

CMSC5733 Social Computing

Irwin King

The Chinese University of Hong Kong

king@cse.cuhk.edu.hk

©2013-2015 All Rights Reserved.



Information and more Information!

flickr™



amazon.com.



You Tube



ebay

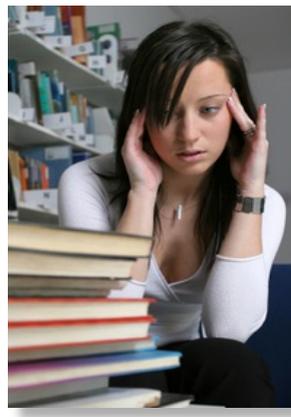
facebook

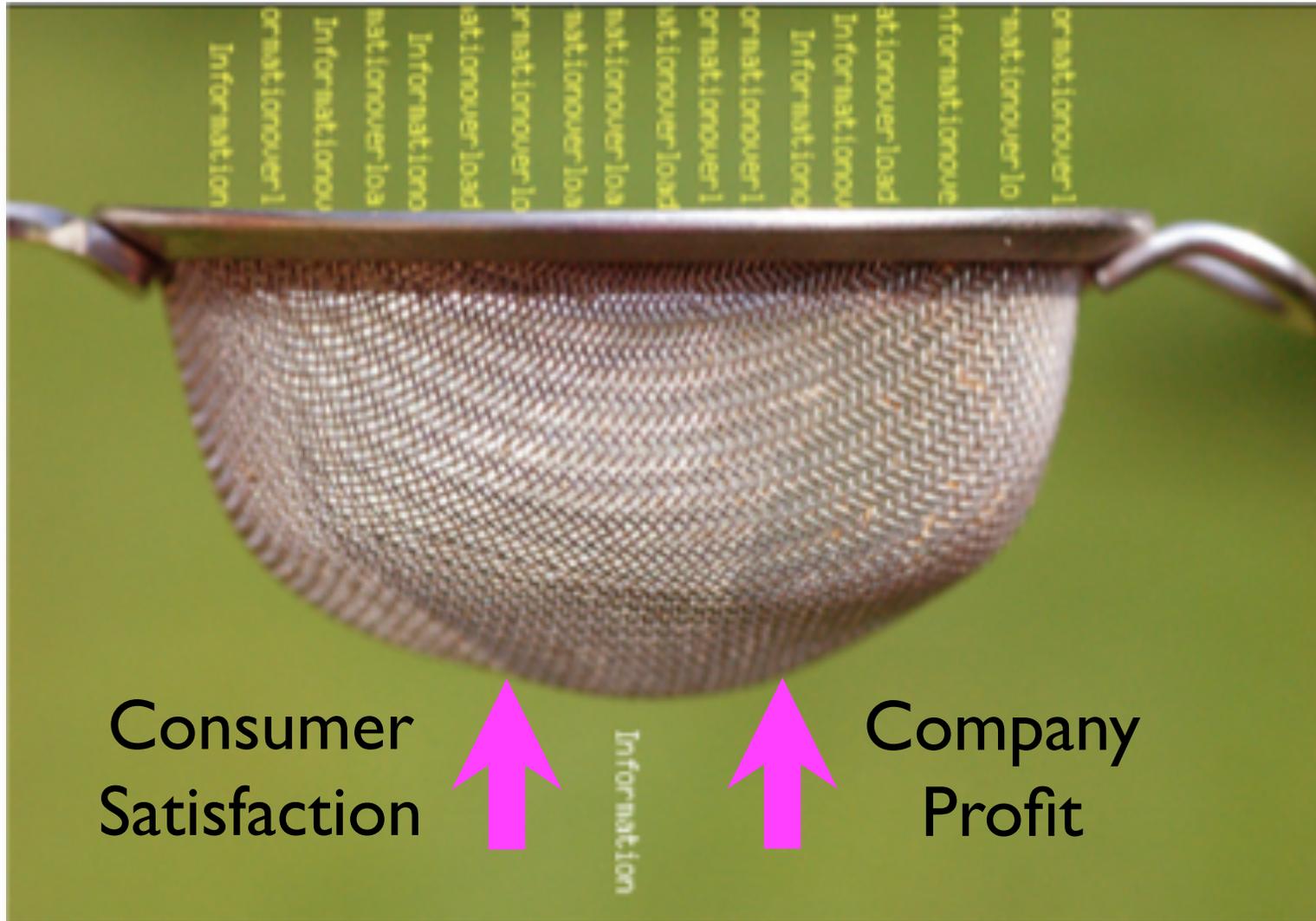
hulu™

twitter



Information Overload







Real Life Examples

Amazon.com: Social Computing, Behavioral Modeling, and Prediction: Huan Liu, John J. Salerno, Michael J. Young: Books

http://www.amazon.com/Social-Computing-Behavioral-Modeling-Prediction/ amazon

Apple Yahoo! Google Maps YouTube Wikipedia News (26) Popular

Amazon.com: Social Comp...

