6-3: Item-Item on Unary Data

Introduction to Recommender Systems

Introduction

- We've talked about item-item over rating data
- Also works well on unary data (implicit feedback)
  - clicks
  - plays
  - purchases
- But some tweaks are needed

Data Representation

- Rating values: user-item rating matrix
- Need some matrix to represent data
  - Logical (1/0) user-item ‘purchase’ matrix
  - Purchase count matrix
- Problem: what is a 0?
  - We just ignore that for item-item

Data Normalization

- Standard mean-centering not meaningful
- But we can normalize user vectors to unit vectors
  - Intuition: users who like many items provide less information about any particular pair
- Could also consider: logging counts

Computing Similarities

- Cosine similarity still works
- Can also use conditional probability
  - see Deshpande and Karypis paper

Aggregating Scores

- Weighted average works for non-binary
  - counts
- For binary (0/1), just sum neighbor similarities
  - fixed neighborhood size means this isn't unbounded

$$\text{score}(u, i) = \sum_{j \in N} \text{sim}(i, j)$$

  - Neighborhood selection unchanged (most similar)

Conclusion

- Item-item basically works for unary data
- A few tweaks to algorithm components needed to make it well-behaved
- Test variants with your data/context
  - Evaluation tools we talked about last module help with this

6-3: Item-Item on Unary Data

Cosine

$$sim(i,j) = \frac{\vec{i}\cdot\vec{j}}{||\vec{i}||\,||\vec{j}||}$$

$U(i)$ - users who bought i

Cond Prob

$$sim(i,j)=P(j|i)=\frac{P(i,j)}{P(i)} = \frac{\frac{|U(i)\cap U(j)|}{n}}{\frac{|U(i)|}{n}}$$

$$= \frac{|U(i)\cap U(j)|}{|U(i)|}$$

$-1 \le sim \le 1$

$$\frac{P(i,j)}{P(i)\cdot P(j)^{\alpha}}$$

$0<\alpha<1$