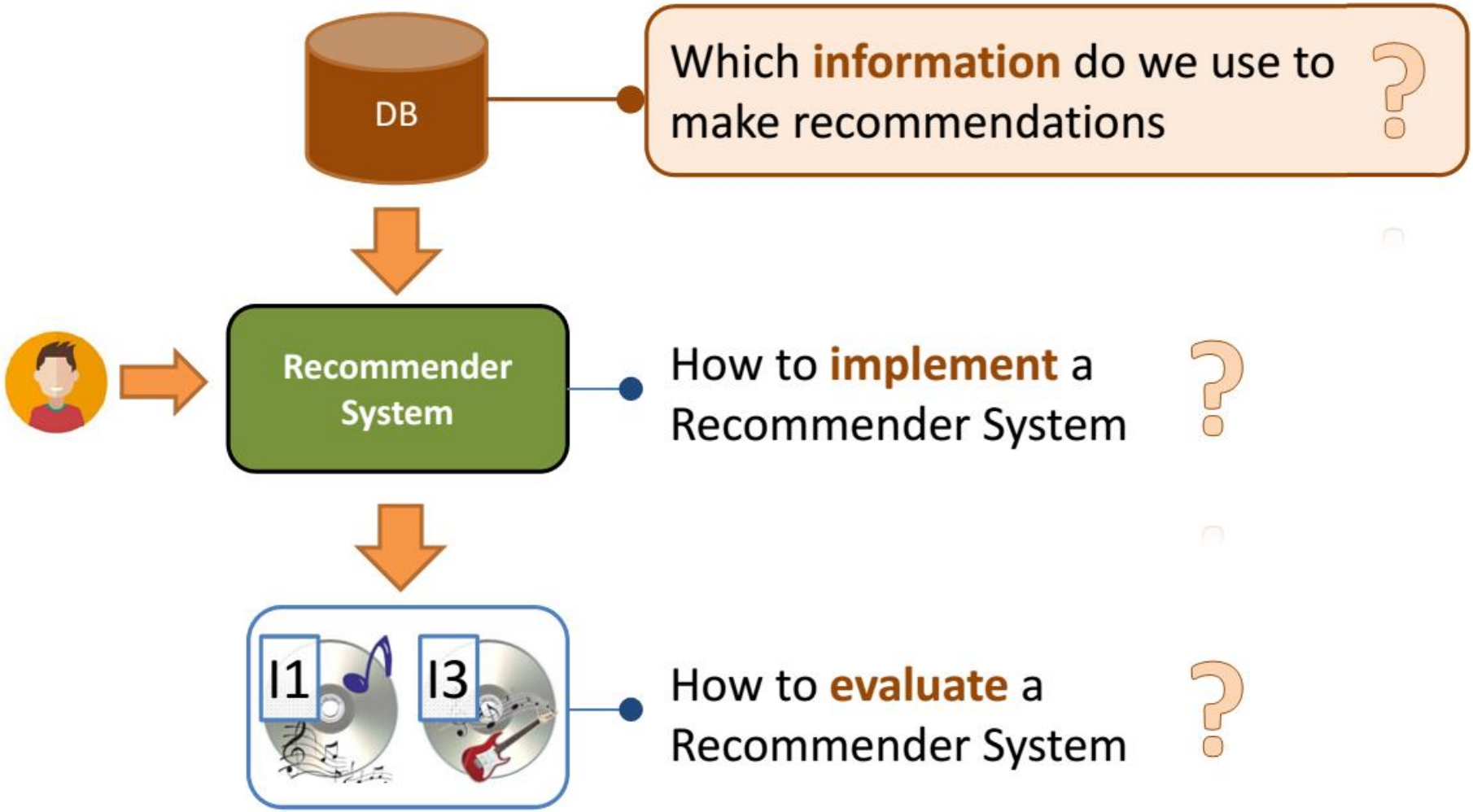
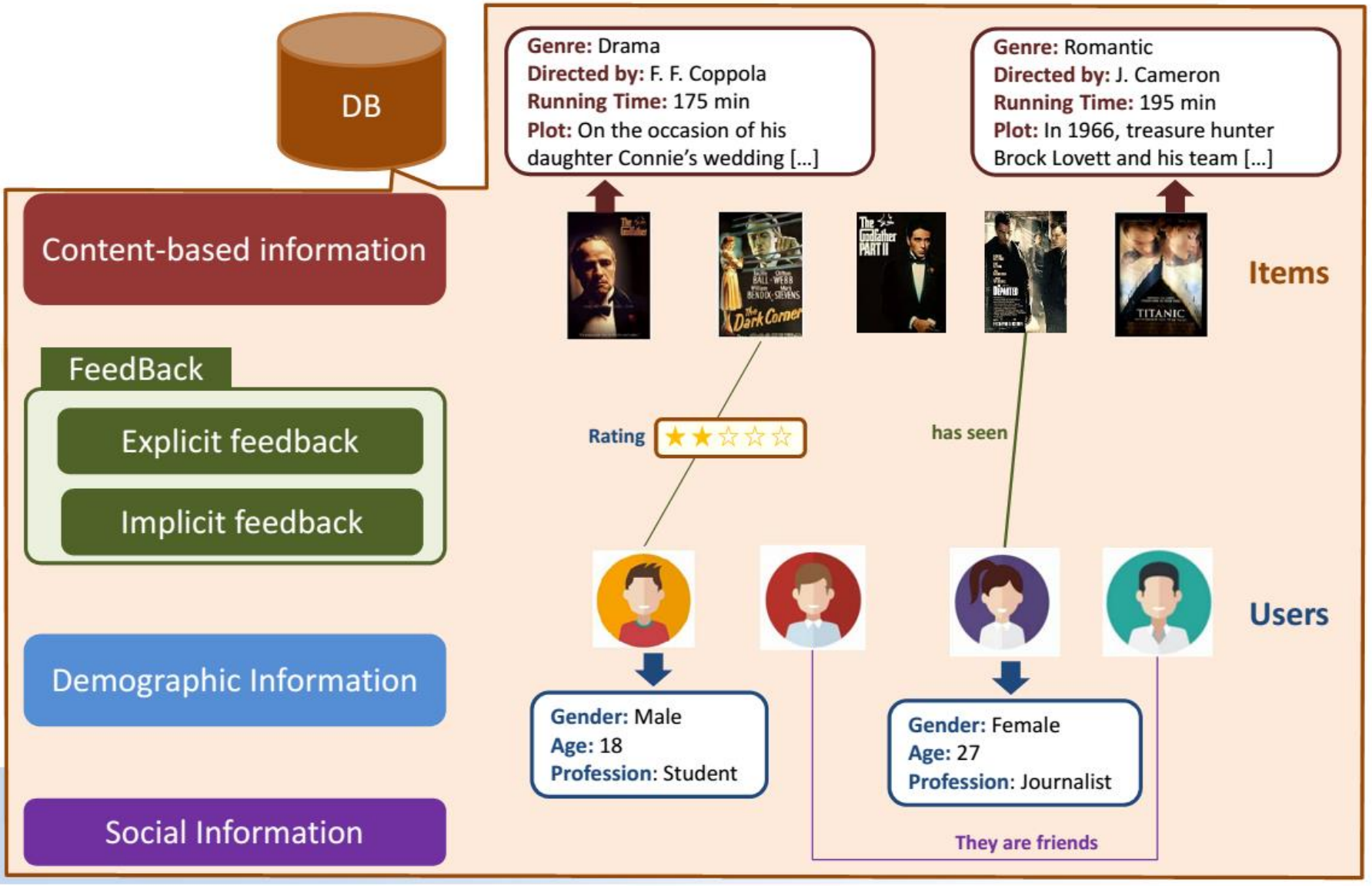
IT 202: Topics in Recommender System

