

Collaborative Filtering

Peter Brusilovsky

with slides by Danielle Lee and Sue Syn

Where we are?

	Search	Navigation	Recommendation
Content-based			
Semantics / Metadata			
Social			

Agenda

- Context
 - Concepts
 - Uses
 - CF vs. CB
 - Algorithms
 - Practical Issues
 - Evaluation Metrics
 - Future Issues
-

Types of Recommender Systems

- Collaborative Filtering Recommender System
 - “Word-of-Mouth” phenomenon.
- Content-based Recommender System
 - Recommendation generated from the content features associated with products and the ratings from a user.
- Case-based Recommender System
 - A kind of content-based recommendation. Information are represented as case and the system recommends the cases that are most similar to a user’s preference.
- Hybrid Recommender System
 - Combination of two or more recommendation techniques to gain better performance with fewer of the drawbacks of any individual one (Burke, 2002).

Recommendation Procedure

1. Understand and model users
2. Collect candidate items to recommend.
3. Based on your recommendation method, predict target users' preferences for each candidate item.
4. Sort the candidate items according to the prediction probability and recommend them.

Example: Amazon.com

amazon.com Hello, Peter Brusilovsky. We have [recommendations](#) for you. ([Not Peter?](#)) FREE 2-Day Shipping: [See details](#)

Peter's Amazon.com | [Today's Deals](#) | [Gifts & Wish Lists](#) | [Gift Cards](#) [Your Digital Items](#) | [Your Account](#) | [Help](#)

Shop All Departments All Departments

Your Amazon.com | [Your Browsing History](#) | [Recommended For You](#) | [Rate These Items](#) | [Improve Your Recommendations](#) | [Your Profile](#) | [Learn More](#)

Peter, Welcome to Your Amazon.com ([if you're not Peter Brusilovsky, click here.](#))

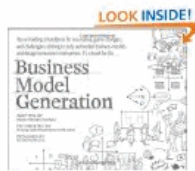
Today's Recommendations For You

Here's a daily sample of items recommended for you. Click here to [see all recommendations](#).

Page 1 of 35



[Shrek Forever After \(Single-Disc... DVD](#) ~ Mike Myers
★★★★☆ (143) \$19.49
[Fix this recommendation](#)



[Business Model Generati...](#)
(Paperback) by Alexander Osterwalder
★★★★☆ (102) \$20.38
[Fix this recommendation](#)



[Kingston 4 GB Class 4 SDHC Flash Memory Card SD...](#)
★★★★☆ (677) \$6.49
[Fix this recommendation](#)



[SE 19 PCS Watch Tool Kit](#)
★★★★☆ (44) \$21.22
[Fix this recommendation](#)



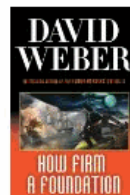
[Sony 2 GB Memory Stick PRO Duo Flash Memory Car...](#)
★★★★☆ (675) \$4.98
[Fix this recommendation](#)

Coming Soon for You

Page 1 of 2



[Source Code \[Blu-ray\]](#) Blu-ray ~ Jake Gyllenhaal
★★★★☆ (41) \$19.99
[Fix this recommendation](#)



[How Firm a Foundation \(Safehold\)](#) (Hardcover) by David Weber
\$16.79



[Bleach Uncut Box Set 9 DVD](#) ~ Artist Not Provided
★★★★☆ (6) \$36.99
[Fix this recommendation](#)

Tap into Your Friends

BETA




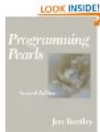


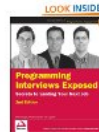



Connect to Facebook to get Amazon recommendations for you and discover your friends' Favorites and Likes

Amazon's Source of Wisdom

Customers Who Bought This Item Also Bought

Page



 <p>Programming Pearls (2nd Edition) by Jon Bentley ★★★★☆ (39) \$28.56</p>	 <p>Cracking the Coding Interview, Fourth Edition:... by Gayle Laakmann ★★★★☆ (28)</p>	 <p>The Algorithm Design Manual by Steven S. Skiena ★★★★☆ (42) \$64.75</p>	 <p>Programming Interviews Exposed: Secrets to Landing... by John Mongan ★★★★☆ (39) \$16.77</p>	 <p>Concrete Mathematics: A Foundation for Comput... by Ronald L. Graham ★★★★☆ (31) \$116.20</p>	 <p>Artificial Intelligence: A Modern Approach (3rd... by Stuart Russell ★★★★☆ (19) \$116.20</p>	 <p>Algorithm Design by Jon Kleinberg ★★★★☆ (23) \$105.00</p>
--	--	--	--	--	--	---

Frequently Bought Together



Price For All Three: **\$117.36**

 Add all three to Cart

 Add all three to Wish List

[Show availability and shipping details](#)

- This item:** Introduction to Algorithms, Third Edition by Thomas H. Cormen Hardcover **\$60.00**
- Programming Pearls (2nd Edition) by Jon Bentley Paperback **\$28.56**
- Cracking the Coding Interview, Fourth Edition: 150 Programming Interview Questions and Solutions by Gayle Laakmann Paperback **\$28.80**

Additional Items to Explore

You viewed

Customers who viewed this also viewed



Nikon D3100 14.2MP Digital SLR Camera...



