

Seminar Collaborative Filtering

KDD Cup

Ziawasch Abedjan, Arvid Heise, Felix Naumann

Collaborative Filtering

Recommendation systems

3

The image shows a Google search interface with the search term 'inception'. The search bar contains 'inception' and a search button. Below the search bar, it indicates 'About 64,800,000 results (0.11 seconds)' and a link to 'Advanced search'. On the left side, there is a navigation menu with options: 'Everything', 'Images', 'Videos', 'News', 'Shopping', 'Realtime', and 'More'. Below the menu, there is a section for 'Any time'. The search results are displayed in a list format. The first result is for 'Inception (2010) - IMDb' with a rating of 8.9/10. The description mentions it is directed by Christopher Nolan and stars Leonardo DiCaprio, Joseph Gordon-Levitt, and Ellen Page. Below the description, there are links for 'Full cast and crew', 'Plot Summary', 'Trivia', 'Synopsis', and 'www.imdb.com/title/tt1375666/'. The word 'Similar' is circled in red. The second result is for 'Inception - Wikipedia, the free encyclopedia' with a brief description of the film and links for 'Joseph Gordon-Levitt', 'Christopher Nolan', 'Ellen Page', and 'Marion Cotillard'.

Google

inception × Search

About 64,800,000 results (0.11 seconds) [Advanced search](#)

Everything
Images
Videos
News
Shopping
Realtime
More

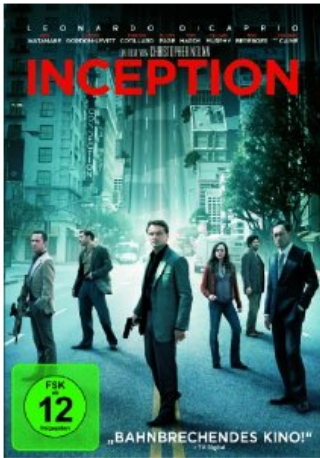
Any time

Inception (2010) - IMDb 🔍
★★★★☆ Rating: 8.9/10 - from 332,186 users
Directed by [Christopher Nolan](#). Starring [Leonardo DiCaprio](#), [Joseph Gordon-Levitt](#), [Ellen Page](#).
In a world where technology exists to enter the human mind through dream invasion, a highly skilled thief is given a final chance at ...
[Full cast and crew](#) - [Plot Summary](#) - [Trivia](#) - [Synopsis](#)
www.imdb.com/title/tt1375666/ - [Cached](#) - [Similar](#)

Inception - Wikipedia, the free encyclopedia 🔍
Inception is a 2010 science fiction action heist film, written, co-produced, and directed by Christopher Nolan. The film features an ensemble cast starring ...
[Joseph Gordon-Levitt](#) - [Christopher Nolan](#) - [Ellen Page](#) - [Marion Cotillard](#)
en.wikipedia.org/wiki/Inception - [Cached](#) - [Similar](#)

Recommendation systems

4



Inception

[Leonardo DiCaprio](#) (Darsteller), [Joseph Gordon-Levitt](#) (Darsteller), [Christopher Nolan](#) (Regisseur) | Alterseinstufung: Freigegeben ab 12 Jahren | Format: DVD

★★★★☆ (505 Kundenrezensionen) **Gefällt mir** (39)

Preis: **EUR 7,95** Kostenlose Lieferung ab EUR 20 (auch bei Verkäufern mit "Versand durch Amazon.de").

Alle Bücher und Blu-rays versandkostenfrei. [Details](#)

Alle Preisangaben inkl. MwSt.

Auf Lager.

Verkauf und Versand durch **Amazon.de**. Geschenkverpackung verfügbar.

Lieferung bis Dienstag, 12. April: Bestellen Sie in den nächsten 6 Stunden und 17 Minuten per **Overnight-Express**.

[Siehe Details.](#)

68 neu ab EUR 4,71 **50 gebraucht** ab EUR 3,00 **1 Sammlerstück(e)** ab EUR 12,96



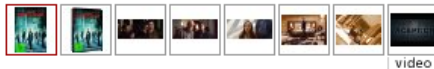
Leihen Sie DVDs & Blu-rays bei LOVEFiLM

LOVEFiLM ist Europas größte Online-Videothek mit einer Auswahl von über 40.000 Titeln.

