

Recommender-Systeme

Kollaboratives Filtern &
inhaltsbasierte Empfehlungen

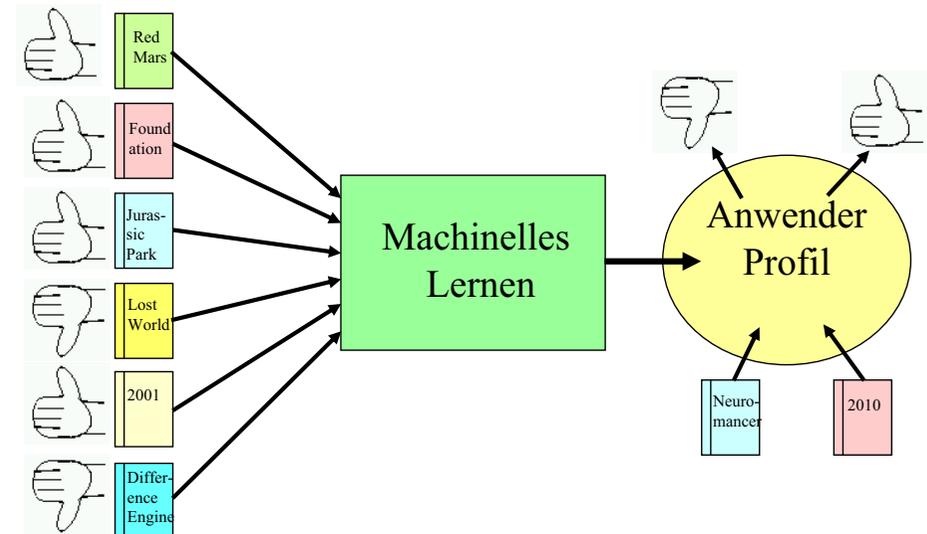
1

Empfehlungs-Systeme

- Systeme, um Nutzern Dinge zu empfehlen (z.B. Bücher, Filme, CDs, Webseiten, Newsgroup Nachrichten), die auf ihren vorigen Präferenzen basieren.
- Viele On-line-Läden liefern Empfehlungen (z.B. Amazon, CDNow).
- Recommender haben den Umsatz von On-Line-Läden erheblich erhöht.
- Es gibt zwei grundlegende Ansätze für Empfehlungen:
 - kollaboratives Filtern (a.k.a. soziales Filtern)
 - inhaltsbasiert

2

Buch-Recommender



3

Personalisierung

- Recommender sind spezielle Personalisierungs-Software.
- Die Personalisierung betrifft die Anpassung an individuelle Bedürfnisse, Interessen und Präferenzen jedes Anwenders.
- Recommender umfassen:
 - Empfehlen
 - Filtern
 - Vorhersagen (z.B. Formular-Vervollständigungen)
- Aus geschäftlicher Perspektive werden Recommender als Teil des Customer Relationship Management (CRM) angesehen.

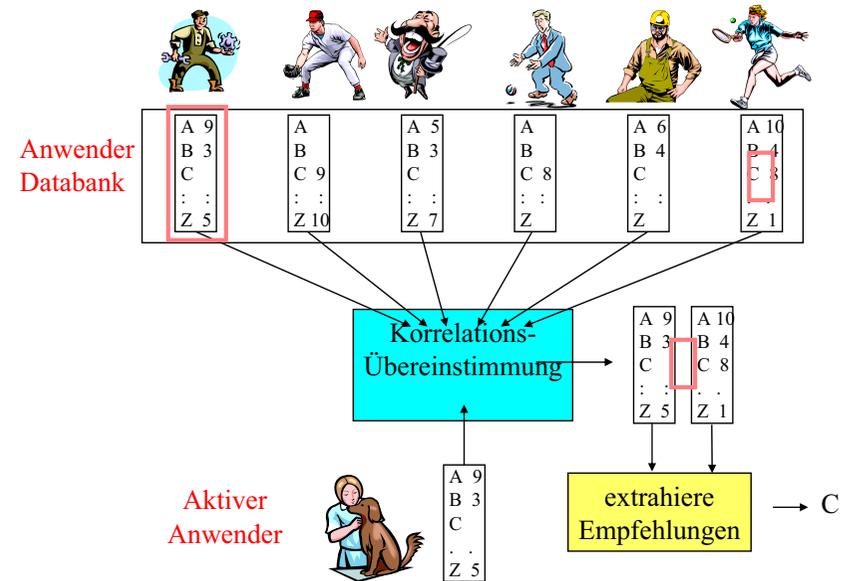
4

Machinelles Lernen und Personalisierung

- Machinelles Lernen kann das Lernen eines *Anwendermodells* oder *Profils* eines bestimmten Benutzers unterstützen, basierend auf:
 - Interaktionsmustern
 - bewerteten Beispielen
- Dieses Modell oder Profil kann dann verwendet werden um:
 - Objekte zu empfehlen
 - Informationen zu filtern
 - Verhalten vorherzusagen

5

Kollaboratives Filtern



7

Kollaboratives Filtern

- Pflegen einer Datenbank mit Anwenderbewertungen einer Vielzahl von Objekten.
- Finde für einen gegebenen Anwender andere, ähnliche Anwender, deren Bewertungen stark mit dem aktuellen Anwender korrelieren.
- Empfehle Objekte, die von diesen ähnlichen Anwendern hoch eingestuft werden, aber vom aktuellen Anwender noch nicht bewertet wurden.
- Nahezu alle vorhandenen kommerziellen Recommender verwenden diesen Ansatz (z.B. Amazon).