Nikon 55-200mm f/4-5.6G ED IF AF-S DX...
~~\$249.99~~ Click for details



Nikon D7000 16.2MP DX-Format CMOS...
\$1,239.99



Nikon D7000-D5000-D3100-D3000 DSLR...
~~\$76.00~~ **\$37.99**

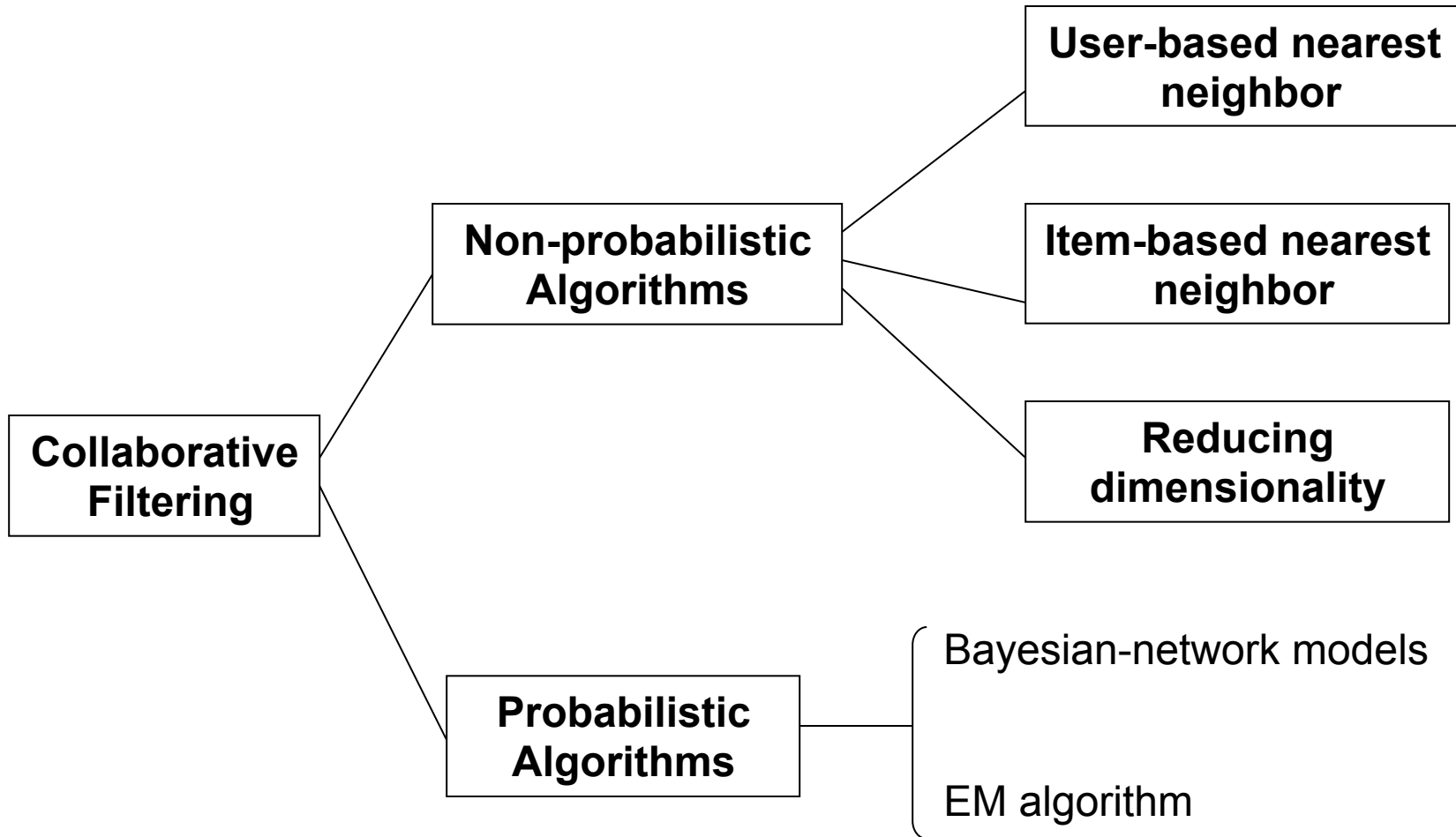


Nikon D3100 14.2MP Digital SLR Camera...
~~\$949.00~~ **\$746.95**

What is Collaborative Filtering?

- ▶ Traced back to the *Information Tapestry* project at Xerox PARC
 - ▶ It allowed its users to annotate the documents that they read and system recommends
- ▶ Expanded to “automatic” CF in the works of Resnick, Riedl, Maes
- ▶ More general definition as ‘the process of filtering or evaluating items using the opinions of other people.’
- ▶ CF recommends items which are likely interesting to a target user based on the evaluation averaging the opinions of people with similar tastes
- ▶ Key idea: people who agreed with me in the past, will also agree in the future.
 - ▶ On the other hand, the assumption of Content-based recommendation is that Items with similar objective features will be rated similarly.

Algorithms



Concepts

- Collaborative Filtering

- The goal of collaborative filtering is to predict how well a user will like an item that he has not rated given a set of historical preference judgments for a community of users.