Click to **LOOK INSIDE!**

Social Computing, Behavioral Modeling, and Prediction (Hardcover)
by [Huan Liu](#) (Editor), [John J. Salerno](#) (Editor), [Michael J. Young](#) (Editor)
Key Phrases: [social network analysis](#), [electronic institutions](#), [cognitive modeling](#), [New York](#), [Cambridge University Press](#), [Virtual World](#) (more...)
No customer reviews yet. [Be the first.](#)

List Price: \$129.00
Price: **\$129.00** & this item ships for **FREE with Super Saver Shipping.** [Details](#)

In Stock.
Ships from and sold by **Amazon.com**. Gift-wrap available.

Want it delivered Thursday, February 19? Order it in the next 14 hours and 52 minutes, and choose **One-Day Shipping** at checkout. [Details](#)
12 new from \$95.29 **4 used** from \$103.33

[Share your own customer images](#)
[Search inside this book](#)

Tell the Publisher!
[I'd like to read this book on Kindle](#)
Don't have a Kindle? [Get yours here.](#)

Quantity: 1

[Add to Shopping Cart](#)
or
[Sign in](#) to turn on 1-Click order
or
[Add to Cart with FREE Two-Day Shipping](#)
Amazon Prime Free Trial required. Sign up when you check out. [Learn More](#)

More Buying Choices
16 used & new from \$95.29
Have one to sell? [Sell yours here](#)

[Add to Wish List](#)
[Add to Shopping List](#)
[Add to Wedding Registry](#)
[Add to Baby Registry](#)
[Share with Friends](#)

Frequently Bought Together
Customers buy this book with [Social Network Analysis: A Handbook](#) by John P Scott
Price For Both: \$174.85



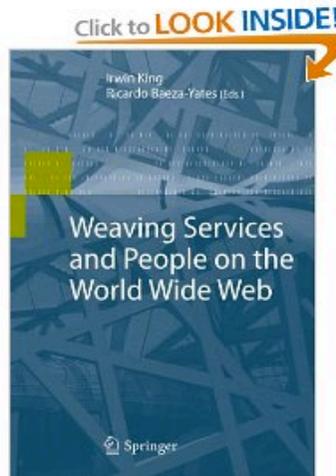
Real Life Examples

amazon.com Hello. [Sign in](#) to get personalized recommendations. New customer? [Start here](#). **FREE 2-Day Shipping: See details**

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

Shop All Departments

Books | [Advanced Search](#) | [Browse Subjects](#) | [New Releases](#) | [Bestsellers](#) | [The New York Times® Bestsellers](#) | [Libros en español](#) | [Bargain Books](#) | [Textbooks](#)



[Share your own customer images](#)
[Search inside this book](#)

Weaving Services and People on the World Wide Web [Hardcover]

[Irwin King](#) (Editor), [Ricardo Baeza-Yates](#) (Editor)

[Be the first to review this item](#) | Like (0)

List Price: ~~\$99.00~~

Price: **\$79.11** & this item ships for **FREE with Super Saver Shipping**. [Details](#)

You Save: **\$19.89 (20%)**

In Stock.

Ships from and sold by **Amazon.com**. Gift-wrap available.

Only 2 left in stock--order soon (more on the way).

Want it delivered Thursday, July 21? Order it in the next 0 hours and 19 minutes, and choose **One-Day Shipping** at checkout. [Details](#)

18 new from \$14.62 **13 used** from \$14.62



FREE Two-Day Shipping for Students. [Learn more](#)

Quantity:

or

[Sign in](#) to turn on 1-Click ordering.

or

Amazon Prime Free Trial required. Sign up when you check out. [Learn More](#)

More Buying Choices

31 used & new from \$14.62

Have one to sell?

[Share](#)

Customers Who Bought This Item Also Bought



Real Life Examples

The screenshot shows an Amazon.com product page for the book "Social Computing, Behavioral Modeling, and Prediction" by Huan Liu, John J. Salerno, and Michael J. Young. The browser address bar shows the URL: <http://www.amazon.com/Social-Computing-Behavioral-Modeling-Prediction/>. The page features a "Frequently Bought Together" section with a price of \$174.85 for two books. Below this, the section "Customers Who Bought This Item Also Bought" is circled in red. It displays three books: "Predictably Irrational: Hidden Forces That... by Dan Ariely" (4.5 stars, 18 reviews, \$16.34), "Complex Systems: ... by ...", and "Generative Social Science: Studies in Agent-Bas... by Joshua M. Epstein" (5 stars, 5 reviews, \$42.00). A blue box with a star rating scale is overlaid on the page, with the text "Five scales rating" above it. The scale consists of five rows of stars, with the number of stars in each row decreasing from left to right. The corresponding text for each row is: "I hate it", "I don't like it", "It's ok", "I like it", and "I love it". The "Predictably Irrational" book's star rating is also circled in red. The page includes standard Amazon navigation elements like "Add to Cart", "Add to Wedding Registry", and "Share with Friends".



