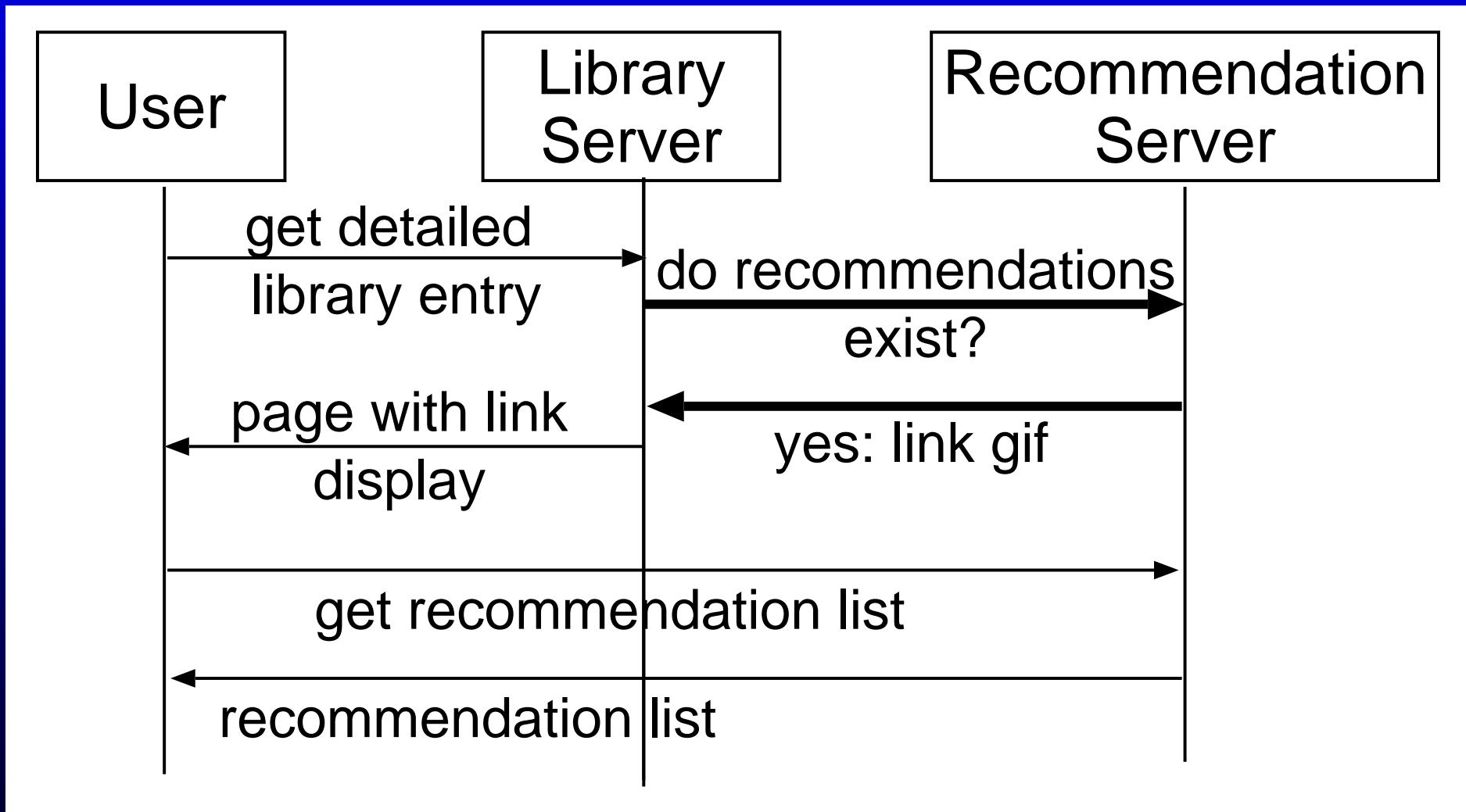


An Architecture for Distributed Recommender Services for Digital and Hybrid Library Systems

The Architecture of the Recommender System II

- User Observation Agent:
 - ★ http-logs with link embedded session IDs
 - ★ preprocessing:
 - * extraction of GET requests
 - * session splitting to take care of public access terminals
- Aggregation Agent:
 - ★ information market: selecting information = purchase
 - ★ computes market-baskets
 - ★ estimates a logarithmic series distribution (LSD)
 - ★ identifies and extracts outliers as recommendations
 - ★ performs incremental updates periodically
- Recommendation Agent:
 - ★ CGI-script on the recommendation server
 - ★ generates recommendation pages
 - ★ accessed via embedded links in the OPAC