- User

- Any individual who provides ratings to a system

- Items

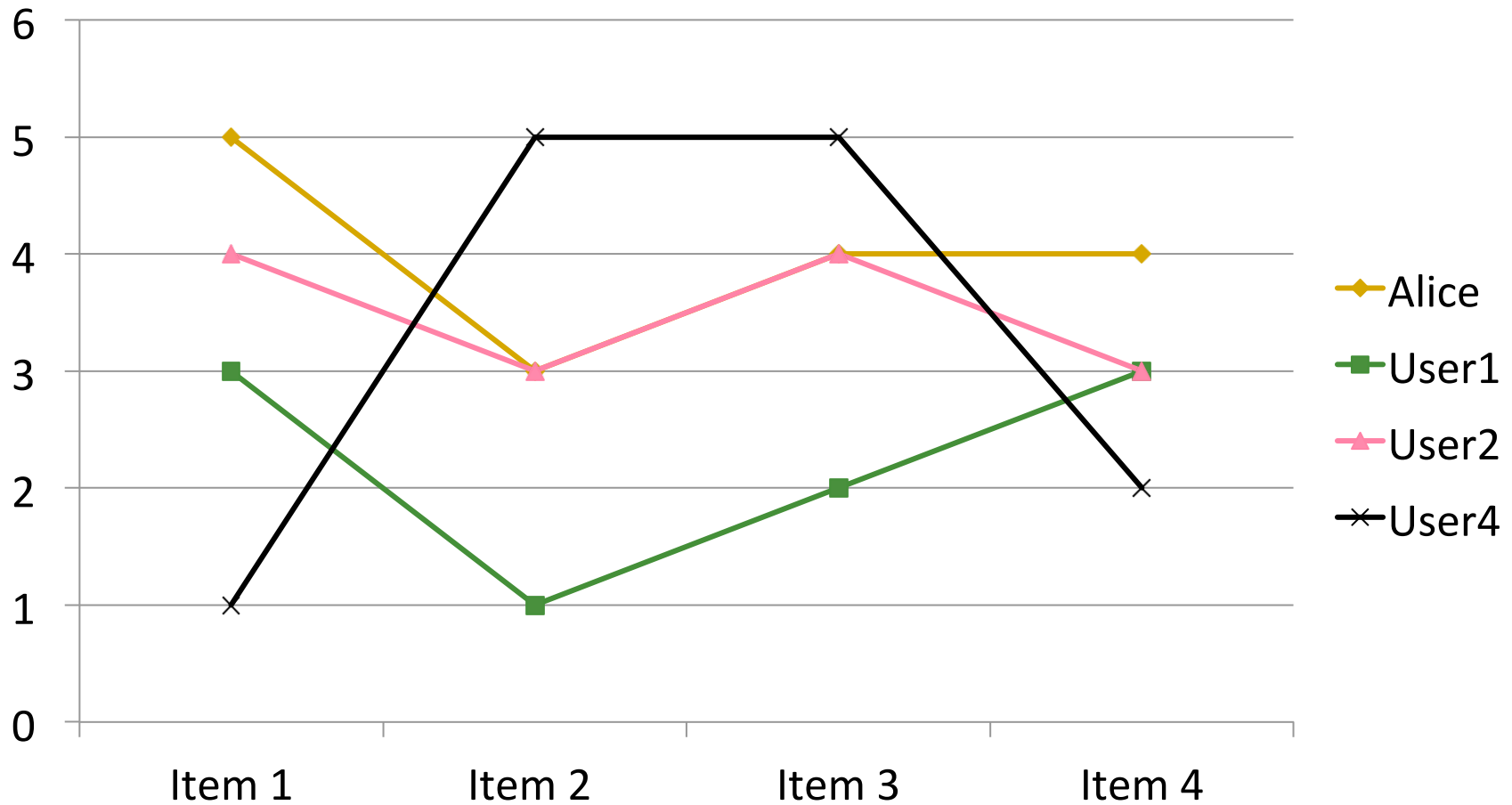
- Anything for which a human can provide a rating
-

User-based CF

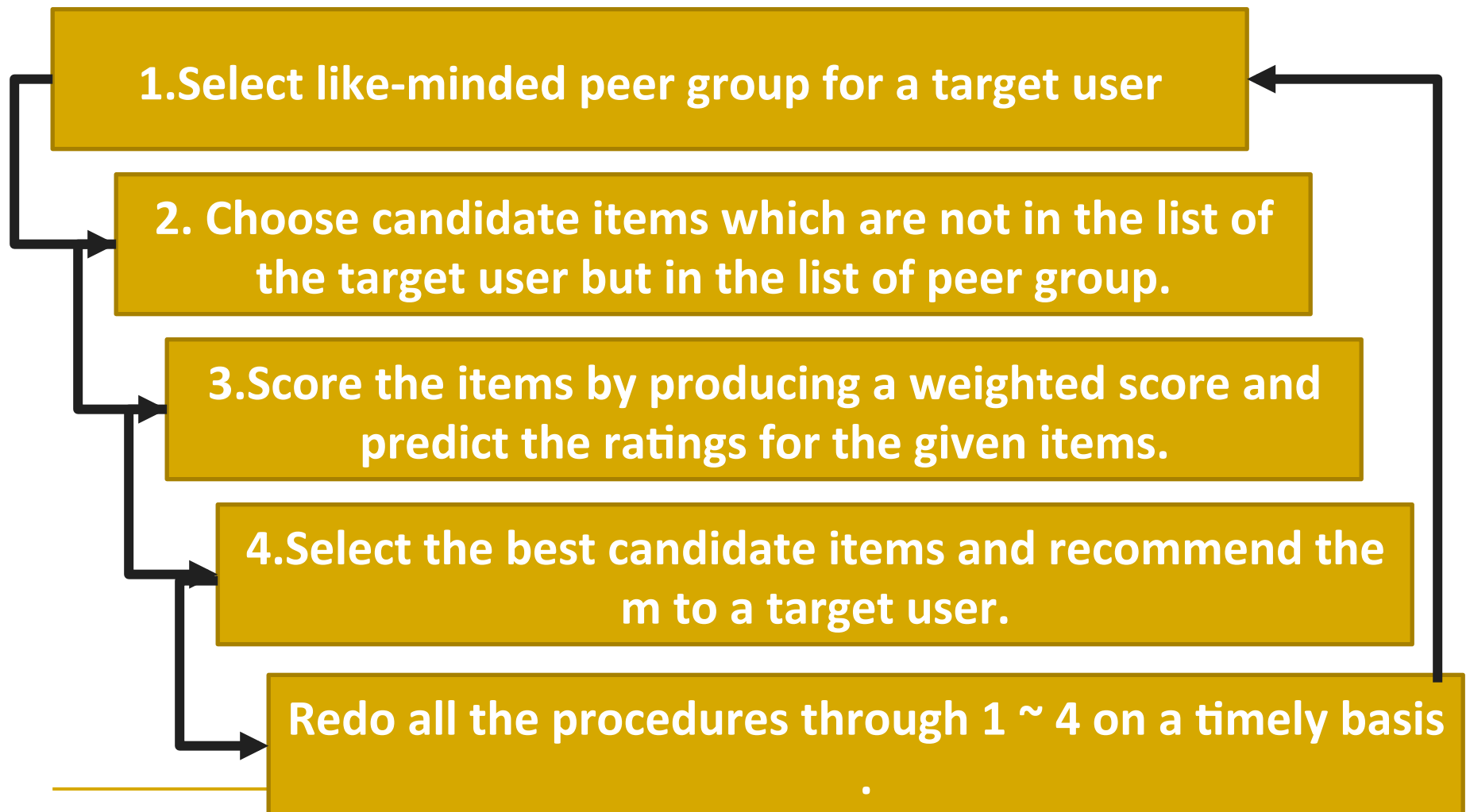
The input for the CF prediction algorithms is a matrix of users' ratings on items, referred as the **ratings matrix**.

