

## Recommender-Systeme Teil 2

### Kollaboratives Filtern & inhaltsbasierte Empfehlungen

1

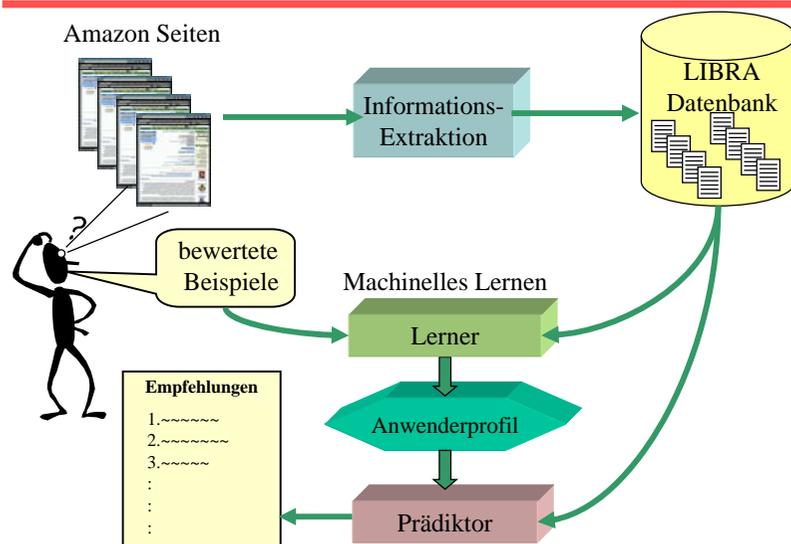
## LIBRA

### Learning Intelligent Book Recommending Agent

- Inhaltsbasierender Recommender für Bücher, der Informationen über Titel verwendet, die von Amazon extrahiert wurden.
- Verwendet Informations-Exktration aus dem Web, um Text in Feldern zu organisieren:
  - Autor
  - Titel
  - Redaktionelle Reviews
  - Stellungnahmen von Kunden
  - Themenbezeichnungen
  - Zugehörige Autoren
  - Zugehörige Titel

2

## LIBRA System



3

The screenshot shows the Amazon.com product page for the book "The Age of Spiritual Machines: When Computers Exceed Human Intelligence (Paperback)" by Ray Kurzweil. The page includes the book cover, a price of \$10.99 (a 20% discount from \$13.74), and options to buy it. Below the main product, there are sections for "Better Together" (related books), "Customers who bought this item also bought" (a list of other books by Kurzweil and related titles), and "Editorial Reviews" (a review snippet from Amazon.com).

## Bsp.: von Amazon extrahierte Informationen

---

Titel: <The Age of Spiritual Machines: When Computers Exceed Human Intelligence>  
Autor: <Ray Kurzweil>  
Preis: <11.96>  
Datum der Veröffentlichung: <Januar 2000>  
ISBN: <0140282025>  
zugehörige Titel: <Titel: <Robot: Mere Machine or Transcendent Mind>  
                  Autor: <Hans Moravec> >  
...  
Reviews: <Autor: <Amazon.com Reviews> Text: <How much do we humans...> >  
...  
Stellungnahmen: <Stars: <4> Autor: <Stephen A. Haines> Text:<Kurzweil has ...> >  
...  
zugehörige Autoren: <Hans P. Moravec> <K. Eric Drexler>...  
Betreffs: <Science/Mathematics> <Computers> <Artificial Intelligence> ...

5

## Libra: Inhaltsinformationen

---

- Libra verwendet diese Information, um die folgenden Slots mit “bags of words” zu füllen:
  - Autor
  - Titel
  - Beschreibung (Stellungnahmen und Kommentare)
  - Betreffs
  - verwandte Titel
  - verwandte Autoren

6

## Libra: Übersicht

---

- Anwender bewertet ausgewählte Titel auf einer Skala von 1 bis 10.
- Libra verwendet einen Naiven-Bayes-Text-Kategorisierungs-Algorithmus, um daraus ein Profil zu erlernen.
  - Bewertung 6–10: positiv
  - Bewertung 1–5: negativ
- Von den anderen Büchern werden diejenigen als Empfehlungen klassifiziert, bei denen die errechnete Wahrscheinlichkeit für eine positive Einschätzung groß genug ist.
- Der Anwender kann auch explizite positive/negative Schlüsselwörter liefern, die dann höher gewichtet werden.