Real Life Examples

amazon.com | Hao's Amazon.com | See All 40 Product Categories | Your Account |  Cart | Your Lists  | Help | 

Your Browsing History | Recommended For You | Rate These Items | Improve Your Recommendations | Your Profile | Learn More

Today's Recommendations For You

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

Page 1 of 25



[Invincible](#) ~ Michael Jackson
★★★★☆ (880) \$7.99



[In Search of Sunrise, Vol. 7: Asia](#) ~ DJ Tiesto
★★★★☆ (53) \$15.99



[Fallen](#) ~ Evanescence
★★★★☆ (2,447) \$8.99



[Amar Es Combatir](#) ~ Maná
★★★★☆ (55) \$8.49



Real Life Examples



My Movies: **gabe_ma** [Edit Profile](#)

Recommendations For You

[Receive Recommendations by Email](#)

Movies in Theaters: 94089



Burn After Reading (R)

[Showtimes & Tickets](#) | [Add to My Lists](#)

Yahoo! Users: **B-** 4794 ratings
The Critics: **B** 14 reviews

Don't Recommend Again Seen It? Rate It!

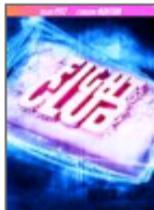


Pride and Glory (R)

[Showtimes & Tickets](#) | [Add to My Lists](#)

Yahoo! Users: **A-** 59 ratings
The Critics: **C+** 6 reviews

Don't Recommend Again Seen It? Rate It!

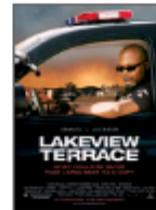


Fight Club (R)

[Showtimes & Tickets](#) | [Add to My Lists](#)

Yahoo! Users: **B+** 52392 ratings
The Critics: **B** 12 reviews

Don't Recommend Again Seen It? Rate It!



Lakeview Terrace (PG-13)

[Showtimes & Tickets](#) | [Add to My Lists](#)

Yahoo! Users: **B** 3229 ratings
The Critics: **C** 12 reviews

Don't Recommend Again Seen It? Rate It!

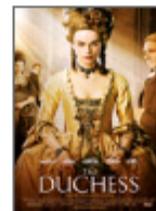


Vicky Cristina Barcelona (PG-13)

[Showtimes & Tickets](#) | [Add to My Lists](#)

Yahoo! Users: **B** 1923 ratings
The Critics: **B+** 13 reviews

Don't Recommend Again Seen It? Rate It!



The Duchess (PG-13)

[Showtimes & Tickets](#) | [Add to My Lists](#)

Yahoo! Users: **B+** 953 ratings
The Critics: **B-** 10 reviews

Don't Recommend Again Seen It? Rate It!

[See All Recommendations](#)



Real Life Examples



Real Life Examples

iLike™

Songs from friends and similar people

[▶ Play All](#)  | [Buy all](#) 

-  [▶ Victims](#) by [The Oppressed](#)
New! Traditional Byrd69
-  [▶ Skinhead Girl](#) by [The Oppressed](#)
New! Traditional Byrd69
-  [▶ King Of The Jungle](#) by [Last Resort](#)
New! Traditional Byrd69
-  [▶ Violence In Our Minds](#) by [Last Resort](#)
New! Traditional Byrd69
-  [▶ Violence](#) by [The Templars](#)
New! Traditional Byrd69

[View all](#) | [invite more friends](#)



5-scale Ratings

Search for items to rate

Music

Enrique

GO!

Search results for **Enrique** in Music:

1.



Escape

~ Enrique Iglesias

Your tags:

Add

(What's this?)

Saved

X|★★★★★

I Own It

2.



Enrique

~ Enrique Iglesias

Your tags:

Add

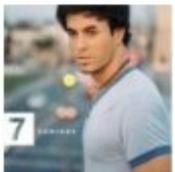
(What's this?)

Saved

X|★★★★★

I Own It

3.



Seven

~ Enrique Iglesias

Your tags:

Add

(What's this?)

Saved

X|★★★★★

I Own It



5-scale Ratings

Sign In

What is your e-mail address?

My e-mail address is

Do you have an Amazon.com password?

No, I am a new customer.

Yes, I have a password:

[Sign in using our secure server](#)

amazon.com | Hao's Amazon.com | See All 40 Product Categories

Your Browsing History | Recommended For You | **Rate These Items**

Search for items to rate

1 Use the search box above to find your favorite books, movies, albums, artists, authors and brands.

2 Tell us what you think of the items we return for your search by rating the item or telling us you already own them.



3 Repeat until the Recommendations you find in Your Amazon.com reflect your tastes and interests.

Search for items to rate

1 Use the search box above to find your favorite books, movies, albums, artists, authors and brands.

2 Tell us what you think of the items we return for your search by rating the item or telling us you already own them.



3 Repeat until the Recommendations you find in Your Amazon.com reflect your tastes and interests.



On The Menu

- Introduction
- Basic Techniques
 - Collaborative filtering
 - Matrix factorization
- Different Models
 - Social graph
 - Social ensemble
 - Social distrust
 - Website recommendation



Basic Approaches

- Content-based Filtering
 - Recommend items based on **key-words**
 - More appropriate for information retrieval
- Collaborative Filtering (CF)
 - Look at users with similar rating styles
 - Look at similar items for each item

Underling assumption: personal tastes are correlated--
Active users will prefer those items which the
similar users prefer!