Pramit Mazumdar

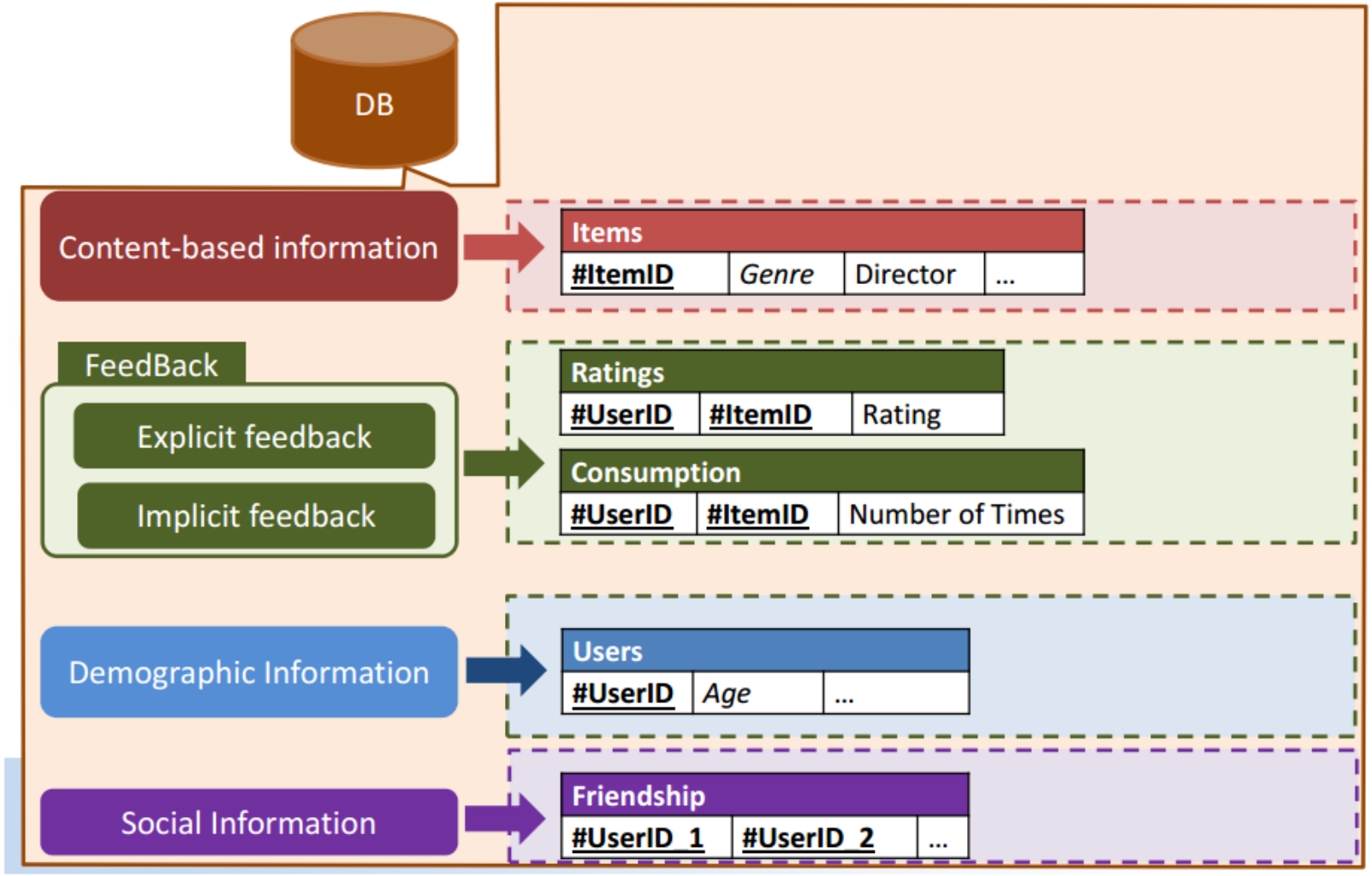
Recommender System Overview



Sources of Information



Information in Database



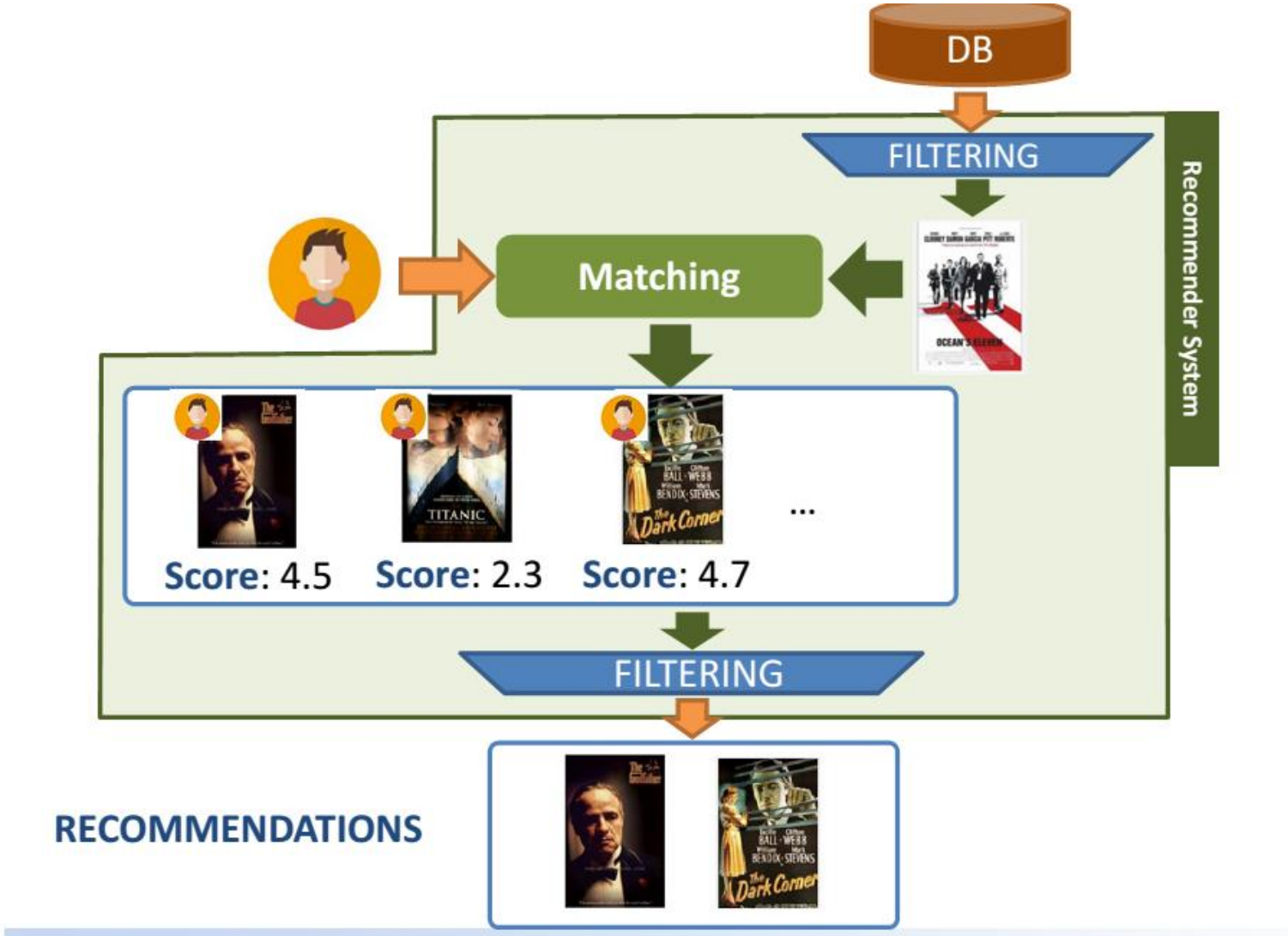
Types of Recommender Systems

- Collaborative Filtering-based systems (Memory-based approach)
  - K-NN-based approach
    - Oriented towards Users
    - Oriented towards Items
- Content-based recommender systems (Model-based approach)
  - Matrix Factorization
  - Learning-based approach