7

## Bayesian Kategorisierung in LIBRA

---

- Das Model ist verallgemeinert, um einen **Vektor** von “bags of words” zu erzeugen (ein “bag” für jeden Slot).
  - Instanzen desselben Worts in verschiedenen Slots werden als separate Merkmale behandelt:
    - “Chrichton” bei Autor – “Chrichton” in der Beschreibung
- Trainingsbeispiele werden als positiv oder negativ *gewichtete* Beispiele bei der Schätzung der bedingten Wahrscheinlichkeitsparameter behandelt:
  - Gegeben ist ein Beispiel mit Bewertung  $1 \leq r \leq 10$ :
    - positive Wahrscheinlichkeit:  $(r - 1)/9$
    - negative Wahrscheinlichkeit:  $(10 - r)/9$

8

## Implementierung

---

- Stopwörter wurden von allen “bags of words” entfernt.
- Buchtitel und Autoren werden auch zu den Slots “verwandte Titel” bzw. “verwandte Autoren” hinzugefügt.
- Alle Wahrscheinlichkeiten werden wegen der kleinen Datenbasis mit der Laplace-Schätzung geglättet.
- Lisp-Implementierung ist ziemlich effizient:
  - Training: 20 Datensätze in 0.4 s, 840 in 11.5 s
  - Test: 200 Bücher pro Sekunde

9

## Erläuterung von Profilen und Empfehlungen

---

- Stärke des Auftretens von Wort  $w_k$  in Slot  $s_j$ :

$$\text{strength}(w_k, s_j) = \log \frac{P(w_k | \text{positive}, s_j)}{P(w_k | \text{negative}, s_j)}$$

10

## Experimentelle Daten

---

- Bücher aus verschiedenen Genres von Amazon.
- Titel, die zumindest eine Stellungnahme oder einen Kommentar haben, wurden ausgewählt.
- Datensätze:
  - Literaturfiktion: 3,061 titles
  - Mysterium: 7,285 titles
  - Wissenschaft: 3,813 titles
  - Science Fiction: 3,813 titles

11

## Bewertete Daten

---

- Vier Anwender bewerteten zufällige Beispiele innerhalb eines Genres, indem sie die Amazon-Seiten hinsichtlich der Titel durchsahen:
  - LIT1 936 titles
  - LIT2 935 titles
  - MYST 500 titles
  - SCI 500 titles
  - SF 500 titles

12

## Experimentelle Methode

- 10-fache Kreuz-Validierung, um Lernkurven zu erzeugen.
- Auf unabhängigen Testdaten wurde gemessen:
  - **Präzision der Top 3**: % der besten 3, die positiv sind
  - **Bewertung der Top 3**: Durchschnittsbewertung, die den besten 3 zugeordnet ist
  - **Klassifizierungs-Korrelation**: Spearman's,  $r_s$ , zwischen vollständigen Klassifizierungen von System und Anwender.
- Test ohne die Slots “verwandter Autor” und “verwandter Titel” (LIBRA-NR).
  - Testet den Einfluss von Informationen, die durch den kollaborativen Ansatz von Amazon erzeugt wurden.