Target User	Item 1	Item 2	Item 3	Item 4	Item 5	Average
Alice	5	3	4	4	???	16/4
User1	3	1	2	3	3	9/4
User2	4	3	4	3	5	14/4
User3	3	3	1	5	4	12/4
User4	1	5	5	2	1	13/4

User-based CF (2)



User-Based NN Recommendation



User-based NN: User Similarity

- Pearson's Correlation Coefficient for User a and User b for all rated Products, P .

$$sim(a, b) = \frac{\sum_{p \in product(P)} (r_{a, p} - \bar{r}_a)(r_{b, p} - \bar{r}_b)}{\sqrt{\sum_{p \in product(P)} (r_{a, p} - \bar{r}_a)^2} \sqrt{\sum_{p \in product(P)} (r_{b, p} - \bar{r}_b)^2}}$$

Average rating of user b

- Pearson correlation takes values from +1 (Perfectly positive correlation) to -1 (Perfectly negative correlation) .

User-based NN: Rating Prediction

$$\text{pred}(a, p) = \bar{r}_a + \frac{\sum_{b \in \text{neighbors}(n)} \text{sim}(a, b) \cdot (r_{b, p} - \bar{r}_b)}{\sum_{b \in \text{neighbors}(n)} \text{sim}(a, b)}$$

One Typical CF recommendation

movielens
helping you find the *right* movies

Welcome **suleehs@yahoo.co.kr** ([Log Out](#))
You've rated **10** movies.
You're the 26th visitor in the past hour.

★★★★★ = Must See
★★★★☆ = Will Enjoy
★★★☆☆ = It's OK
★★☆☆☆ = Fairly Bad
★☆☆☆☆ = Awful

So far you have rated **10** movies.
MovieLens needs at least **15** ratings from you to generate predictions for you.
Please rate as many movies as you can from the list below.

[next >](#)

Your Rating	Movie Information
★★★★☆ 4.5 stars	Butterfly Effect, The (2004) Drama, Sci-Fi, Thriller
??? Not seen	Chain Reaction (1996) Action, Adventure, Thriller
??? Not seen	High Noon (1952) Western
★★★★☆ 3.5 stars	Joy Luck Club, The (1993) Drama
??? Not seen	Matilda (1996) Children, Comedy, Fantasy
★★★★☆ 4.0 stars	Out of Africa (1985) Drama, Romance
??? Not seen	Scrooged (1988) Comedy, Fantasy, Romance
★★☆ 1.5 stars	Six Days Seven Nights (1998) Adventure, Comedy, Romance
??? Not seen	Sword in the Stone, The (1963) Animation, Children, Fantasy, Musical
★★★★☆ 4.0 stars	Working Girl (1988) Comedy, Drama, Romance

[next >](#)

One Typical CF recommendation

Tags Related to Your Search: [classic \(636\)](#), [70mm \(455\)](#), [action \(436\)](#), [Tumey's DVDs \(424\)](#), [comedy \(413\)](#), ([about tags](#))

Page 1 of 618 1 2 3 4 ... 618 next Skip to page #:

Predictions for you ↕	Your Ratings	Movie Information	Wish List
★★★★★	Not seen ▾	Bells of St. Mary's, The (1945) DVD VHS info imdb Drama [add tag] Popular tags: nuns [add tag]	<input type="checkbox"/>
★★★★★	Not seen ▾	Ship of Fools (1965) DVD info imdb Drama [add tag] Popular tags: talky [add tag] literate [add tag] Oscar (Best Cinematography) [add tag]	<input type="checkbox"/>
★★★★★	Not seen ▾	Random Harvest (1942) DVD VHS info imdb Drama, Romance [add tag] Popular tags: Quite Romantic [add tag] Friends Should See [add tag] made me cry [add tag]	<input type="checkbox"/>
★★★★★	Not seen ▾	In This Our Life (1942) info imdb add tag Drama	<input type="checkbox"/>
★★★★★	Not seen ▾	Revolution OS (2001) DVD info imdb Documentary [add tag] Popular tags: Open Source [add tag] own [add tag] Free Software Foundation [add tag]	<input type="checkbox"/>
★★★★★	Not seen ▾	Imaginary Heroes (2004) DVD info imdb add tag Comedy, Drama	<input type="checkbox"/>
★★★★★	Not seen ▾	Losing Chase (1996) info imdb add tag Drama	<input type="checkbox"/>
★★★★★	Not seen ▾	Seducing Doctor Lewis (Grande séduction, La) (2003) DVD VHS info imdb Comedy - French [add tag] Popular tags: to-rent [add tag] disk [add tag] TOO SEE [add tag]	<input type="checkbox"/>
★★★★★	Not seen ▾	Rose Tattoo, The (1955) DVD info imdb Drama, Romance [add tag] Popular tags: Oscar (Best Cinematography) [add tag] Oscar (Best Actress) [add tag]	<input type="checkbox"/>
★★★★★	Not seen ▾	Cry Freedom (1987) info imdb Drama [add tag] Popular tags: 70mm [add tag] Betamax [add tag] based on a true story [add tag]	<input type="checkbox"/>
★★★★★	Not seen ▾	Man from Earth, The (2007) info imdb Sci-Fi	<input type="checkbox"/>

Internet | Protected Mode: Off 100%

Benefits of Collaborative Filtering

- Collaborative filtering systems work by people in system, and it is expected that people to be better at evaluating information than a computed function
- CF doesn't require content analysis & extraction
- Independent of any machine-readable representation of the objects being recommended.
 - Works well for complex objects (or multimedia) such as music, pictures and movies
- More diverse and serendipitous recommendation

CF vs. CB

	CF	CB
Compare	Users interest	Item info
Similarity	Set of users User profile	Item info Text document
Shortcoming	Needs other users' feedback -> cold start Coverage Unusual interest	Feature matters Over-specialize Eliciting user feedback

Uses for CF : Domains

- Many items
 - Many ratings
 - Many more users than items recommended
 - Users rate multiple items
 - For each user of the community, there are other users with common needs or tastes
 - Item evaluation requires personal taste
 - Items persists
 - Taste persists
 - Items are homogenous
-

More on User-Based NN

- Other difference measures: Adjusted Cosine Similarity, Spearman's rank correlation coefficient
- Reduce the relative importance of the agreement on universally liked items : *inverse user frequency* (Breese, et al., 1998) and *variance weighting factor* (Herlocker, et al., 1999).
- Skewed neighboring is possible: *Significance weighting* (Herlocker, et al., 1999).
- Calculating a user's perfect neighborhood is immensely resource intensive calculations

Item-based NN Recommendation

Target User	Item 1	Item 2	Item 3	Item 4	Item 5	Average
Alice	5	3	4	4	???	4.0
User1	3	1	2	3	3	2.4
User2	4	3	4	3	5	3.8
User3	3	3	1	5	4	3.2
User4	1	5	5	2	1	2.8

Item-based Nearest Neighbor

	Item 1	Item 2	Item 3	Item 4	Item 5
Alice	1	-1	0	0	
User1	0.6	-1.4	-0.4	0.6	0.6
User2	0.2	-0.8	0.2	-0.8	1.2
User3	-0.2	-0.2	-2.2	1.8	0.8
User4	-1.8	2.2	2.2	-0.8	-1.8

Item-Based NN Recommendation

- Generate predictions based on similarities between items
 - Usually a cosine similarity used
- Prediction for a user u and item i is composed of a weighted sum of the user u 's ratings for items most similar to i .

$$pred(u, i) = \frac{\sum_{j \in ratedItems(u)} sim(i, j) \cdot r_{uj}}{\sum_{j \in ratedItems(u)} sim(i, j)}$$

Item-based Nearest Neighbor

- ▶ More computationally efficient than user-based nearest neighbors.
- ▶ Compared with user-based approach that is affected by the small change of users' ratings, item-based approach is more stable.
- ▶ Recommendation algorithm used by Amazon.com (Linden et al., 2003).

Uses for CF : User Tasks

- What tasks users may wish to accomplish
 - Help me find new items I might like
 - Advise me on a particular item
 - Help me find a user (or some users) I might like
 - Help our group find something new that we might like
 - Domain-specific tasks
 - Help me find an item, new or not
-

Uses for CF : System Tasks

- What CF systems support
 - Recommend items
 - Eg. Amazon.com
 - Predict rating for a given item
 - Constrained recommendations
 - Recommend best items from a set of items
-

Other Non-Probabilistic CF Algorithms

■ Association Rule Mining

- I.e., “If a customer purchases baby food then the customer also buys diapers in 70% of the cases.”
- Build Models based on commonly occurring patterns in the ratings matrix.
- “If user X liked both item 1 and item 2, then X will most probably also like item 5.”

$$\text{Support}(X \rightarrow Y) = \frac{\text{Number of Transactions containing } X \cup Y}{\text{Number of Transactions}}$$

$$\text{Confident}(X \rightarrow Y) = \frac{\text{Number of Transactions containing } X \cup Y}{\text{Number of Transactions containing } X}$$

Simple Probabilistic Algorithms

- Represent probability distributions
- Given a user u and a rated item i , the user assigned the item a rating of r : $p(r|u, i)$.

$$E(r | u, i) = \sum_r r \cdot p(r | u, i)$$

- Bayesian-network models, Expectation maximization (EM) algorithm
-

Dimensionality Reduction Algorithms

- Map item space to a smaller number of underlying “dimensions”
- Matrix Factorization/Latent Factor models:
 - Singular Value Decomposition,
 - Principal Component Analysis,
 - Latent Semantic Analysis, etc.
- Expensive offline computation and mathematical complexity
- Will be presented in a separate lecture

Dimensionality Reduction Algorithms

- Matrix Factorization got an attention since Netflix Prize competition.