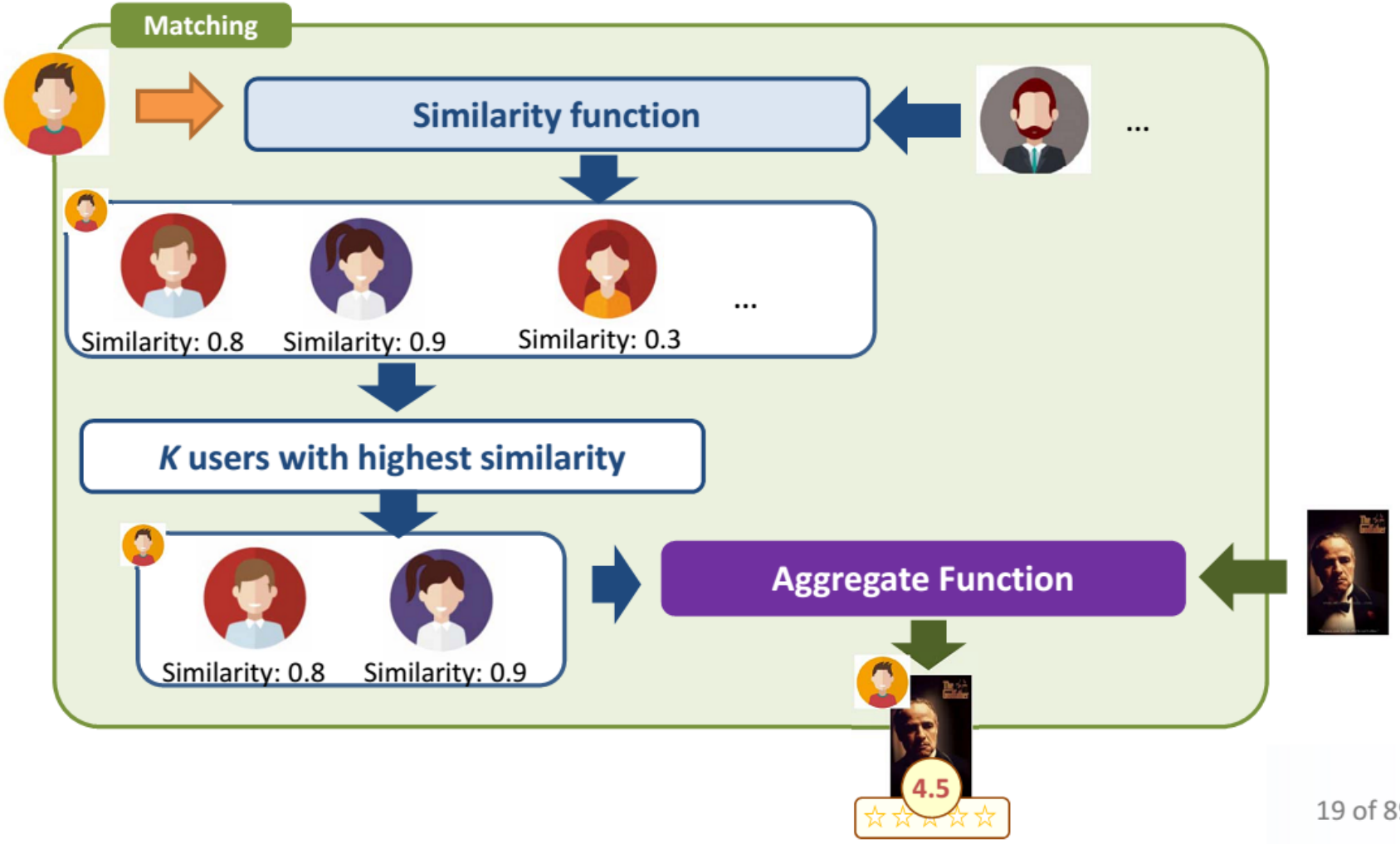
Evaluation Metrics

- Accuracy
- Diversity
- Trust
  - Reliability
- Serendipity
  - Novelty

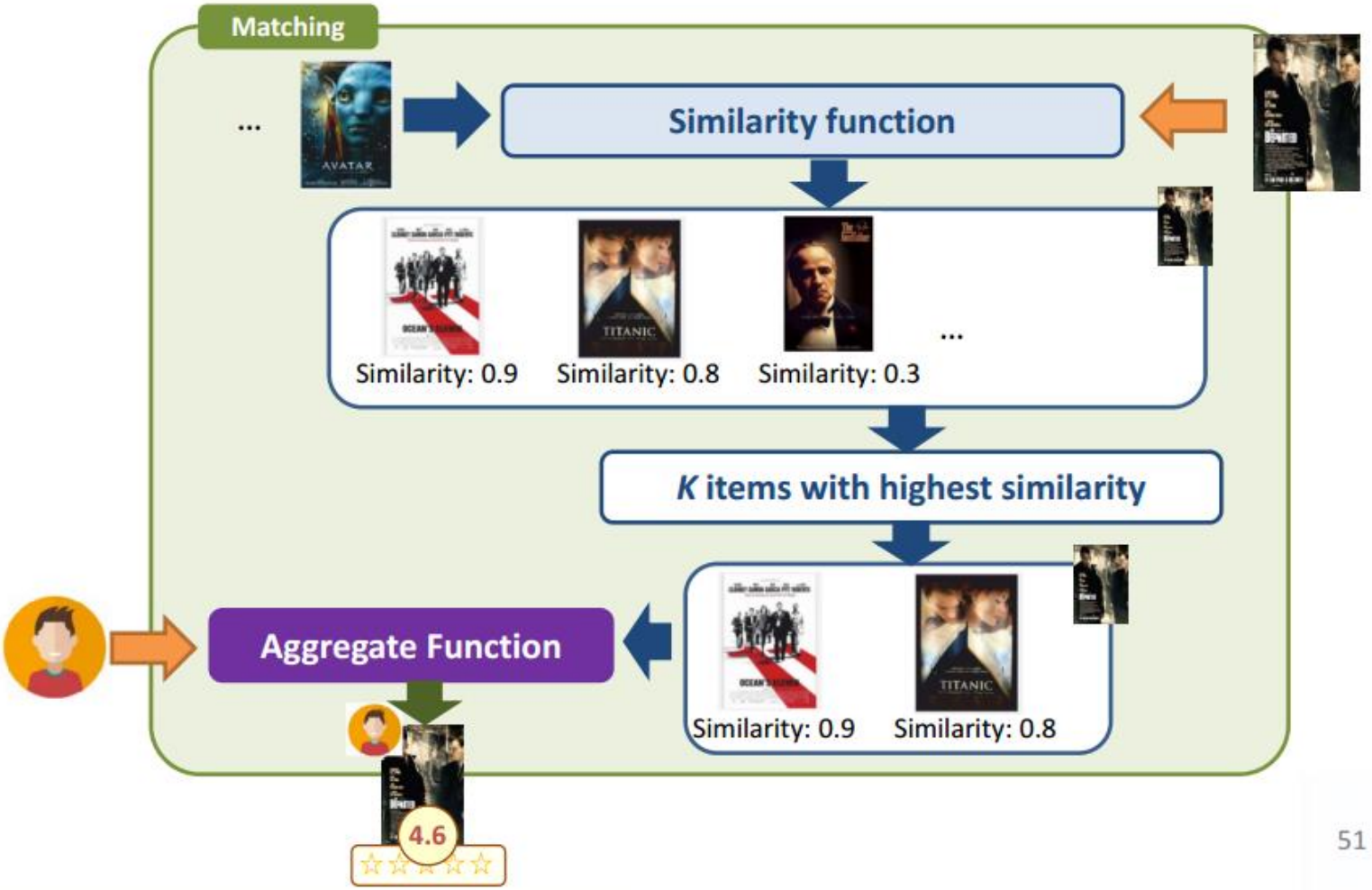
General Architecture



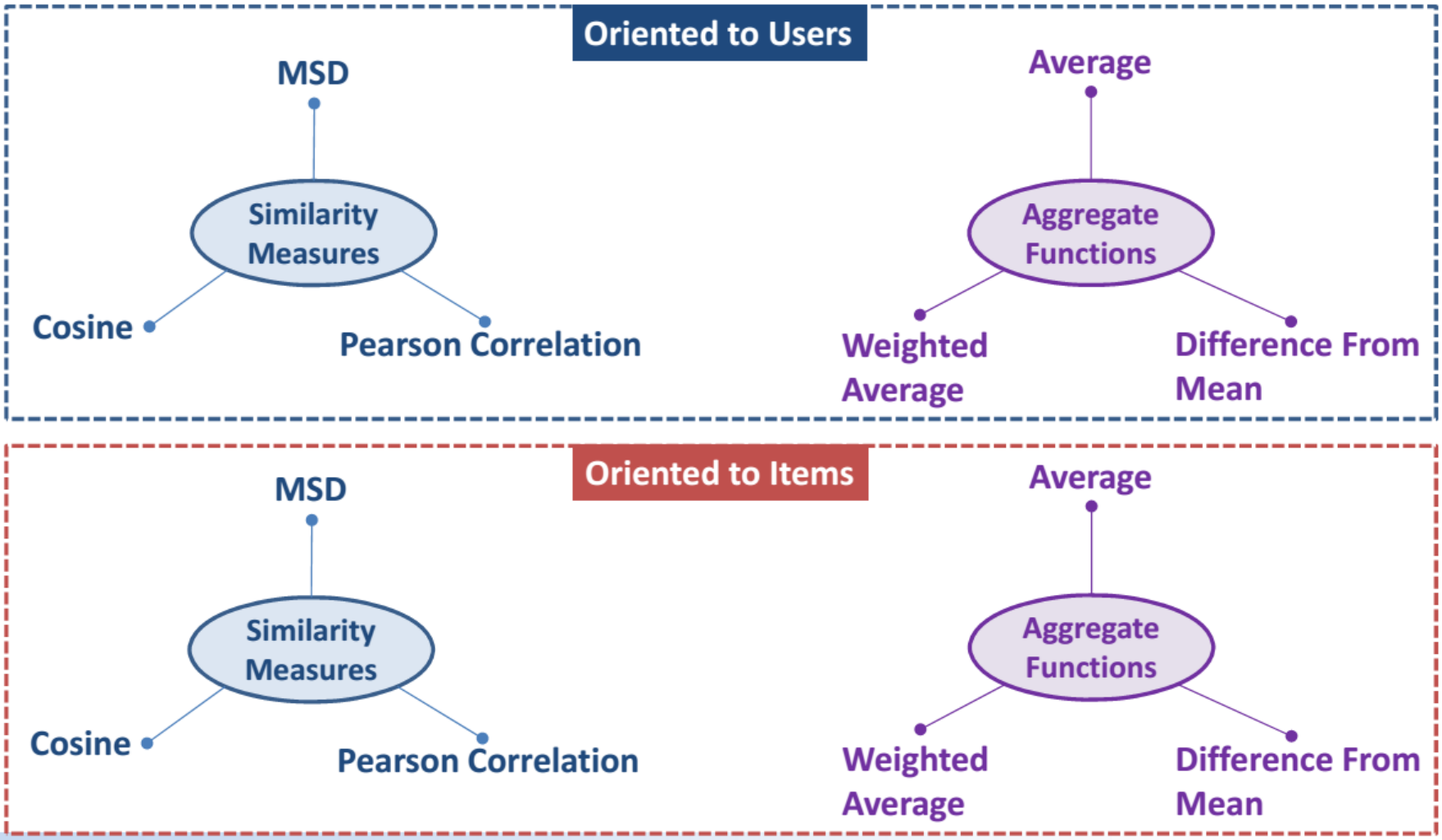
KNN-oriented towards Users



KNN-oriented towards Items



Similarity and Aggregate Functions



Similarity Measures

MSD between Users

Mean Square Difference

$$MSD(u,v) = \frac{\sum_{\{i|\, r_{ui}\neq\cdot;\, r_{vi}\neq\cdot\}} (r_{ui} - r_{vi})^2}{|\{i|\, r_{ui} \neq\cdot;\, r_{vi} \neq\cdot\}|}$$