Framework

Items

	i_1	i_2			i_j							i_m
u_1												
u_2	1	3		4		2		5			3	4
u_i		3		4		r_{ij}	3	4		3	4	4
u_n	1			3	5	2		4	1			3

Users

• The tasks

- Find the **unknown** rating!
- Which **item(s)** should be recommended?

The Chinese University of Hong Kong, CMSC5733 Social Computing, Irwin King

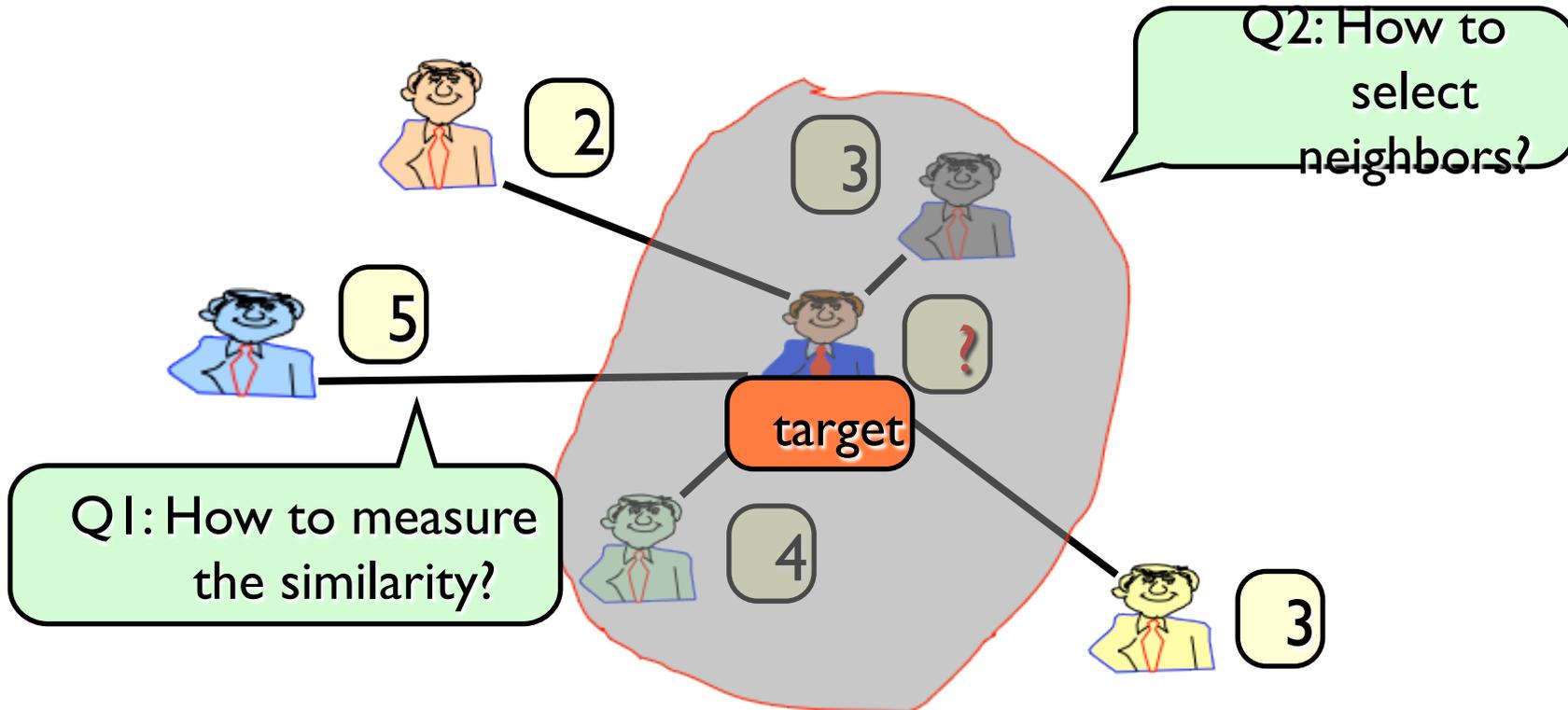


Collaborative Filtering

- Memory-based (Neighborhood-based)
 - User-based
 - Item-based
- Model-based
 - Clustering Methods
 - Bayesian Methods
 - Matrix Factorization
 - etc.