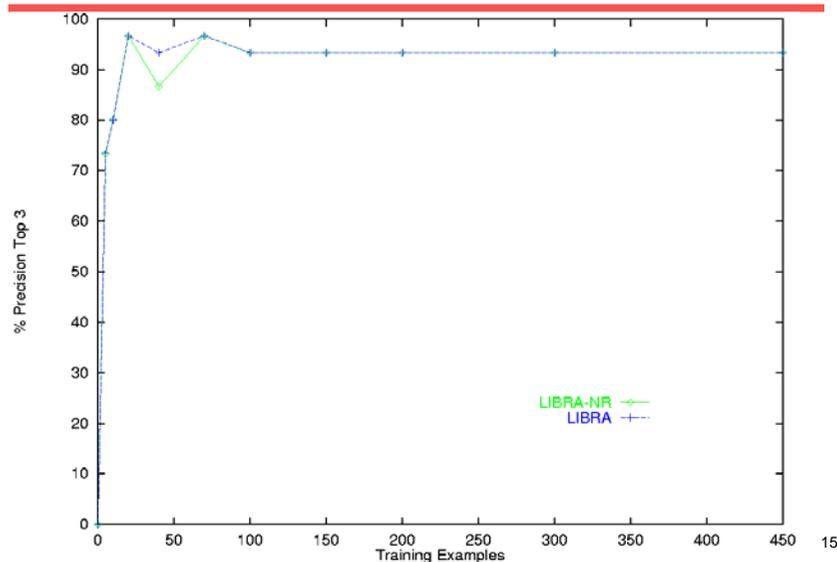
13

## Experimentelle Ergebnisse: Zusammenfassung

- **Präzision der Top 3** ist nach nur **20 Beispielen** mit um die **90 %** ziemlich konsistent.
- **Bewertung der Top 3** ist nach nur **20 Beispielen** mit über **8** ziemlich konsistent.
- Alle Ergebnisse sind nach nur **5 Beispielen** immer bedeutend besser als eine zufällige Auswahl.
- **Klassifizierungs-Korrelation** liegt nach nur **10 Beispielen** generell über **0.3 (gemäßigt)**.
- **Klassifizierungs-Korrelation** liegt nach nur **40 Beispielen** generell über **0.6 (hoch)**.

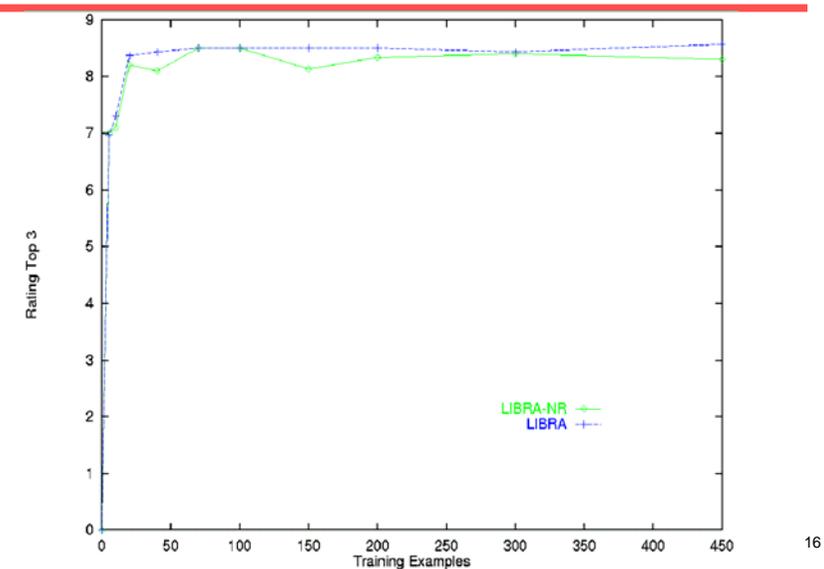
14

## Präzision der Top 3 für “Science”



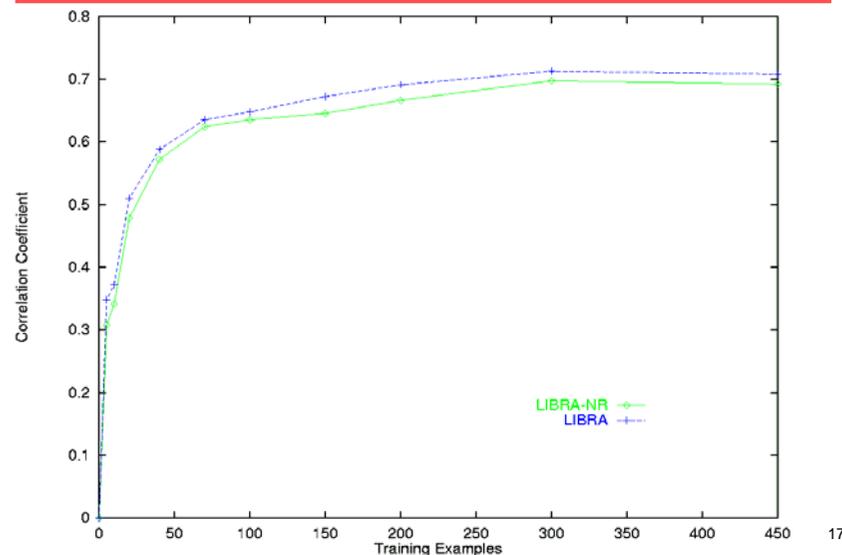
15

## Bewertung der Top 3 für “Science”



16

## Klassifizierungs-Korrelation für "Science"



17

## Anwenderstudie

### Nutzer

- wurden gebeten, Libra zu verwenden und Empfehlungen anzufragen.
- wurden zu mehreren Feedback-Runden ermutigt.
- bewerteten alle Bücher der endgültigen Empfehlungs-Liste.
- wählten zwei Bücher zum Kauf.
- schickten nach Lesen der Auswahl Stellungnahmen zurück.
- vervollständigten Fragebogen über das System.