6

Kollaborative Filtermethode

- Gewichte alle Anwender in Bezug auf ihre Ähnlichkeit mit dem aktiven Anwender.
- Wähle eine Teilmenge der Anwender aus (*Nachbarn*), um sie zur Vorhersage zu verwenden.
- Normalisiere Bewertungen und berechne eine Vorhersage aus einem gewichteten Mittel der ausgewählten Nachbar-Bewertungen.
- Präsentiere Objekte mit höchsten vorhergesagten Bewertungen als Empfehlungen.

8

Ähnlichkeitsgewichtung

- Verwende typischerweise den Pearson-Korrelationskoeffizienten zwischen Bewertungen für den aktiven Anwender a und einem weiteren Anwender u .

$$c_{a,u} = \frac{\text{covar}(r_a, r_u)}{\sigma_{r_a} \sigma_{r_u}}$$

r_a und r_u sind die Bewertungsvektoren für die m Objekte, die **sowohl** von a als auch von u bewertet sind.

$r_{u,j}$ ist die Bewertung von Anwender u für das Objekt j .

9

Kovarianz und Standard-Abweichung

- Kovarianz:

$$\text{covar}(r_a, r_u) = \frac{\sum_{i=1}^m (r_{a,i} - \bar{r}_a)(r_{u,i} - \bar{r}_u)}{m}$$

$$\bar{r}_x = \frac{\sum_{i=1}^m r_{x,i}}{m}$$

- Standard-Abweichung:

$$\sigma_{r_x} = \sqrt{\frac{\sum_{i=1}^m (r_{x,i} - \bar{r}_x)^2}{m}}$$

10

Signifikanz-Gewichtung

- Es ist wichtig, keinen Korrelationen zu vertrauen, die nur auf sehr wenigen gemeinsam bewerteten Objekten basieren.
- Verwende *Signifikanzgewichte* $s_{a,u}$, die auf der Anzahl vom gemeinsam bewerteten Objekten, m basieren.

$$w_{a,u} = s_{a,u} c_{a,u} \quad \text{mit} \quad s_{a,u} = \begin{cases} 1 & \text{if } m > 50 \\ \frac{m}{50} & \text{if } m \leq 50 \end{cases}$$

11

Nachbar-Selektion

- Auswahl der zu dem aktiven Anwender a am stärksten korrelierenden Anwender, die dann als Quelle der Vorhersagen dienen.
- Der Standardansatz ist, die am ähnlichsten n Anwender u zu verwenden, basierend auf den Ähnlichkeitsgewichten $w_{a,u}$.
- Ein alternativer Ansatz ist es, alle Anwender einzuschließen, deren Ähnlichkeit zu a über einer gegebenen Schwelle liegt.

12

Bewertungs-Vorhersage

- Sage unter Verwendung der n ausgewählten Nachbar-Anwender $u \in \{1, 2, \dots, n\}$ für den aktiven Anwender a eine Bewertung $p_{a,i}$ für jedes Objekt i voraus .
- Um die verschiedenen Bewertungsniveaus unterschiedlicher Anwender zu berücksichtigen, basieren wir die Vorhersagen auf der Differenz zum Durchschnitt der Bewertungen aller Bewertungen des jeweiligen Nutzers.
- Gewichte die Bewertungsbeiträge des Anwenders nach ihrer Ähnlichkeit mit dem aktiven Anwender.

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n w_{a,u} (r_{u,i} - \bar{r}_u)}{\sum_{u=1}^n w_{a,u}}$$

13

Probleme mit kollaborativem Filtern

- **Kaltstart:** Es müssen bereits genug andere Anwender im System sein, um eine Übereinstimmung zu finden.
- **Seltenheit:** Wenn viele Objekte empfohlen werden sollen, ist die Anwender/ Bewertungsmatrix dünn besetzt – selbst wenn es viele Anwender gibt – und es ist schwierig, Anwender zu finden, die die gleichen Objekte bewertet haben.
- **Erster Beurteiler:** K.F. kann kein Objekt empfehlen, das nicht zuvor bewertet worden ist.
 - Neue Objekte
 - Exotische Objekte
- **Popularitäts-Ausrichtung:** koll. Filtern kann jemandem mit sehr speziellen Vorlieben keine Objekte empfehlen.
 - Die Methode neigt dazu, populäre Objekte zu empfehlen.

14

Inhaltsbasiertes Empfehlen

- Empfehlungen basieren hier eher auf Informationen über den **Inhalt** (die Eigenschaften) von Objekten als auf den Meinungen anderer Anwender.
- Verwendet einen Algorithmus für maschinelles Lernen, um ein Profil der Anwenderpräferenzen aus Beispielen zu erzeugen, die auf Merkmalsbeschreibungen des Inhalts basieren.
- Einige existierende Anwendungen:
 - Newsweeder (Lang, 1995)
 - Syskill und Webert (Pazzani et al., 1996)

15

Vorteile eines inhaltsbasierten Ansatzes

- Kein Bedarf an Daten über andere Anwender.
 - Kein Kaltstart-Problem und keine Seltenheitsprobleme.
- Ist fähig, Anwendern mit eindeutigen Vorlieben Empfehlungen auszusprechen
- Ist fähig, neue und unpopuläre Objekte zu empfehlen
 - Kein Erster-Beurteiler-Problem.
- Kann Erläuterungen zu den empfohlenen Objekten durch die Auflistung der Inhaltsmerkmale liefern, die die Empfehlung bewirkten.

16

Nachteile der inhaltsbasierten Methode

- Erfordert Inhalt, der sinnvoll durch Merkmale kodiert werden kann.
- Anwender-Vorlieben müssen als lernbare Funktion dieser Inhaltsmerkmale dargestellt werden können.
- Nicht fähig, Qualitätsbeurteilungen anderer Anwender auszuwerten.
 - Es sei denn, diese sind irgendwie in den Inhaltsmerkmalen enthalten.