User-User Similarity



User-based Collaborative Filtering

Items

Users

u1												
u2	1	3		4		2		5			3	4
u3												
u4		3		4			3	4		3	4	4
u5												
u6	1			3	5	2		4	1			3



User-based Collaborative Filtering

Items

Users

u1													
u2	1	3		4		2		5			3	4	
u3													
u4		3		4			3	4		3	4		4
u5													
u6	1			3	5	2		4	1			3	



User-based Collaborative Filtering

Items

Users

u1												
u2	1	3	4	2	5			3	4			
u3												
u4		3	4		3	4		3	4		4	
u5												
u6	1		3	5	2	4	1			3		



User-based Collaborative Filtering

Items

u ₁													
u ₂	1	3		4		2		5			3	4	
u ₃													
u ₄		3		4			3	4		3	4		4
u ₅													
u ₆	1			3	5	2		4	1			3	

Users



User-based Collaborative Filtering

Items

Users

u1												
u2	1	3	4		2		5			3	4	
u3												
u4		3	4			3	4		3	4		4
u5												
u6	1		3	5	2		4	1			3	



User-based Collaborative Filtering

- Predict the ratings of active users based on the ratings of similar users found in the user-item matrix
- Pearson correlation coefficient

$$w(a, i) = \frac{\sum_j (r_{aj} - \bar{r}_a)(r_{ij} - \bar{r}_i)}{\sqrt{\sum_j (r_{aj} - \bar{r}_a)^2 \sum_j (r_{ij} - \bar{r}_i)^2}} \quad j \in I(a) \cap I(i)$$

- Cosine measure

$$c(a, i) = \frac{r_a \cdot r_i}{\|r_a\|_2 * \|r_i\|_2}$$

u_i	1	3	4	2	5		3	4		
u_a	3	4		3	4	3	4	4		
	1		3	5	2		4	1		3



Nearest Neighbor Approaches

[Sarwar, 00a]

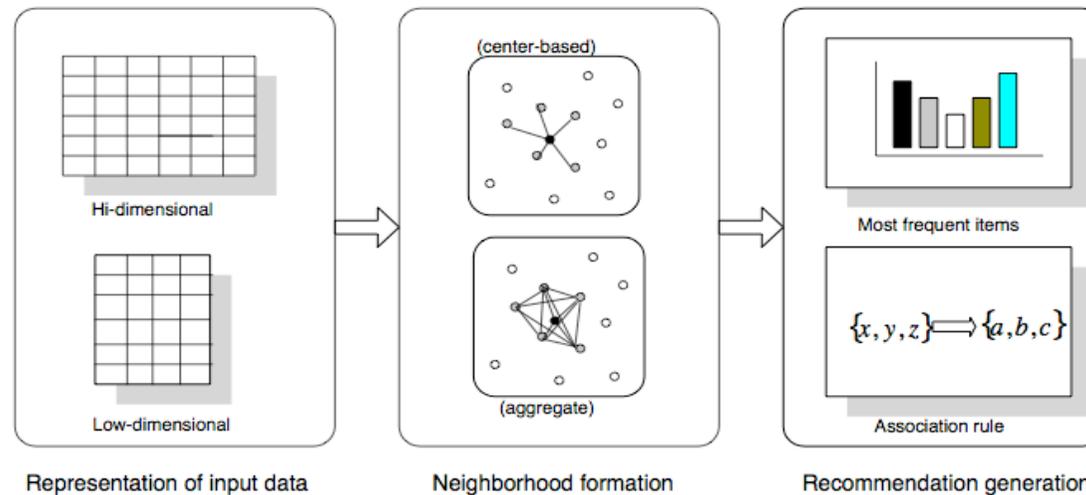


Figure 1: Three main parts of a Recommender System.

- Identify highly similar users to the active one
- All with a measure greater than a threshold

- Best K ones

$$r_{aj} = \bar{r}_a + \frac{\sum_i w(a, i)(r_{ij} - \bar{r}_i)}{\sum_i w(a, i)}$$

- Prediction

The Chinese University of Hong Kong, CMSC5733 Social Computing, Irwin King



Collaborative Filtering

- **Memory-based Method (Simple)**
 - User-based Method [Xue et al., SIGIR '05]
 - Item-based [Deshpande et al., TOIS '04]
- **Model-based (Robust)**
 - Clustering Methods [Hkors et al, CIMCA '99]
 - Bayesian Methods [Chien et al., IWAIIS '99]
 - Aspect Method [Hofmann, SIFIR '03]
 - Matrix Factorization [Sarwar et al., WWW '01]



Collaborative Filtering

- Memory-based (Neighborhood-based)
 - User-based
 - Item-based
- Model-based
 - Clustering Methods
 - Bayesian Methods
 - Matrix Factorization
 - etc.