18

## Verbinden von Inhalt und Kollaboration

- Inhaltsbasierte und kollaborative Methoden haben komplementäre Stärken und Schwächen.
- Kombiniere Methoden, um das Beste von beiden zu erhalten.
- Es gibt verschiedene hybride Ansätze:
  - Wende beide Methoden an und verknüpfe die Empfehlungen.
  - Verwende kollaborative Daten als Inhalt.
  - Verwende inhaltsbasierte Empfehlungen als weiteren Nutzer, dessen Verhalten in weitere Vorhersagen einfließt.
  - **Verwende einen inhaltsbasierten Empfehler, um kollaborative Daten zu vervollständigen.**

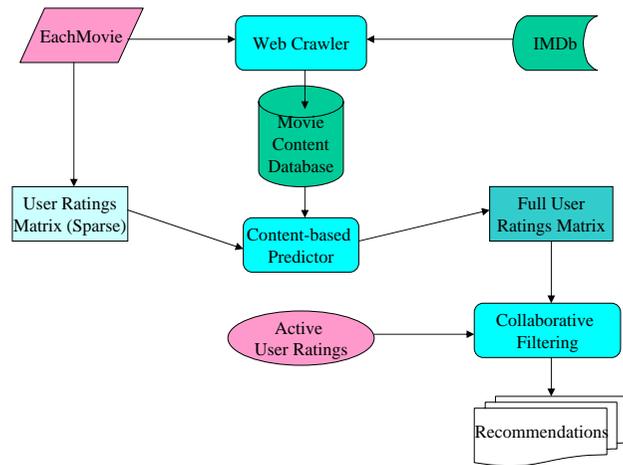
19

## Anwendung: Movies

- *EachMovie* Datensatz [Compaq Research Labs]
  - enthält Anwenderbewertungen für Filme auf einer Skala von 0 bis 5.
  - 72,916 Anwender (mit durchschn. 39 Bewertungen).
  - 1,628 Filme.
  - Spärliche besetzte Anwender-Bewertungsmatrix – (zu 2.6% besetzt).
- Internet Movie Database (*IMDb*)
  - Wurde für die Titel aus *EachMovie* gecrawlt.
- Wesentliche Filminformationen:
  - Titel, Direktor, Rollenbesetzung, Genre, etc.
- Populäre Meinungen:
  - Nutzerkommentare, Zeitungs- und Newsgroup-Berichte, etc.

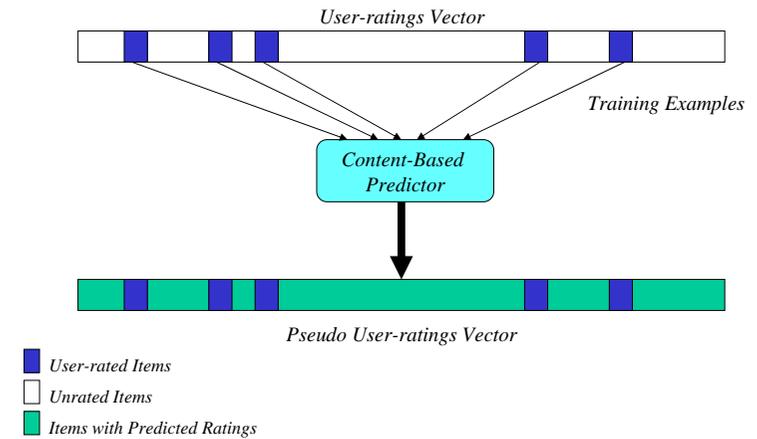
20

## Content-Boosted kollaboratives Filtern



21

## Content-Boosted CF - I



22

## Content-Boosted CF - II



- Berechne Pseudo-Anwenderbewertungsmatrix
  - Volle Matrix – approximiert die konkreten Nutzerbewertungen
- Führe CF aus
  - Unter Verwendung von Pearson-Korr.-Koeffizient zwischen den Pseudo-Anwender-Bewertungsvektoren

