

## So How Can We Do It?

The content based approachThe user feedback based approach

#### **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

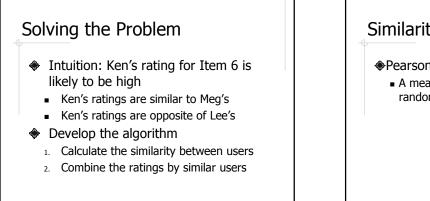
People who agreed in the past are likely to agree in the future

# **Problem Formulation**

User-Item Matrix

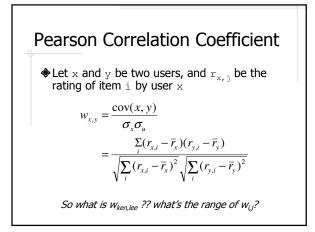
Item	Ken	Lee	Meg	Nan
1	1	4	2	2
2	5	2	4	4
3			4	
4	2	5		5
5	4	1		1
6	??	2	5	

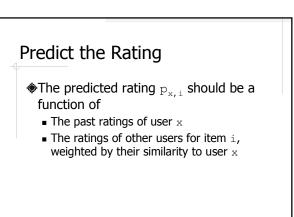
So what would be Ken's rating for Item 6??



## Similarity Measure

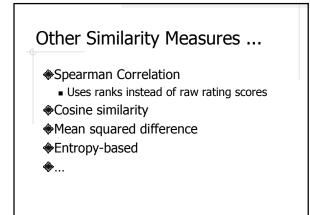
 Pearson Correlation Coefficient
 A measure of linear correlation of two random variables

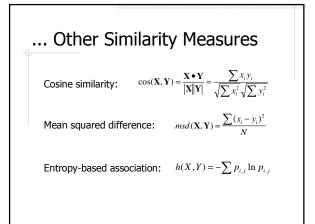


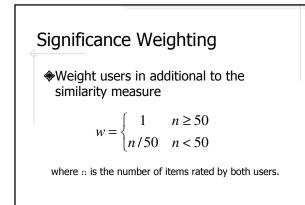


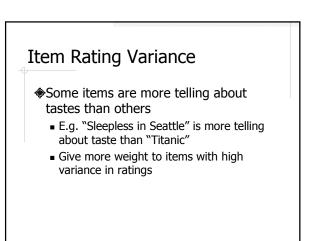
#### Variations and Optimizations

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









#### Neighborhood Selection

- Select a subset of users for better performance and *accuracy* 
  - Correlation threshold
  - Best n neighbors

#### Combine Neighborhood Ratings

- Deviation from mean
- Weighted average
- Weighted average of z-scores

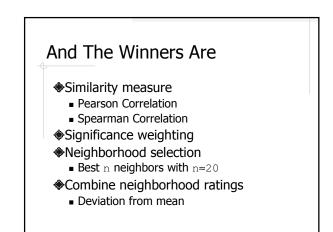
```
s = \frac{1}{n} \sum_{i=1}^{n} \left| r_i - \overline{r} \right|
```

Standardized measurement (*z-score*):

Mean absolute deviation:

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

# Algorithm Quality Metrics Coverage – percentage of items for which the system can produce a prediction Accuracy Statistical metrics Mean Absolute Error (MAE) Decision-support metrics Efficiency Throughput – number of recommendations per second



# Other Recommendation Algorithms

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

...

#### References

- GroupLens: An Open Architecture for Collaborative Filtering of Netnews by P. Resnick et. al, 1994.
- An Algorithmic Framework for Performing Collaborative Filtering by J. Herlocker et. Al, 1999.
- *E-Commerce Recommendation Applications* by J. B. Schafer et. al, 2001.