Item-Item Similarity

- Search for similarities among items
- Item-Item similarity is more stable than user-user similarity



Correlation-based Method

[Sarwar, 2001]

- Same as in user-user similarity but on item vectors
- Pearson correlation coefficient
- Look for users who rated both items

$$s_{ij} = \frac{\sum_u (r_{uj} - \bar{r}_j)(r_{ui} - \bar{r}_i)}{\sqrt{\sum_u (r_{uj} - \bar{r}_j)^2 \sum_u (r_{ui} - \bar{r}_i)^2}}$$

- u : users rated both items

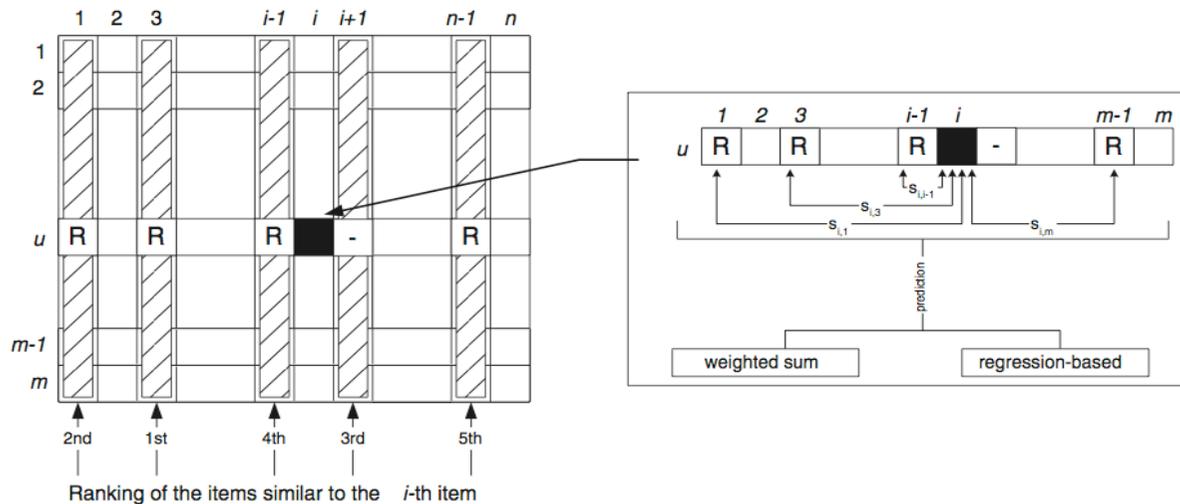
	i_1	i_2		i_i		i_j				i_m	
u_1											
u_2	1	3		4	2	5			3	4	
u_i		3		4		3	4		3	4	4
u_n	1			3	5	2	4	1		3	



Correlation-based Method

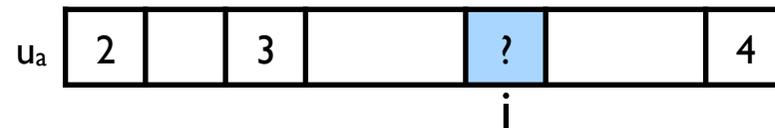
[Sarwar, 2001]

- Calculate item similarity, then determine its **k-most similar items**



- Predict rating for a given user-item pair as a **weighted sum** over **similar items** that he **rated**

$$r_{ai} = \frac{\sum_j s_{ij} r_{aj}}{\sum_j s_{ij}}$$



Collaborative Filtering

- Memory-based (Neighborhood-based)
 - User-based
 - Item-based
- Model-based
 - Clustering Methods
 - Bayesian Methods
 - **Matrix Factorization**
 - etc...



Matrix Factorization

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
u_1	5	2		3		4		
u_2	4	3			5			
u_3	4		2				2	4
u_4								
u_5	5	1	2		4	3		
u_6	4	3		2	4		3	5

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
u_1	5	2	2.5	3	4.8	4	2.2	4.8
u_2	4	3	2.4	2.9	5	4.1	2.6	4.7
u_3	4	1.7	2	3.2	3.9	3.0	2	4
u_4	4.8	2.1	2.7	2.6	4.7	3.8	2.4	4.9
u_5	5	1	2	3.4	4	3	1.5	4.6
u_6	4	3	2.9	2	4	3.4	3	5

$$U = \begin{bmatrix} 1.55 & 1.22 & 0.37 & 0.81 & 0.62 & -0.01 \\ 0.36 & 0.91 & 1.21 & 0.39 & 1.10 & 0.25 \\ 0.59 & 0.20 & 0.14 & 0.83 & 0.27 & 1.51 \\ 0.39 & 1.33 & -0.43 & 0.70 & -0.90 & 0.68 \\ 1.05 & 0.11 & 0.17 & 1.18 & 1.81 & 0.40 \end{bmatrix}$$

$$V = \begin{bmatrix} 1.00 & -0.05 & -0.24 & 0.26 & 1.28 & 0.54 & -0.31 & 0.52 \\ 0.19 & -0.86 & -0.72 & 0.05 & 0.68 & 0.02 & -0.61 & 0.70 \\ 0.49 & 0.09 & -0.05 & -0.62 & 0.12 & 0.08 & 0.02 & 1.60 \\ -0.40 & 0.70 & 0.27 & -0.27 & 0.99 & 0.44 & 0.39 & 0.74 \\ 1.49 & -1.00 & 0.06 & 0.05 & 0.23 & 0.01 & -0.36 & 0.80 \end{bmatrix}$$