23

## Experimentelle Evaluierung

- Teilmenge von *EachMovie* wurde verwendet:
  - 7,893 Anwender; 299,997 Bewertungen
- Testmenge: 10% der Anwender wurden zufällig ausgewählt.
  - Sie bewerteten je mindestens 40 Filme.
  - Training auf der verbleibenden Menge.
- “Hold-out”-Menge: 25% der Objekte für jeden Testanwender.
  - Vorhersage der Bewertung für jedes Objekt in der “Hold-out”-Menge.
- Vergleich CBCF mit anderen Ansätzen:
  - Reines CF
  - Rein inhaltsbasiert
  - hybrider Naïve Bayes (Durchschnitts-CF und inhaltsbasierende Vorhersagen)

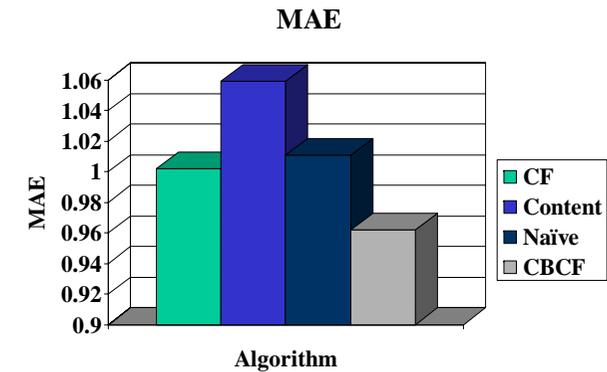
24

## Maße

- Mittlerer absoluter Fehler (MAE)
  - Vergleicht die Vorhersagen mit Anwenderbewertungen
- ROC Empfindlichkeit [Herlocker 99]
  - Wie gut helfen die Vorhersagen den Anwendern, eine gute Auswahl zu treffen?
  - Bewertungen  $\geq 4$  als “gut” betrachtet;  $< 4$  als “schlecht” betrachtet
- Gepaarte T-test stellen die statistische Signifikanz fest.

25

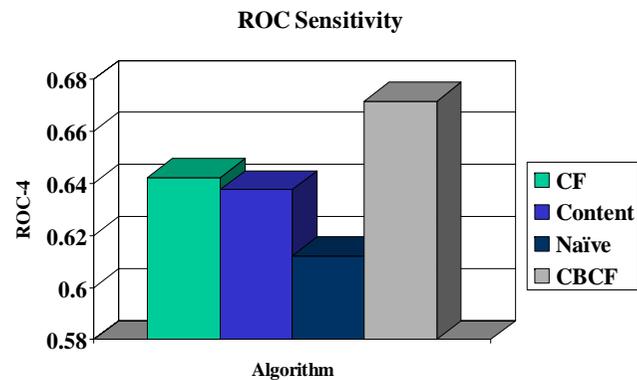
## Ergebnisse - I



*CBCF ist bedeutend besser (4% über CF) bei ( $p < 0.001$ )*

26

## Ergebnis - II



*CBCF übertrifft Rest (5% Verbesserung über CF)*

27

## Aktives Lernen

- Wird verwendet, um die Anzahl der Trainingsbeispiele zu verringern.
- System fordert Bewertungen für spezifische Objekte, von denen es vermutet, am meisten zu lernen.
- Verschiedene existierende Methoden:
  - Uncertainty Sampling
  - Komitee-basiertes Sampling

28

## Halbüberwachtes Lernen (weakly supervised, Bootstrapping)

---

- Verwende die große Anzahl ungekennzeichneter Beispiele, um das Lernen von einer kleinen Menge von gekennzeichneten Beispielen zu unterstützen.
- Einige neue Methoden:
  - Halbüberwachtes EM (Erwartungsmaximierung)
  - Ko-Training
  - Transduktive SVM's

29

## Schlussfolgerungen

---

- Empfehlungen und Personalisierung sind wichtige Ansätze zur Bekämpfung der Informations-Überfrachtung.
- Machinelles Lernen ist ein wichtiger Teil der Systeme für diese Aufgaben.
- Kollaboratives Filtern hat Probleme.
- Inhaltsbasierende Methoden sprechen diese Probleme an (haben aber eigene Probleme).
- Das beste ist, beides zu integrieren.

30