The Architecture of the Recommender System III

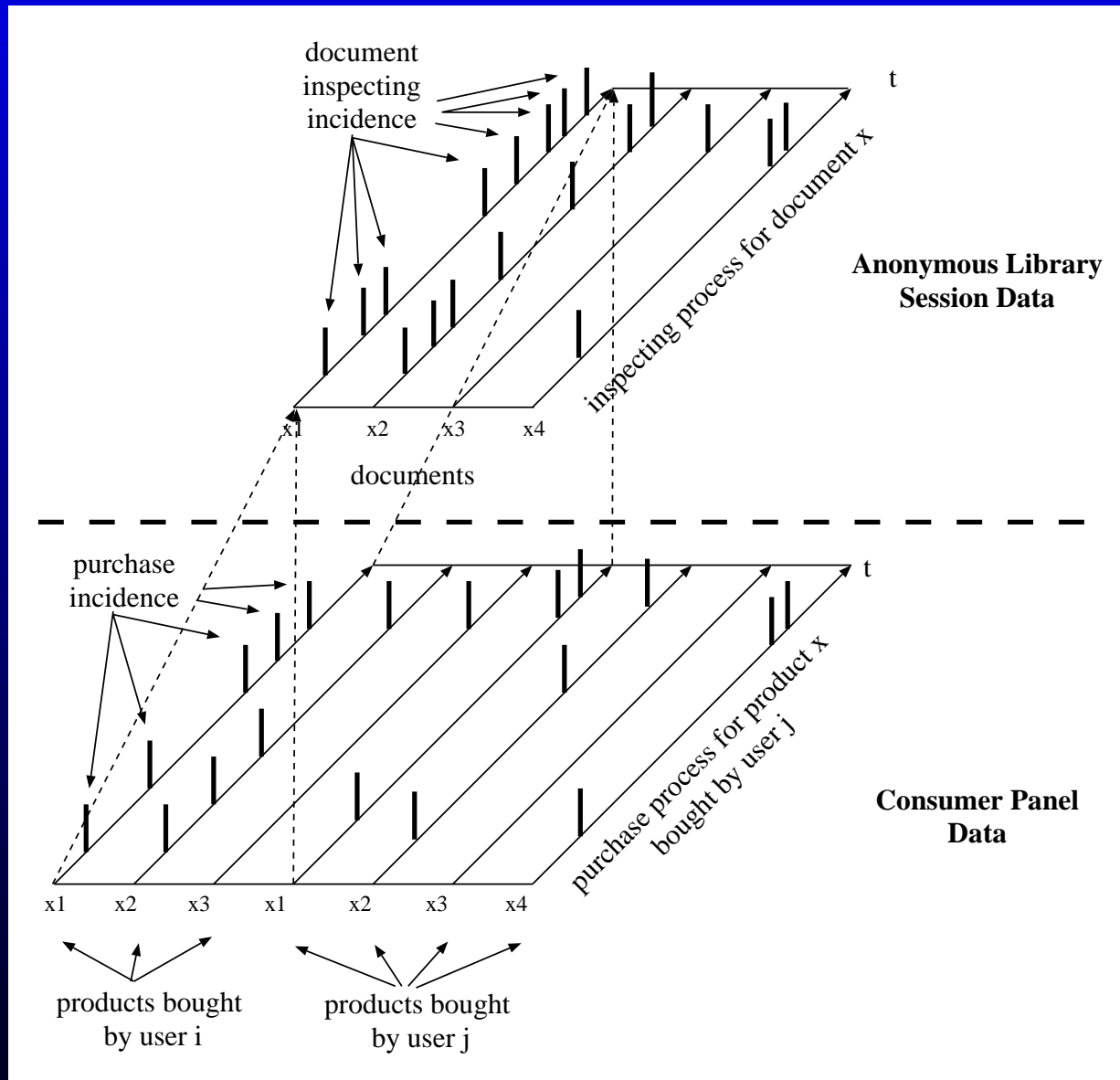


Message Trace for Detailed View and Recommendation Request

Stochastic Purchase Incidence Model

- Based on Ehrenberg's Repeat-Buying Theory (1988), descriptive theory of consumer behavior
- Anonymous customers (patrons)
- Analysis of market baskets (user sessions)
- Purchase (detailed document inspection) frequency of product pairs → Frequency distribution follows a logarithmic series distribution (LSD)
- Calculate observed and the expected LSD distribution
- Apply a χ^2 -goodness-of-fit test between observed and expected
- Determine the outliers in the tail (pairs with high repeat buys) → Recommendations