Matrix Factorization

- Matrix Factorization in Collaborative Filtering
 - To fit the product of two (low rank) matrices to the observed rating matrix.
 - To find two latent user and item feature matrices.
 - To use the fitted matrix to predict the unobserved ratings.

$$Y \approx UV = \begin{pmatrix} \mathbf{u}_{11} & \cdots & \mathbf{u}_{1k} \\ \vdots & \ddots & \vdots \\ \mathbf{u}_{m1} & \cdots & \mathbf{u}_{mk} \end{pmatrix} \begin{pmatrix} \mathbf{v}_{11} & \cdots & \mathbf{v}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{v}_{k1} & \cdots & \mathbf{v}_{kn} \end{pmatrix}$$

Diagram illustrating the matrix factorization equation $Y \approx UV$. The matrix U is labeled "User-specific latent feature vector" and the matrix V is labeled "Item-specific latent feature column vector".



Matrix Factorization

- Optimization Problem
- Given a $m \times n$ rating matrix R , to find two matrices $U \in \mathbb{R}^{l \times m}$ and $V \in \mathbb{R}^{l \times n}$

$$R \approx U^T V,$$

where $l < \min(m, n)$, is the number of factors



Matrix Factorization

- Models
 - SVD-like Algorithm
 - Regularized Matrix Factorization (RMF)
 - Probabilistic Matrix Factorization (PMF)
 - Non-negative Matrix Factorization (NMF)



SVD-like Algorithm

- Minimizing

$$\frac{1}{2} \|R - U^T V\|_F^2,$$

- For collaborative filtering

$$\min_{U, V} \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij} (R_{ij} - U_i^T V_j)^2$$

where I_{ij} is the indicator function that is equal to 1 if user u_i rated item v_j and equal to 0 otherwise.



Regularized Matrix Factorization

- Minimize the loss based on the observed ratings with regularization terms to avoid over-fitting problem

$$\min_{U,V} \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij} (R_{ij} - U_i^T V_j)^2 + \frac{\lambda_1}{2} \|U\|_F^2 + \frac{\lambda_2}{2} \|V\|_F^2$$

Regularization terms

where $\lambda_1, \lambda_2 > 0$

- The problem can be solved by simple gradient descent algorithm.



Regularized Matrix Factorization

- Algorithm for RMF
 - Not convex & local optimal
 - Gradient-descent algorithm
 - Gradient computation with randomly initialized U and V

$$\frac{\partial L}{\partial \mathbf{u}_{il}} = \lambda \mathbf{u}_{il} - \sum_{j|(i,j) \in S} (y_{ij} - \widehat{y}_{ij}) \mathbf{v}_{jl}$$
$$\frac{\partial L}{\partial \mathbf{v}_{jl}} = \lambda \mathbf{v}_{jl} - \sum_{i|(i,j) \in S} (y_{ij} - \widehat{y}_{ij}) \mathbf{u}_{il}$$

- Update U and V alternatively

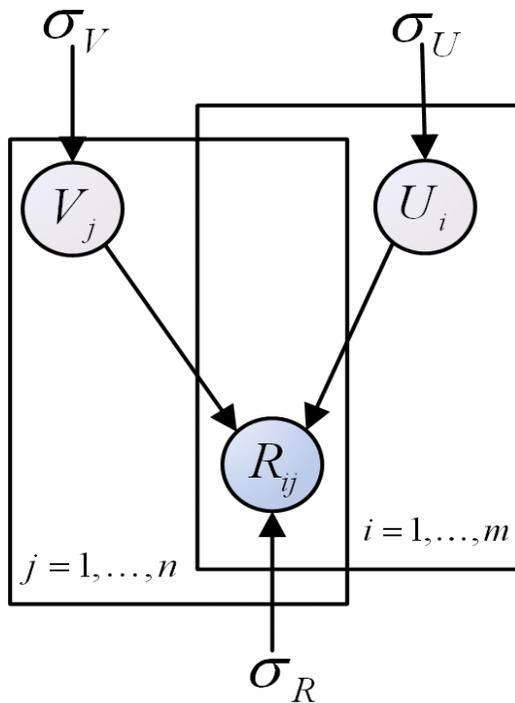
$$\mathbf{u}_{il}^{(t+1)} = \mathbf{u}_{il}^{(t)} - \tau \frac{\partial L}{\partial \mathbf{u}_{il}^{(t)}}$$
$$\mathbf{v}_{jl}^{(t+1)} = \mathbf{v}_{jl}^{(t)} - \tau \frac{\partial L}{\partial \mathbf{v}_{jl}^{(t)}}$$

τ is the step size of gradient decent.