Mean Square Difference

$$MSD(i,j) = \frac{\sum_{\{u|\, r_{ui}\neq\cdot;\, r_{uj}\neq\cdot\}} (r_{ui} - r_{uj})^2}{|\{u|\, r_{ui} \neq\cdot;\, r_{uj} \neq\cdot\}|}$$

MSD between Items

Similarity Measures

Cosine between Users

$$cos(u,v) = \frac{\sum_{\{i|\, r_{ui}\neq\cdot;\, r_{vi}\neq\cdot\}} r_{ui} r_{vi}}{\sqrt{\sum_{\{i|\, r_{ui}\neq\cdot;\, r_{vi}\neq\cdot\}} r_{ui}^2}\sqrt{\sum_{\{i|\, r_{vi}\neq\cdot;\, r_{vi}\neq\cdot\}} r_{vi}^2}}$$

$$cos(i,j) = \frac{\sum_{\{u|\, r_{ui}\neq\cdot;\, r_{uj}\neq\cdot\}} r_{ui} r_{uj}}{\sqrt{\sum_{\{u|\, r_{ui}\neq\cdot;\, r_{uj}\neq\cdot\}} r_{ui}^2}\sqrt{\sum_{\{u|\, r_{ui}\neq\cdot;\, r_{uj}\neq\cdot\}} r_{uj}^2}}$$

Cosine between Items

Similarity Measures

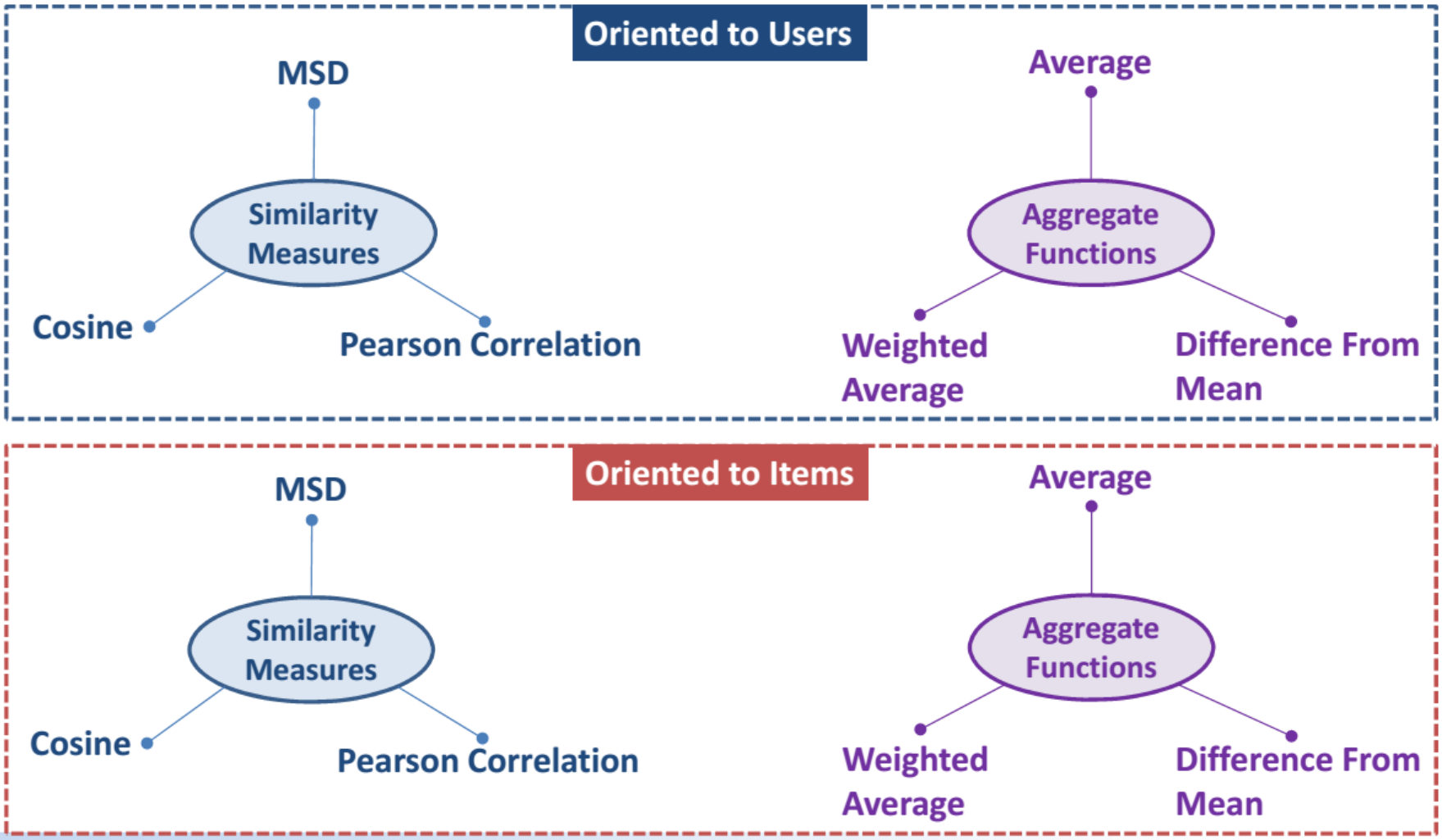
Pearson Correlation between Users

$$\rho(u,v) = \cos(u^*, v^*) = \frac{\sum_{\{i|\, r_{ui}\neq\cdot;\, r_{vi}\neq\cdot\}} (r_{ui} - \overline{r_u})(r_{vi} - \overline{r_v})}{\sqrt{\sum_{\{i|\, r_{ui}\neq\cdot;\, r_{vi}\neq\cdot\}} (r_{ui} - \overline{r_u})^2}\ \sqrt{\sum_{\{i|\, r_{ui}\neq\cdot;\, r_{vi}\neq\cdot\}} (r_{vi} - \overline{r_v})^2}}$$