Practical Issues : Ratings

- Explicit vs. Implicit ratings
 - Explicit ratings
 - Users rate themselves for an item
 - Most accurate descriptions of a user's preference
 - Challenging in collecting data
 - Implicit ratings
 - Observations of user behavior
 - Can be collected with little or no cost to user
 - Ratings inference may be imprecise.
-

Practical Issues : Ratings

- Rating Scales

- Scalar ratings

- Numerical scales
 - 1-5, 1-7, etc.

- Binary ratings

- Agree/Disagree, Good/Bad, etc.

- Unary ratings

- Good, Purchase, etc.
 - Absence of rating indicates no information
-

Practical Issues : Cold Start

- New user
 - Rate some initial items
 - Non-personalized recommendations
 - Describe tastes
 - Demographic info.
 - New Item
 - Non-CF : content analysis, metadata
 - Randomly selecting items
 - New Community
 - Provide rating incentives to subset of community
 - Initially generate non-CF recommendation
 - Start with other set of ratings from another source outside community
-

Evaluation Metrics

- Accuracy

- Predict accuracy

- The ability of a CF system to predict a user's rating for an item
 - Mean absolute error (MAE)
 - Classic, but now often criticized

- Rank accuracy

- Precision – percentage of items in a recommendation list that the user would rate as useful
 - Half-life utility – percentage of the maximum utility achieved by the ranked list in question
-

Evaluation Metrics

- Novelty
 - The ability of a CF system to recommend items that the user was not already aware of.
 - Serendipity
 - Users are given recommendations for items that they would not have seen given their existing channels of discovery.
 - Coverage
 - The percentage of the items known to the CF system for which the CF system can generate predictions.
-

Evaluation Metrics

- Learning Rate

- How quickly the CF system becomes an effective predictor of taste as data begins to arrive.

- Confidence

- Ability to evaluate the likely quality of its predictions.

- User Satisfaction

- By surveying the users or measuring retention and use statistics
-

Additional Issues : Interfaces

- Social Navigation

- Make the behavior of community visible
 - Leaving “footprints” : read-wear / edit-wear
 - Attempt to mimic more accurately the social process of word-of-mouth recommendations
 - Epinions.com
-

Additional Issues : Interfaces

Epinions.com (<http://www.epinions.com>)

Compare Prices and Read Reviews on Canon PowerShot SD400 / IXUS 50 Digital Camera at Epinions.c - Microsoft Internet Explorer

Address: http://www.epinions.com/Canon_PowerShot_SD400_Digital_Camera

Canon PowerShot SD400 / IXUS 50 Digital Camera

Overall rating: ★★★★★
Reviewed by 14 Epinions users

Ease of Use: ██████████
Durability: ██████████
Battery Life: ██████████
Photo Quality: ██████████
Shutter Lag: ██████████

[Compare Prices](#)
[View Details](#)
[Read Reviews](#)

[Write a Review](#)

[Add to wish list](#)
[Subscribe to reviews on this product](#)

Read Reviews

Showing 1-12 of 12 reviews

Sort by: [Product Rating](#) | [Review Date](#)

Product Rating: ★★★★★
Ease of Use: ██████████
Durability: ██████████
Battery Life: ██████████
Photo Quality: ██████████
Shutter Lag: ██████████

Canon's New SD400, best balance of performance, features, and cost of any current Micro-Cam
by [Howard_Creech](#) **LEAD** in Electronics, Apr 25 '05
(A Very Helpful review)
Pros: Fast, ultra compact, user friendly, very good image quality, tough stainless steel body
Cons: Mediocre battery life, weak flash, red-eye problems, chromatic aberration, and noisy ISO 400 images
Canon has dominated the ultra compact digicam market since the first Digital Elphs hit camera store shelves and the new Canon Powershot SD400 demonstrates why that trend isn't likely to change anytime soon. The SD400 provides users with 5-megapixel ...
[Read the full review](#)

Canon PowerShot SD400 / IXUS 50 Digital Camera - Cool, Compact and Fast
by [dkozin](#) **LEAD** in Electronics, Aug 08 '05
(A Very Helpful review)
Pros: Cool looks, sturdy, compact, fast operation, excellent picture quality, large bright LCD
Cons: No manual control, flash leaves shadow in corner upclose in macro mode, no battery status
I have bought the Canon PowerShot SD400 Digital Elph for only \$270 using coupons after it was advertised in an Office Depot catalog for a price that was probably a mistake. The camera is no longer available at this ...
[Read the full review](#)

Good digital camera with a little bit issue
by [unchinois](#), Sep 08 '05
(A Helpful review)
Pros: Easy of use, size, good movie option
Cons: We had issues on the zooming for a few weeks.
I have purchased the Canon PowerShot SD400 as a gift to my uncle who knows practically nothing about computers. So far, he has been pretty happy about it overall except the issue that I am going to talk about later on in the review. Size You ...