Testen Sie LOVEFiLM.de jetzt 14 Tage lang kostenlos und erhalten Sie einen 20 EUR

Gutschein für Amazon.de. [Alles zur Aktion bei LOVEFiLM.de](#)

Für Bildvergrößerung und andere Ansichten klicken



[2 weitere Bilder und Videos ansehen](#)

[Eigene Bilder hinzufügen](#)

Weitere Ausgaben: Preis:

[Blu-ray](#)

Weitere Angebote:

EUR 11,99 [76 Angebote](#) ab EUR 7,50

Kunden, die diesen Artikel gekauft haben, kauften auch



Salt (Deluxe Extended Edition) [Deluxe Edition] DVD ~ Angelina Jolie

★★★★☆ (146)

EUR 8,97



Knight and Day - Agentenpaar wider Willen (Exten... DVD ~ Tom Cruise

★★★★☆ (115)

EUR 7,97



Das A-Team - Der Film (Extended Cut) DVD ~ Liam Neeson

★★★★☆ (177)

EUR 14,99



Shutter Island DVD ~ Leonardo DiCaprio

★★★★☆ (195)

EUR 8,91




Avatar (Extended Collector's Edition, 3... DVD ~ Sam Worthington

★★★★☆ (241)

EUR 17,95

Recommendation systems

5



Boulevard of Broken Dreams

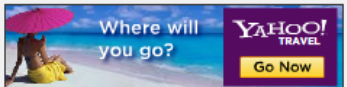
Rate: ☆☆☆☆☆ [Share](#) 374 [retweet](#) 4

by Green Day

[Add To Playlist](#)


© 2011 Reprise

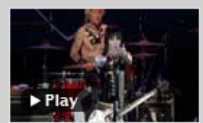
Categories: Rock, Video Premieres, Alternative Rock





Embed This Video:


```
<object width="400" height="255" id="uvp_fop" allowFull
```

- 

Boulevard of Broken ...
Green Day
- 


I Hate Myself For Lovi...
Joan Jett & The Blackhe...
- 


Riptide
Sick Puppies
- 


Been To Hell
Hollywood Undead
- 


Rope
Foo Fighters

More Green Day Videos

- 

21 Guns
Green Day
- 

Boulevard Of Broken Dr...
Green Day
- 

American Idiot
Green Day
- 

When I Come Around
Green Day



ALL YOU CAN EAT
in Deiner Stadt
ab
5€

Recommendation systems

6

- Content-based recommendations
 - Relevance is based on content
 - Similar items
 - Uses also metadata

- Collaborative filtering
 - Recommend items suiting one's taste
 - Based on community ratings
 - Serendipity

Collaborative Filtering Algorithms

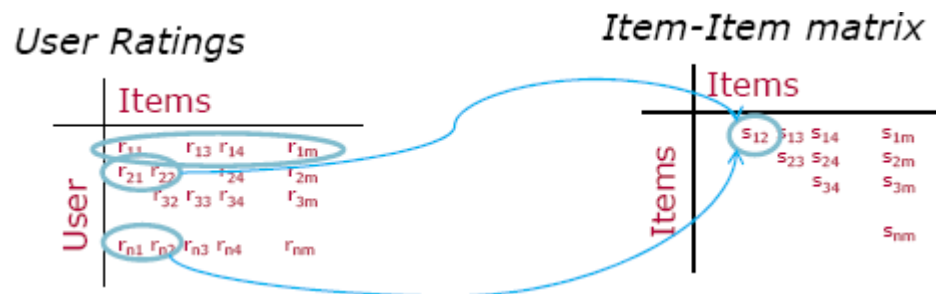
7

- Memory-based CF
 - Item/user-based top N
 - New items and user can be added easily
 - Scales well on dense data sets
 - Limited scalability on sparse data sets
- Model-based CF
 - Singular Value Decomposition (SVD)
 - Better scalability on sparse data sets
 - Little evidence for the model
- Hybrid recommenders
 - Content-boosted CF
 - Mixture of memory-based and model-based techniques

Neighbor-based CF

8

- Process: filtering -> prediction -> recommendation
- User-based approach
 - Create cluster of similar users
 - Recommendations depend on ratings of similar users
 - Problem: missing ratings
- Item-based approach
 - Create item-item matrix based on rating similarities
 - Retrieve top K similar items and aggregating their ratings by the user