Regularized Matrix Factorization

- PMF
 - Define a conditional distribution over the observed ratings as:

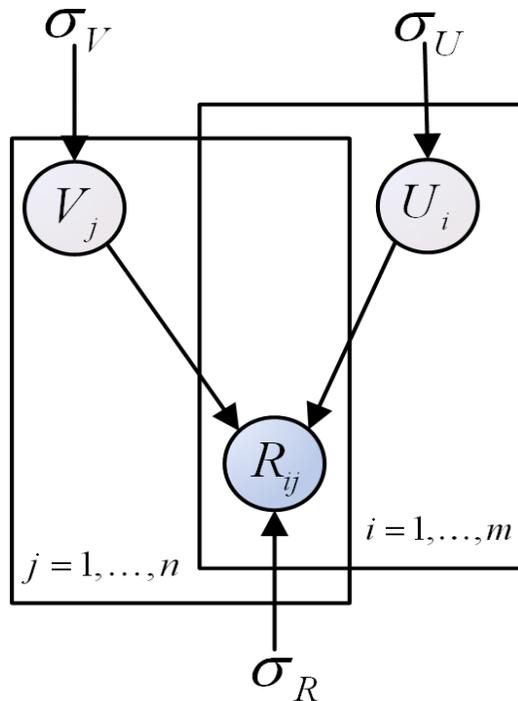


$$p(R|U, V, \sigma_R^2) = \prod_{i=1}^m \prod_{j=1}^n \left[\mathcal{N} \left(R_{ij} | g(U_i^T V_j), \sigma_R^2 \right) \right]^{I_{ij}^R}$$



Regularized Matrix Factorization

- PMF
 - Assume zero-mean spherical Gaussian priors on user and item feature:



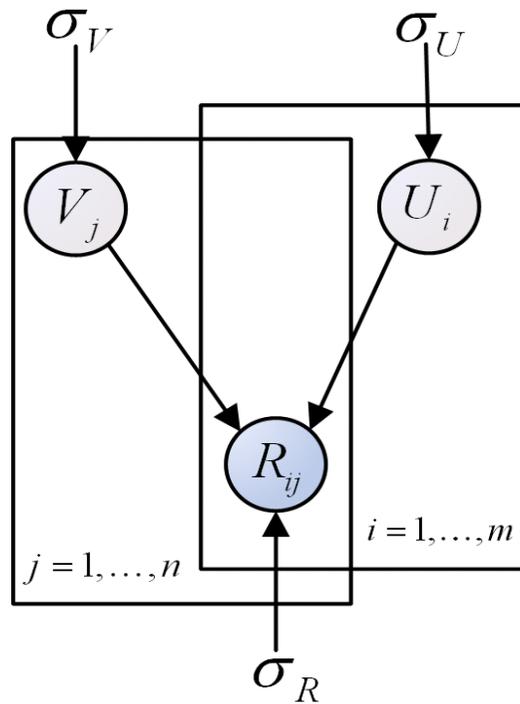
$$p(U|\sigma_U^2) = \prod_{i=1}^m \mathcal{N}(U_i|0, \sigma_U^2 \mathbf{I})$$

$$p(V|\sigma_V^2) = \prod_{j=1}^n \mathcal{N}(V_j|0, \sigma_V^2 \mathbf{I})$$



Regularized Matrix Factorization

- PMF
 - Bayesian inference



$$\begin{aligned}
 p(U, V | R, \sigma_R^2, \sigma_U^2, \sigma_V^2) &\propto p(R | U, V, \sigma_R^2) p(U | \sigma_U^2) p(V | \sigma_V^2) \\
 &= \prod_{i=1}^m \prod_{j=1}^n \left[\mathcal{N} \left(R_{ij} | g(U_i^T V_j), \sigma_R^2 \right) \right]^{I_{ij}^R} \\
 &\times \prod_{i=1}^m \mathcal{N}(U_i | 0, \sigma_U^2 \mathbf{I}) \times \prod_{j=1}^n \mathcal{N}(V_j | 0, \sigma_V^2 \mathbf{I}).
 \end{aligned}$$



RMF and PMF

- PMF is the probabilistic interpretation of RMF
- PMF and RMF have the same optimization objective function



Non-negative Matrix Factorization

- NMF
 - Non-negative constraints on all entries of matrices U and V

$$\begin{pmatrix} -10 & \dots & u_{1d} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{md} \end{pmatrix} \times \begin{pmatrix} -10 & \dots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{d1} & \dots & v_{dn} \end{pmatrix} \rightarrow \text{Non-negativity Constraint}$$

$(-10) * (-10) = 100$



Non-negative Matrix Factorization

- NMF
 - Given an observed matrix Y , to find two non-negative matrices U and V
 - Two types of loss functions

- Squared error function

$$\sum_{ij} (R_{ij} - U_i^T V_j)^2$$

- Divergence

$$D(R||U^T V) = \sum_{ij} (R_{ij} \log \frac{R_{ij}}{U_i^T V_j} - R_{ij} + U_i^T V_j)$$

- Solving by multiplicative updating rules



Non-negative Matrix Factorization

- Multiplicative updating rules
 - For divergence objective function

$$u_{il} \leftarrow u_{il} \frac{\sum_j v_{jl} y_{ij} / (\hat{y}_{ij})}{\sum_a v_{al}}$$
$$v_{il} \leftarrow v_{il} \frac{\sum_j u_{jl} y_{ij} / (\hat{y}_{ij})}{\sum_a u_{al}}$$