$$\rho(i,j) = \frac{\sum_{\{u|\, r_{ui}\neq\cdot;\, r_{uj}\neq\cdot\}} (r_{ui} - \overline{r_i})(r_{uj} - \overline{r_j})}{\sqrt{\sum_{\{u|\, r_{ui}\neq\cdot;\, r_{uj}\neq\cdot\}} (r_{ui} - \overline{r_i})^2}\ \sqrt{\sum_{\{u|\, r_{ui}\neq\cdot;\, r_{uj}\neq\cdot\}} (r_{uj} - \overline{r_j})^2}}$$

Pearson Correlation between Items

Similarity and Aggregate Functions



Aggregate Function

Average between Users

$$p_{u,i} = \frac{\sum_{\{v|r_{vi}\neq\cdot\}} r_{vi}}{|v|\{r_{vi} \neq \cdot\}|}$$

$$p_{u,i} = \frac{\sum_{\{j|r_{uj}\neq\cdot\}} r_{uj}}{|\{j|r_{uj} \neq \cdot\}|}$$

Average between Items

Aggregate Function

Weighted Average between Users

$$p_{u,i} = \frac{\sum_{\{v|\, r_{vi} \neq \cdot\}} sim(u,v) r_{vi}}{\sum_{\{v|r_{vi} \neq \cdot\}} sim(u,v)}$$

$$p_{u,i} = \frac{\sum_{\{j|r_{uj} \neq \cdot\}} sim(i,j) r_{uj}}{\sum_{\{j|r_{uj} \neq \cdot\}} sim(i,j)}$$

Weighted Average between Items

Aggregate Function

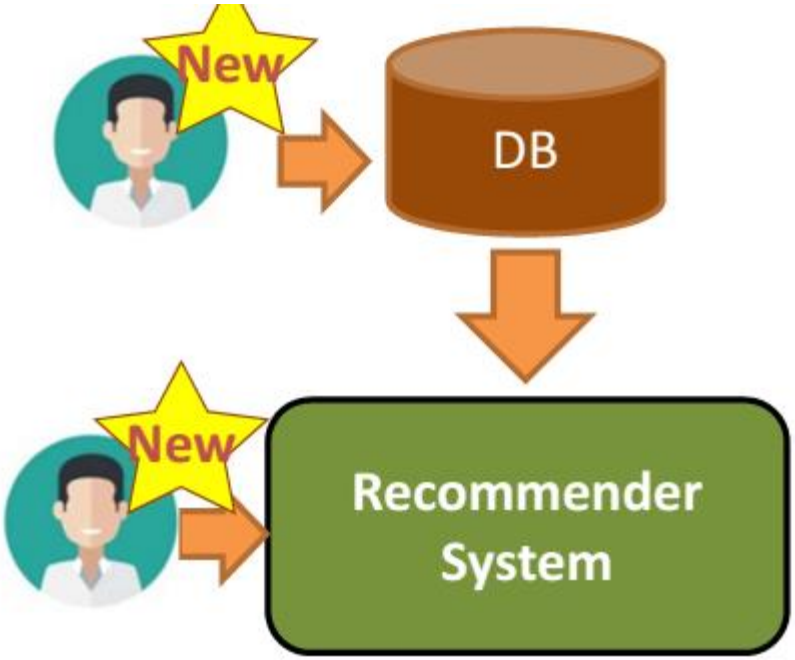
Difference from Mean between Users

$$p_{u,i} = \overline{r_u} + \frac{\sum_{\{v|\, r_{vi} \neq \cdot\}} sim(u,v)(r_{vi} - \overline{r_v})}{\sum_{\{v|\, r_{vi} \neq \cdot\}} sim(u,v)}$$