[http://www.epinions.com/member/community_lists.html/show/~1](#)

Internet

20% off!
Expires 09/21/05.

Being in Just Got Better!
Plans Starting at **\$39.99**
monthly access
ANNUAL AGREEMENT REQUIRED.

JOIN IN

verizon wireless

Limited Time Offer!
cingular
Authorized Dealer
FREE
DATA

Additional Issues : Interfaces

- Explanation

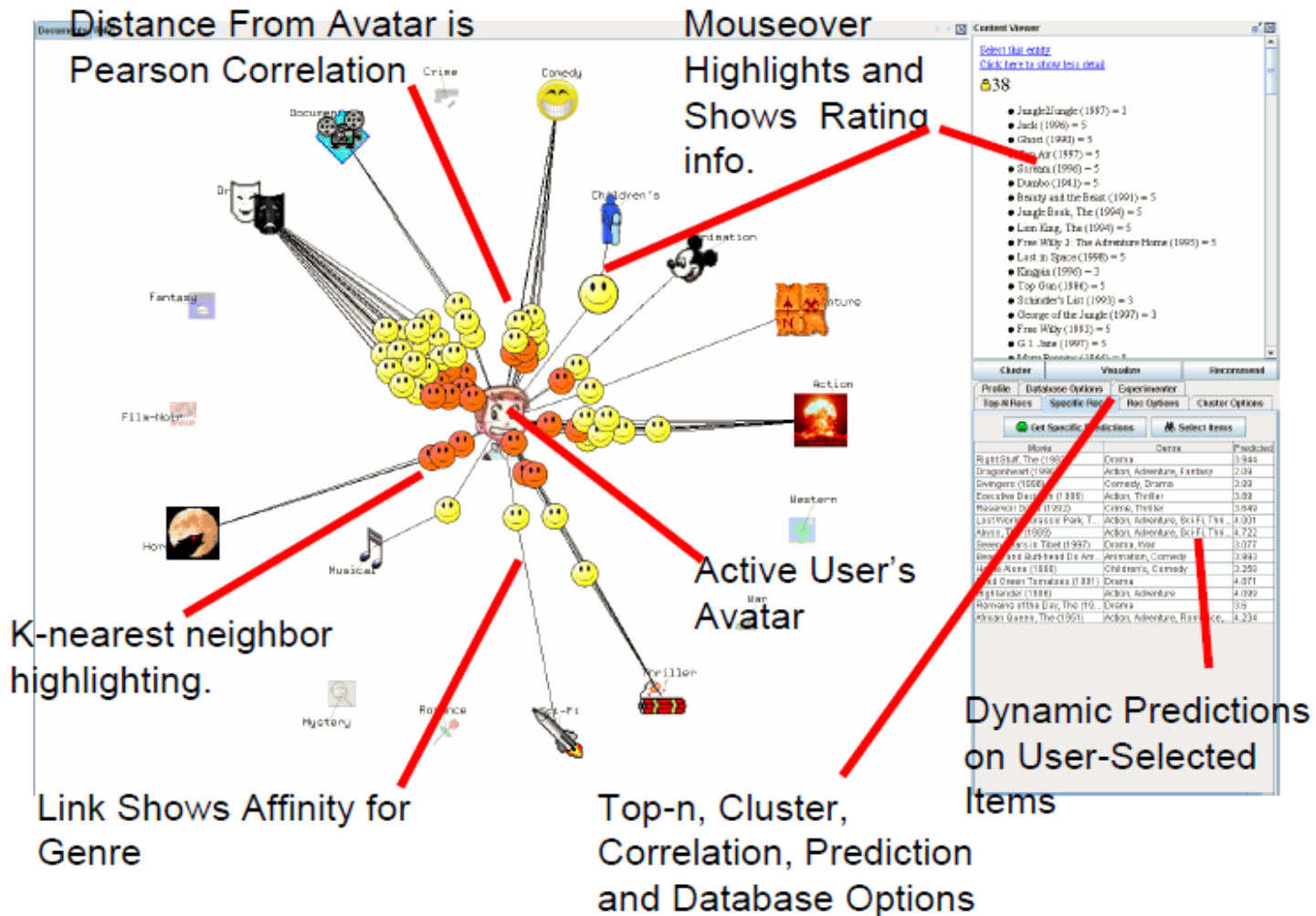
- Where, how, from whom the recommendations are generated.
 - Do not make it too much!
 - Not showing reasoning process
 - Graphs, key items
 - Reviews
-

Additional Issues : Privacy & Trust

- User profiles
 - Personalized information
 - Distributed architecture

 - Recommender system may break trust when malicious users give ratings that are not representative of their true preferences.
-

Choose your Peers



- PeerChooser (CHI 2008) John O'Donovan and Barry Smyth (UCD)

The Concept of *Controllability*

GoodReads: Book recommender system

Here are some book recommendations...

