

Recommender-Systeme

Seminararbeit WS 25/26

Von: Fabian Koßmann

Betreuer: Prof. Dr. Christoph Schmitz

Informatik
Hauptcampus

H O C H
S C H U L E
T R I E R

1. Grundlagen von Recommender-Systemen
2. User-Item-Matrix & Nutzerdaten
3. Kaltstartproblem
4. Ähnlichkeits- und Vergleichsmaße
5. Item- & User-Profile (Feature-Extraktion & -Repräsentation)
6. Content-basierte Empfehlungen
7. Kollaboratives Filtern
8. Erweiterte Modellierungsansätze (Binärisierung, Rating-Normalisierung, Clustering, Klassifikation)
9. Evaluationsmetriken (Precision@k, Recall@k)

- System zur **Vorhersage von Nutzerreaktionen**
- Ziel: **relevante & nützliche Items empfehlen**
- Nutzung von:
 - früheren Interaktionen
 - expliziten / impliziten Präferenzen
 - Kontextinformationen
- Zwei Hauptansätze:
 - **Content-based**: Abgleich von Item-Features mit Nutzerprofil
 - **Collaborative Filtering**: Ähnlichkeiten zwischen Nutzern oder Items

	Der Herr der Ringe	1984	Harry Potter 1	Brave New World
Anna	5	4	4	?
Ben	3	5	?	4
Clara	?	4	5	3
David	4	?	4	5
Emma	2	5	?	4

- **Explizite Nutzerdaten**

Direkt durch aktives Feedback der Nutzer erhoben
(z. B. Bewertungen, Likes, Umfragen)

- **Implizite Nutzerdaten**

Indirekt aus dem Nutzungsverhalten abgeleitet
(z. B. Klicks, Verweildauer, Suchanfragen)

- Problem bei **neuen Nutzern oder Items**
- **Zu wenige Interaktionen** => leere User-Item-Matrix
- Präferenzen **nicht zuverlässig schätzbar**
- Typische Lösungen:
 - **Explizite Nutzerangaben** (Initial-Bewertungen)
 - **Inhaltsbasierte Merkmale / Demographie**
- Nachteile: **hoher Nutzeroaufwand, ungenaue Annahmen**

- Vergleichen Items miteinander
- Bestimmen die Ähnlichkeit zwischen Nutzern
- Bewerten die Übereinstimmung zwischen Nutzer- und Item-Profilen
- **Typische Maße**
 - Kosinus Ähnlichkeit
 - Jaccard Distanz

$$UP_{\text{Clara}} = [0, 1, 1, 1, 1, 2, 0.51]$$

$$IP_{\text{Der Herr der Ringe}} = [1, 0, 0, 0, 1, 0, 1.02]$$

$$d_{\cos}(UP, IP) = \frac{1.5202}{2.874 \cdot 1.744} \approx 0.303$$

Der Herr der Ringe (A) und *Harry Potter* (B)

$A = \{\text{Zauberer, Schule, Magie, Zauberstab}\}$

$B = \{\text{Magie, Schule, Freundschaft, Abenteuer}\}$

$A \cap B = \{\text{Magie, Schule}\}$

$A \cup B = \{\text{Zauberer, Schule, Magie, Zauberstab, Freundschaft, Abenteuer}\}$

$$J(A, B) = \frac{2}{6} = \frac{1}{3}$$

$$d_J(A, B) = 1 - \frac{1}{3} = \frac{2}{3}$$

$$IP_{\text{features}} = [\text{Autor(en)}, \dots, \text{Genre(s)}, \dots, \text{Seitenanzahl}]$$

	Tolkien	Orwell	Rowling	Huxley	Fantasy	Dystopie	Seitenzahl (norm.)
Der Herr der Ringe	1	0	0	0	1	0	1.02
1984	0	1	0	0	0	1	0.61
Harry Potter 1	0	0	1	0	1	0	0.46
Brave New World	0	0	0	1	0	1	0.45

- Items oft **nicht-atomar** (Text, Bilder, Videos)
- Inhalte liegen **nicht strukturiert** vor
- relevante Merkmale müssen aus Rohdaten extrahiert werden

TF-IDF (für Textdokumente)

- Misst Relevanz eines Begriffs im Dokument **relativ zum Korpus**
- Hohe TF-IDF-Werte als **charakteristische Begriffe**
- Geeignet zur **numerischen Dokumentrepräsentation**

Tagging (medienübergreifend)

- Items werden mit **beschreibenden Tags** annotiert
- Tags als **semantische Features**
- Manuell (Nutzer) oder automatisch (Algorithmen)
- Einsetzbar für **Texte, Bilder, Videos**

- Jedes Item wird als **Vektor relevanter Eigenschaften** modelliert

Diskrete Merkmale

- z. B. Genre, Schauspieler, Regisseur
- Darstellung als **binäre Features** (0/1)