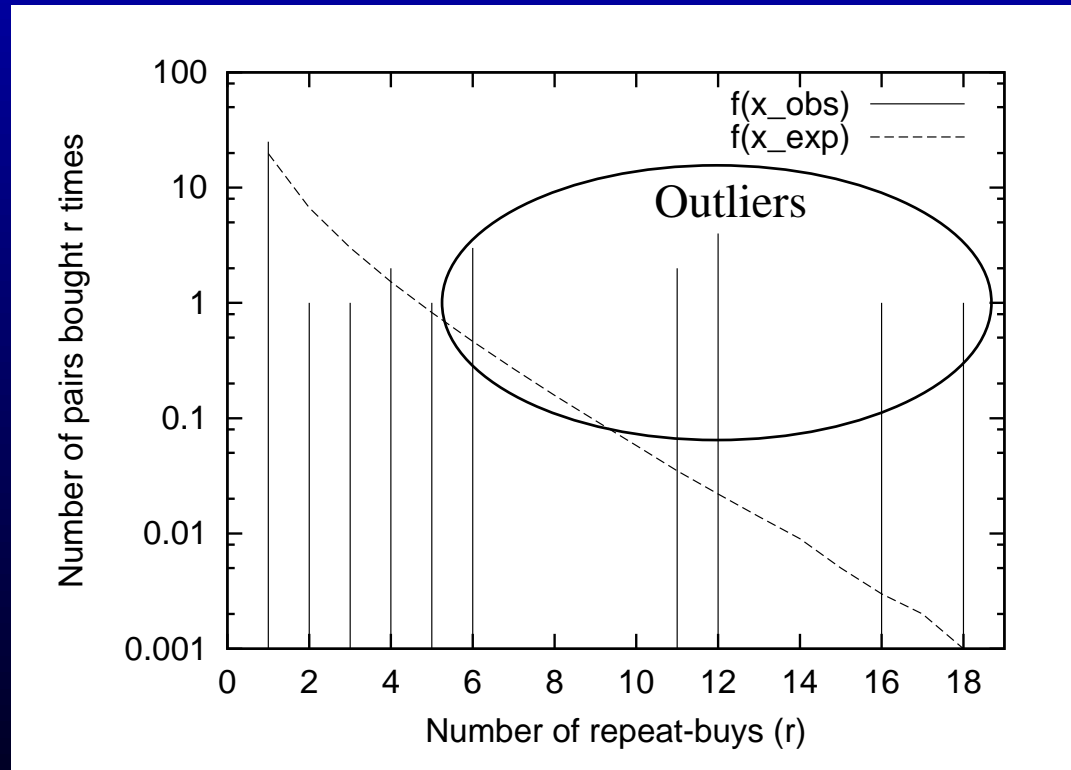
Stochastic Processes



From Consumer Panels to Anonymous Library Session Data

Generating Recommendation

- Recommendations:
 - ★ Products bought together more often than expected by stochastic model
 - ★ Violating the independence assumptions



Plot of Frequency Distribution (Digital preservation and metadata by Lazinger)

Method Features

- Incremental Update
 - ★ $O(n^2)$ in time and space, n number of items updated
 - ★ Improves scalability
 - ★ Memory of the recommender is kept
- Comparison with Association Rules (AR)
 - ★ Lower complexity than association rule algorithm
 - ★ With AR choice of minimum support and confidence is critical and not statistically well founded
 - ★ AR performance need not be stable over time (chosen support and confidence needs to be checked regularly)
 - ★ Computation of the LSD-distribution independently for each document, with AR support and confidence depend on the complete data set

Implementation

- Library of the Universität Karlsruhe (TH)
 - ★ Internet/Catalog Service Provider for many libraries within the south-west of Germany
 - ★ Hosting 23 different catalogs
 - ★ 15 million books - combined library catalog
 - ★ Recommendations are crossing catalog boundaries
- Recommendation Server
 - ★ 1,2 GHz AMD Athlon Processor
 - ★ 1,5 GB Main Memory
 - ★ Mandrake Linux (Kernel Version 2.4.17)
 - ★ Recommender Software: Perl
 - ★ MySQL Database for Author, Title, etc.: (Catalog, Document ID) Key