Recommendations of books

Your interest in **Classics** updated: 1 minute ago

Options of User Control

Recommendations by Shelf

read	50
Tip: Create a new bookshelf of related books that you've enjoyed or want to read. You'll get recommendations based on those books!	
add	
hide	
Classics	50
Contemporary	15
Historical Fiction	6
Philosophy	33
Poetry	3
Psychology	29

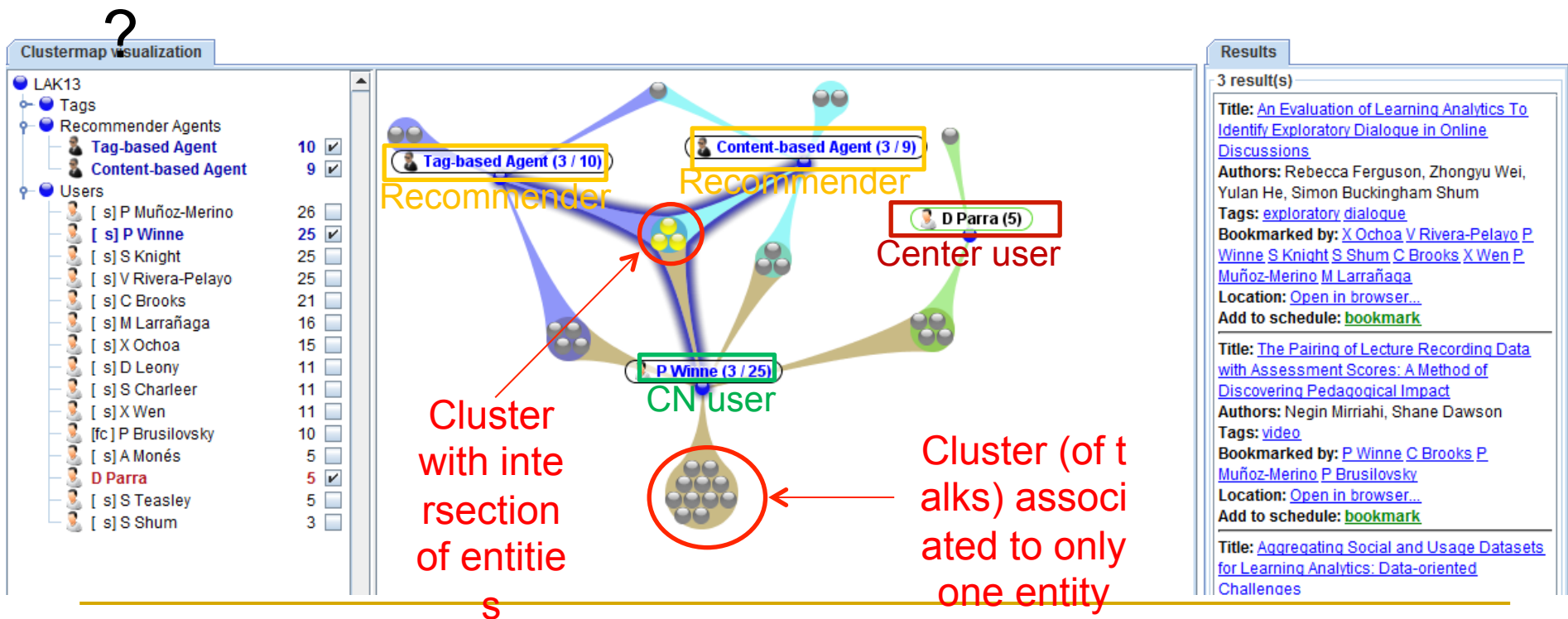
Explainability

Your interest in **Contemporary** updated: 1 minute ago

BRAIN TRAINING GAMES
Improve your memory

TalkExplorer

- Adaptation of Aduna Visualization in CN
- Main research question: Do fusion (intersection) of contexts of relevance improve user experience?



SetFusion

Tune weights of the recommender methods: **b)**

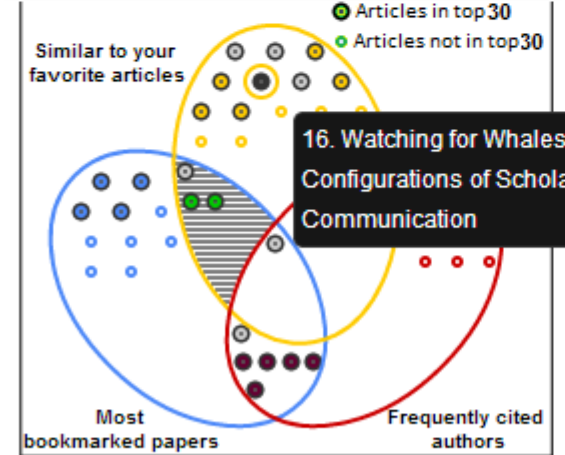
Most bookmarked papers

Similar to your favorite articles

Frequently cited authors in ACM DL

[Update Recommendation List →](#)

* Hover over circles to explore articles
* Click on the diagram to highlight subsets **c)**





Articles in top30
Articles not in top30


16. Watching for Whales: Emerging Configurations of Scholarly Communication

Most bookmarked papers Frequently cited authors

a)

[3. When thumbnails are and are not enough: Factors behind users](#)  [\[see abstract\]](#)
by Dan Albertson

[8. Two Paths to Motivation through Game Design Elements: Reward-Based Gamification and Meaningful Gamification](#)  [\[see abstract\]](#)
by Scott Nicholson

[17. Ebooks and cross generational perceived privacy issues](#)  [\[see abstract\]](#)
by Renee Kapusniak, Jennifer Sue Thiele

a) Recommended Talks, b) Controllable Sliders, c) Venn Diagram to Control (click) & Inspect (hover)

Additional Issues :

Hybrid Approach

- CF + CB
 - Content based system
 - Maintain user profile based on content analysis
 - Collaborative system
 - Directly compare profiles to determine similar users for recommendation
 - Fab system
-

Additional Issues :

Hybrid Approach

Example : Fab System Architecture

