

Recommendation Systems

- Predict items a user may be interested in based on information about the user and the items
- An effective way to help people cope with information overload
- Examples: Amazon, Netflix, Tivo, …

So How Can We Do It?

- The content based approachE.g. full text search results
- The user feedback based approach
 E.g. rating
- Which one is better?? Any room for improvement??

Collaborative Filtering

Rate items based on the ratings of other users who have similar taste as you

Problem Definitions

Prediction

- Given: a user and k items
- Return: predicted rating for each item
- Recommendation
 - Given: a user
 - Return: k items from the database with the highest predicted rating

Basic Assumptions

- Items are evaluated by users explicitly or implicitly
 - Ratings, reviews
 - Purchases, browsing behaviors
 - ...
- We may map explicit and implicit evaluations to a rating scale, e.g. 1-5.

Heuristic

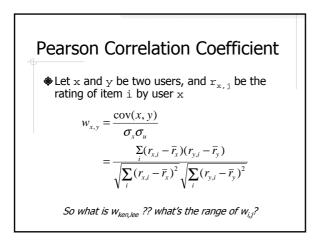
People who agreed in the past are likely to agree in the future

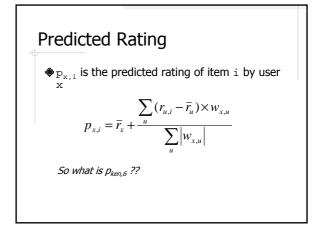
Problem Formulation

User-Item Matrix

Item	Ken	Lee	Meg	Nan
1	1	4	2	2
2	5	2	4	4
3			3	
4	2	5		5
5	4	1		1
6	??	2	5	

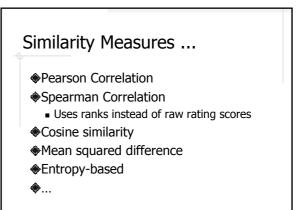
So what would be Ken's rating for Item 6??

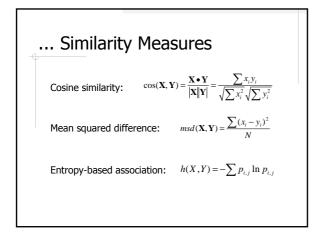


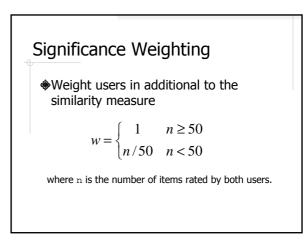


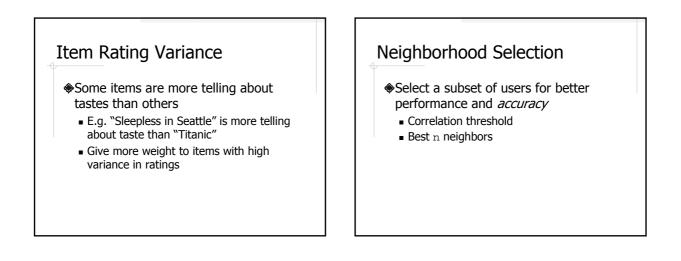
Variations and Optimizations

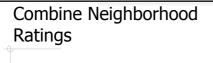
- Similarity measure
- Significance weighting
- Item rating variance
- Neighborhood selection
- Combine neighborhood ratings

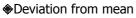


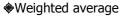










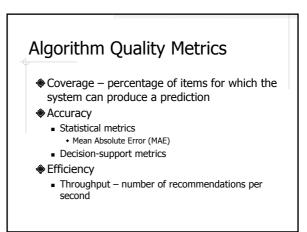


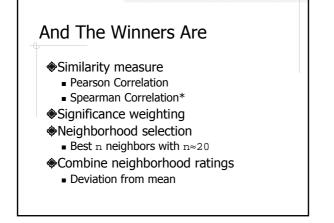
Weighted average of z-scores

Mean absolute deviation: $s = \frac{1}{n} \sum_{i=1}^{n} |r_i - \vec{r}|$

 $z_i = \frac{r_i - r_i}{r_i}$

Standardized measurement (*z-score*):





Other Recommendation Algorithms

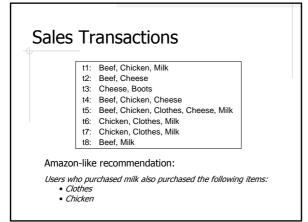
- Combine collaborative and contentbased filtering
- Item-item collaborative filtering
- Bayesian networks
- ...