Numerische Merkmale

- z. B. Bewertungen, Seitenzahlen

Reellwertige Features notwendig

- Binäre Kodierung würde Informationsgehalt verlieren

Praxisaspekt

- Kombination aus **binären & numerischen Features**
- **Skalierung numerischer Merkmale** wichtig für Ähnlichkeitsberechnung

$$UI_{\text{Clara}} = [?, 4, 5, 3]$$

$$IP_{1984} = [0, 1, 0, 0, 0, 1, 0.61]$$

$$IP_{\text{Harry Potter 1}} = [0, 0, 1, 0, 1, 0, 0.46]$$

$$IP_{\text{Brave New World}} = [0, 1, 0, 0, 0, 1, 0.45]$$

$$UP_{\text{Clara}} = [0, 1, 1, 1, 1, 2, 0.51]$$

	Tolkien	Orwell	Rowling	Huxley	Fantasy	Dystopie	Seitenzahl (norm.)
Anna	1.0	1.0	1.0	0.0	2.0	1.0	0.70
Ben	1.0	1.0	0.0	1.0	1.0	2.0	0.69
Clara	0.0	1.0	1.0	1.0	1.0	2.0	0.51
David	1.0	0.0	1.0	1.0	2.0	1.0	0.64
Emma	1.0	1.0	0.0	1.0	1.0	2.0	0.69

$$UP_{\text{Anna}} = [1, 1, 1, 0, 2, 1, 0.70]$$

$$IP_{\text{Brave New World}} = [0, 0, 0, 1, 0, 1, 0.45]$$

$$d_{\cos}(UP_{\text{Anna}}, IP_{\text{Brave New World}}) = \frac{1.315}{2.914 \cdot 1.484} \approx 0.30$$

$$\begin{aligned}d_{\cos}(\text{Anna, Ben}) &\approx 0.937, \quad d_{\cos}(\text{Anna, Clara}) \approx 0.994 \\d_{\cos}(\text{Anna, David}) &\approx 0.994, \quad d_{\cos}(\text{Anna, Emma}) \approx 0.87 \\r_{\text{Clara, Brave New World}} &= 3 \quad \text{und} \quad r_{\text{David, Brave New World}} = 5 \\r_{\text{Anna, Brave New World}} &= \frac{3 + 5}{2} = 4\end{aligned}$$

$$(UI_B)_{u,i} = \begin{cases} 1, & \text{falls } r_{u,i} \geq 3 \\ 0, & \text{sonst.} \end{cases}$$

	Der Herr der Ringe	1984	Harry Potter 1	Brave New World
Anna	1	1	1	0
Ben	1	1	0	1
Clara	0	1	1	1
David	1	0	1	1
Emma	0	1	0	1

	Der Herr der Ringe	1984	Harry Potter 1	Brave New World
Anna	0.67	-0.33	-0.33	?
Ben	-1.00	1.00	?	0.00
Clara	?	0.00	1.00	-1.00
David	-0.33	?	-0.33	0.67
Emma	-1.67	1.33	?	0.33

$$r_{\text{Anna}, C_1} = \frac{5 + 4}{2} = 4.5$$

$$r_{\text{Ben}, C_1} = 3, \quad r_{\text{Clara}, C_1} = 5, \quad r_{\text{David}, C_1} = 4, \quad r_{\text{Emma}, C_1} = 2$$

$$r_{U_1, C_1} = \frac{3 + 2}{2} = 2.5$$

$$r_{U_1, 1984} = 5 \quad \text{und} \quad r_{U_1, \text{Brave New World}} = 4$$

	C_1 (Der Herr der Ringe , Harry Potter 1)	1984	Brave New World
Anna	4.5	4.0	?
Clara	5.0	4.0	3.0
David	4.0	?	5.0
U_1 (Ben, Emma)	2.5	5.0	4.0

- Vorhersage von Nutzer–Item-Interaktionen als **Klassifikationsproblem**
- Bewertungen werden als **Klassen** modelliert
(z. B. *positiv / negativ* oder diskrete Scores)
- Einsatz **überwachter ML-Verfahren**
(z. B. Entscheidungsbäume, neuronale Netze)
- Modelle nutzen **Feature-Repräsentationen** aus
 - Nutzerdaten
 - Item-Eigenschaften
- Lernen implizit **Nutzerpräferenzen**
- Ermöglichen Vorhersagen für **unbekannte Nutzer–Item-Paare**

$$\text{Precision@k}(U) = \frac{|R_U \cap G_U|}{k}$$

$$R_{\text{Anna}} = \{ \textit{Brave New World}, 1984 \}$$

$$G_{\text{Anna}} = \{ \textit{Brave New World} \}$$

$$R_{\text{Anna}} \cap G_{\text{Anna}} = \{ \textit{Brave New World} \}$$

$$\text{Precision@2}(\text{Anna}) = \frac{|R_{\text{Anna}} \cap G_{\text{Anna}}|}{2} = \frac{1}{2} = 0.5$$

$$\text{Recall@k}(U) = \frac{|R_U \cap G_U|}{|G_U|}$$

$$R_{\text{Anna}} = \{ \textit{Brave New World}, 1984 \}$$

$$G_{\text{Anna}} = \{ \textit{Brave New World} \}$$

$$R_{\text{Anna}} \cap G_{\text{Anna}} = \{ \textit{Brave New World} \}$$

$$\text{Recall@2}(\text{Anna}) = \frac{|R_{\text{Anna}} \cap G_{\text{Anna}}|}{|G_{\text{Anna}}|} = \frac{1}{1} = 1.0$$

- Überblick über **Recommender-Systeme**
(Content-based, Collaborative Filtering, moderne Ansätze)
- Wachsende Bedeutung durch **steigende Datenmengen**
- Einfluss auf **Informationszugang & Meinungsbildung**
- **Ausblick:**
ML-Methoden, Fairness & Transparenz, gesellschaftliche Effekte

- **Leskovec et al. (2020)** – *Mining of Massive Datasets*:
Standardwerk zu Datenanalyse, u. a. Grundlagen und Methoden von Recommender-Systemen.
- **Ricci et al. (2011)** – *Recommender Systems Handbook*:
Umfassendes Referenzwerk zu Theorie, Algorithmen, Evaluation und Anwendungen von Recommender-Systemen.
- **Felfernig et al. (2018)** – *Evaluating Group Recommender Systems*:
Methoden und Metriken zur Bewertung von Gruppen-Recommender-Systemen.