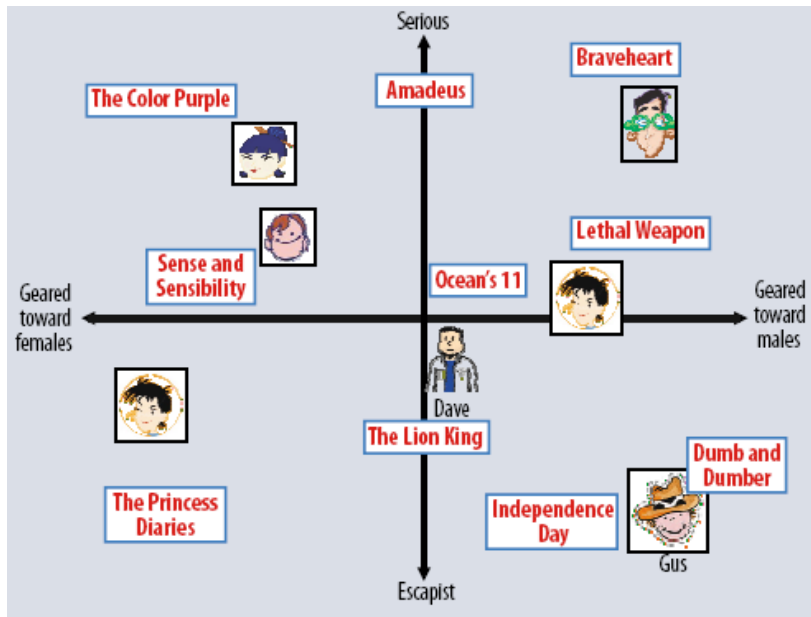


Singular Value Decomposition

9

- Reduce dimensions of user/item matrix by factorization
- Use overlapping features (e.g. gender, genre)

$$\begin{array}{c|cc} & \text{items} & \\ \hline \text{user} & r_{11} & r_{1n} \\ & r_{i1} & r_{ij} \\ & r_{m1} & r_{mn} \end{array} = \begin{array}{c|cc} & \text{features} & \\ \hline \text{user} & f_{11} & f_{12} \\ & f_{m1} & f_{m2} \end{array} \times \begin{array}{c|ccc} & \text{items} & & \\ \hline \text{feature} & g_{11} & g_{12} & g_{1n} \\ & g_{21} & g_{22} & \dots \\ & & & g_{2n} \end{array}$$



KDD Cup

Overview

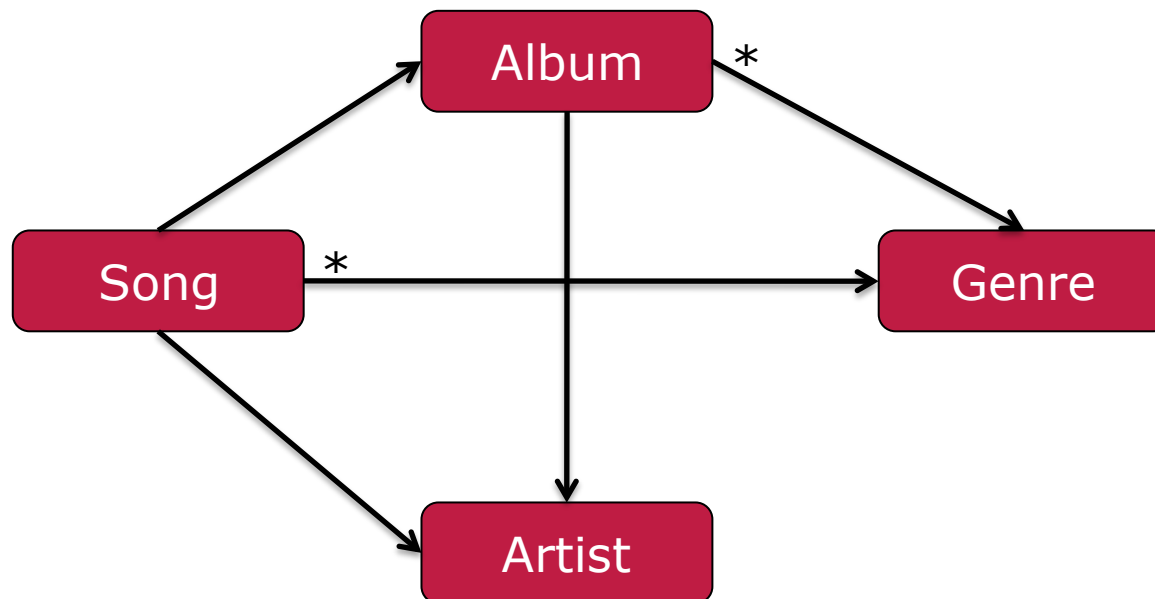
11

- ACM SIGKDD
 - Knowledge Discovery and Data Mining

- Variety of topics
 - 2010 - Student performance evaluation
 - 2008 - Breast cancer
 - 2007 - **Consumer recommendations**
 - 2004 - Particle physics; plus protein homology prediction
 - 2001 - Molecular bioactivity; plus protein locale prediction
 - 1997 - Direct marketing for lift curve optimization