Results for Observation Period 2001-01-01 to 2003-03-13

	I q undef.	II no χ^2 (< 3 classes)	III Sign. $\alpha = 0.05$	IV Sign. $\alpha = 0.01$	V Not sign.	Σ
A Obs. < 10	510.656 (0)	79.175 (14.925)	0 (0)	0 (0)	0 (0)	589.831 (14.925)
B $\bar{x} = 1$	183.456 (0)	0 (0)	0 (0)	0 (0)	0 (0)	183.456 (0)
C $\bar{x} > \sigma^2$ $r \leq 3$	996 (0)	143.441 (34.886)	739 (603)	4.375 (2.202)	3.599 (1.627)	153.132 (39.318)
D $\bar{x} > \sigma^2$ $r > 3$	0 (0)	47.234 (47.121)	4.401 (4.401)	6.809 (6.776)	10.272 (9.590)	68.716 (67.888)
E $\sigma^2 > \bar{x}$	0 (0)	28.643 (28.643)	11.436 (11.436)	9.810 (9.810)	15.072 (15.072)	64.961 (64.961)
Σ	695.108 (0)	298.493 (125.575)	16.576 (16.440)	20.976 (18.788)	28.943 (26.289)	1.060.096 (187.092)

(n) denotes n lists of recommendations. Total number of recommendations: 1.983.908

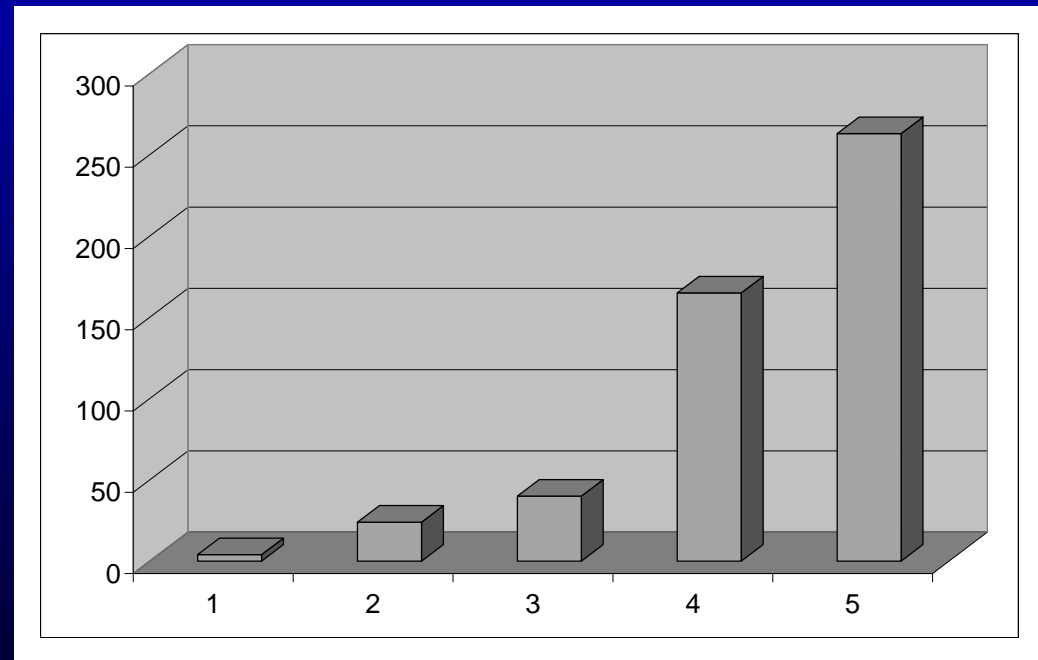
Evaluation I

- Total figures:
 - ★ 15 million documents
 - ★ 1,060,096 have been inspected by users together in a session with at least one other document (2001-01-01 – 2003-03-13)
 - ★ Recommendation lists for 187,092 documents
 - ★ 1,983,908 recommendations
 - ★ Average usage of recommendations: 118.83 times per day (increasing trend)
- Coverage:
 - ★ Coverage of total documents: 1.25 %
 - ★ Coverage of inspected documents: 38.18 %
 - ★ Coverage of detailed inspections (2003-06-18): **We can provide recommendations in 46.21 % of all document inspections**

Evaluation II

- Question to users:
 - ★ I consider the recommendation service in general as:
 - ★ (1) dispensable – (2) needs improvement – (3) usable – (4) good – (5) super

- Result:
 - ★ 496 votes
 - ★ mean: 4.33
 - ★ Standard deviation: 0.88



Further Research

- Further evaluation through users
- Evaluation of recommendations through experts
- Including time information to analyse fashions, emerging trends, . . .
- Diffusion models
- Recommendations of higher order
- Usage of the system for library collection management

Others Also Use: A Robust Recommender System for Scientific Libraries

Recommender System available at:
www.ubka.uni-karlsruhe.de

Questions?

Andreas Geyer-Schulz, Andreas Neumann, Anke Thede

Information Services and Electronic Markets,
Institute for Information Engineering and Management,
Universität Karlsruhe (TH), D-76128 Karlsruhe, Germany
{geyer-schulz, neumann, thede}@em.uni-karlsruhe.de