Some Libraries

- Taste <u>http://taste.sourceforge.net/</u>
 COFE –
- http://eecs.oregonstate.edu/iis/CoFE/
- And more <u>http://en.wikipedia.org/wiki/Collaborativ</u> <u>e filtering#Software libraries</u>

Non-personalized Recommendation

- What if the user is new to the site?
- What if the site itself is new, i.e. no previous user transactions?



Support Count

- The support count, or frequency, of a itemset is the number of the transactions that contain the itemset
 - Item, Itemset, and Transaction
- Examples:
 - support_count({beef})=5
 - support_count({beef,chicken,milk
 - })=??

Frequent Itemset

- An itemset is frequent if its support count is greater than or equals to a minimum support count threshold
 - support_count(X)≥min_sup
- Frequent itemset mining

The Apriori Property

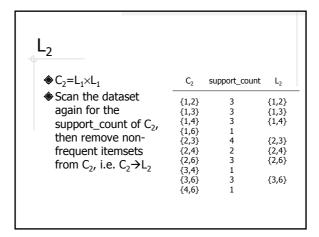
- All nonempty subsets of a frequent itemset must also be frequent
- Or, if an itemset is not frequent, its supersets cannot be frequent either

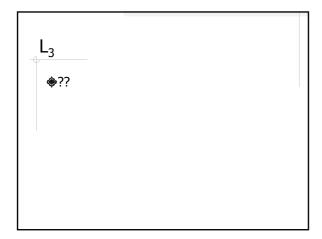
Finding Frequent Itemsets – The Apriori Algorithm

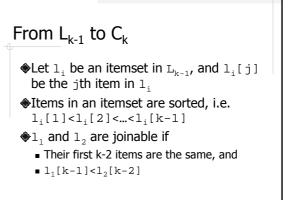
- ♦Given min_sup
- $\ensuremath{\circledast}\xspace$ Find the frequent 1-itemsets $\ensuremath{\mathtt{L}}_{\ensuremath{\mathtt{1}}}$
- $\mbox{\ensuremath{\$}}$ Find the the frequent k-itemsets \mathtt{L}_{k} by joining the itemsets in \mathtt{L}_{k-1}
- $\ensuremath{\circledast}\xspace{\ensuremath{\mathsf{Stop}}\xspace}$ when \mathtt{L}_k is empty

oriori Alg	orithm Exa	mple
	TID	Transactions
beef 1	1	1, 2, 3
chicken 2	2	1, 4
milk 3	- 3	4, 5
cheese 4	-	
boots 5	4	1, 2, 4
clothes 6	5	1, 2, 6, 4, 3
	6	2, 6, 3
	7	2, 6, 3
Support 25	% 8	1, 3

L ₁			
Scan the data once to get the count of	C_1	support_count	L ₁
to get the count of each item	{1}	5	{1}
Remove the items	{2}	5	{2}
that do not meet min_sup	{3}	5	{3}
	{4}	4	{4}
	{5}	1	
	{6}	3	{6}







From C_k to L_k References GroupLens: An Open Architecture for Collaborative Filtering of Netnews by P. $\ensuremath{\circledast}\xspace \ensuremath{\mathsf{Reduce}}$ the size of $c_{_{\!\rm K}}$ using the Apriori property Resnick et. al, 1994. ■ any (k-1)-subset of an candidate must be An Algorithmic Framework for Performing frequent, i.e. in L_{k-1} Collaborative Filtering by J. Herlocker et. Al, 1999. Scan the dataset to get the support E-Commerce Recommendation Applications counts by J. B. Schafer et. al, 2001. Data Mining: Concepts and Techniques by Jiawei Han and Micheline Kamber