- 2 tracks with three places

- On Yahoo Music Dataset
 - Artists, Albums, Songs, Genres
- Track 1: predict user rating
- Track 2: decide whether a user rates a song or not



Track 1

13

- Typical collaboration filtering task
- Predict the rating of a specific user for unrated songs
- Includes hierarchy of items
 - User might have rated other songs of same album/artist
 - No user might have rated the song but the same album/artist
- Includes time stamp of rating
 - Rating behavior might have changed over time
 - Older songs rated differently than newer songs?

#Users	#Items	#Ratings	#Train Ratings	#Validation Ratings	#Test Ratings
1,000,990	624,961	262,810,175	252,800,275	4,003,960	6,005,940

Track 2

14

- Predict if a user would rate a given song highly or not at all
- Need a model for rate behavior
 - Machine learning?
- Ratings on songs only
- No timestamps
- Given six songs, which three will most likely not be rated

#Users	#Items	#Ratings	#Train Ratings	#Test Ratings	#Users
249,012	296,111	62,551,438	61,944,406	607,032	249,012

Seminar

Approaches Track 1

16

- Koren, Y., Bell, R. & Volinsky, C.: "Matrix Factorization Techniques for Recommender Systems"
 - Similar to Singular Value Decomposition
 - Create small abstracted matrix
 - Soft clustering

- Sarwar, B., Karypis, G., Konstan, J. & Reidl, J.: "Item-based collaborative filtering recommendation algorithms"
 - Classic approach
 - Improvements in Robert M. Bell & Yehuda Koren: "Improved Neighborhood-based Collaborative Filtering"
 - Need to be adjusted to include time stamp and hierarchy

Approaches Track 2

17

- Kurucz, M., Benczúr, A.A., Kiss, T., Nagy, I., Szabó, A. & Torma, B.: "Who Rated What: a Combination of SVD, Correlation and Frequent Sequence Mining"
 - Combination of four predictions (only two relevant)
 - ◇ SVD and base prediction
 - Needs estimation on how many ratings

- Sueiras, J.: "A classical predictive modeling approach for Task "Who rated what?" of the KDD CUP 2007"
 - Machine learning (logistic regression)
 - User-specific, movie-related, and user-movie pair features
 - SVD for movie-related

Application

18

- Send mail to ziawasch.abedjan@hpi.uni-potsdam.de
- Subject: [CF Seminar]
- Deadline: April 18th
- Notification: April 19th

- Limited to 8 participants = 4 teams
 - Random selection if more applicants

- Send top 3 wishes on tasks and approaches
 - We are open to other approaches

Time Schedule

19

- April 14th: first seminar, topic presentations
- April **16th: application deadline**, team/paper preferences
- April **17th**: team/paper notification

- April 21th: mandatory consultation
- April 28th: **paper presentation**

- May 12th: initial implementation/idea presentation
- June 9th: intermediate presentation/project consolidation
- June 23th: **final presentations**
- June 30th: KDD cup **submission deadline** (results only)

- July 9th: hand in **short paper** (6 pages)

Grading

20

- 3 LP (half semester, project seminar)
- Paper and final presentations
- Participation during intermediate presentations / discussions
- Implementation strategies/proposed extensions
- Short paper
- Bonus: good results in KDD cup

Options

21

- Mahout or stand-alone

- Single or combined repository
 - We recommend cooperation
 - Many similar issues

References

22

- Official KDD Cup: <http://kddcup.yahoo.com/>
- Su, Xiaoyuan & Khoshgoftaar, Taghi M.: "A survey of collaborative filtering techniques"
- Koren, Yehuda; Bell, Robert & Volinsky, Chris: "Matrix Factorization Techniques for Recommender Systems"
- Sarwar, Badrul; Karypis, George, Konstan, Joseph & Reidl, John: "Item-based collaborative filtering recommendation algorithms"
- Improvements in Robert M. Bell & Yehuda Koren: "Improved Neighborhood-based Collaborative Filtering"
- Miklós Kurucz, András A. Benczúr, Tamás Kiss, István Nagy, Adrienn Szabó & Balázs Torma: "Who Rated What: a Combination of SVD, Correlation and Frequent Sequence Mining"
- Jorge Sueiras: "A classical predictive modeling approach for Task "Who rated what?" of the KDD CUP 2007"