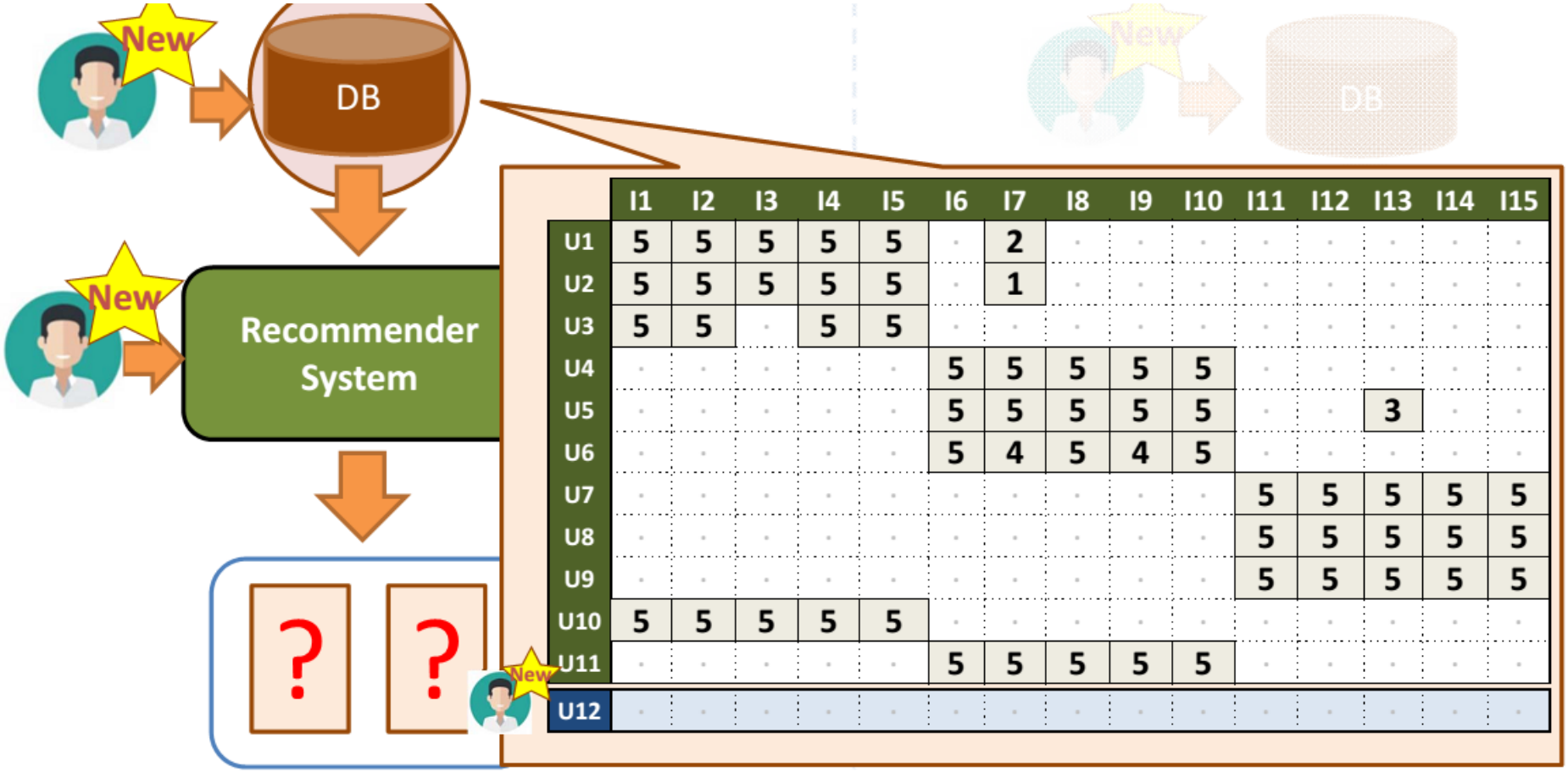
$$p_{u,i} = \overline{r_i} + \frac{\sum_{\{j|r_{uj} \neq \cdot\}} sim(i,j)(r_{uj} - \overline{r_j})}{\sum_{\{j|r_{uj} \neq \cdot\}} sim(i,j)}$$

Difference from Mean between Items

User Cold-start Problem

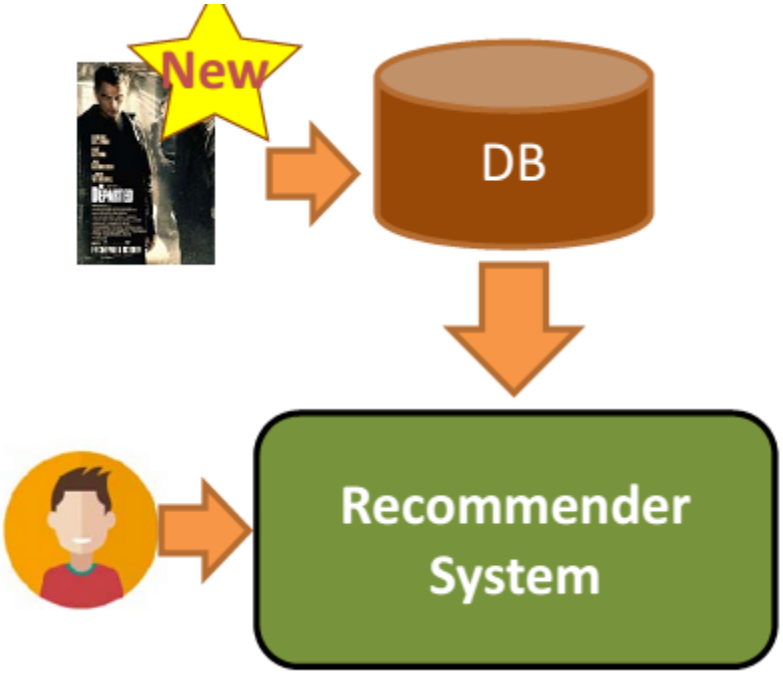


User Cold-start Problem

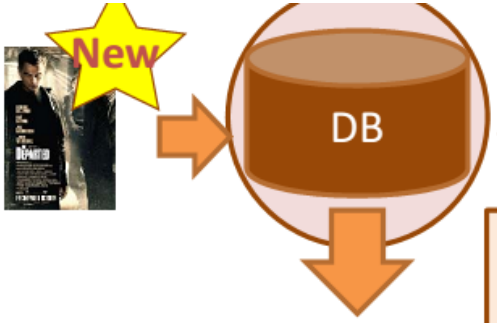


	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15
U1	5	5	5	5	5		2								
U2	5	5	5	5	5		1								
U3	5	5		5	5										
U4						5	5	5	5	5					
U5						5	5	5	5	5			3		
U6						5	4	5	4	5					
U7											5	5	5	5	5
U8											5	5	5	5	5
U9											5	5	5	5	5
U10	5	5	5	5	5										
U11						5	5	5	5	5					
U12															

Item Cold-start Problem



Item Cold-start Problem



This new item is not recommended to any user

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16
U1	5	5	5	5	5		2									
U2	5	5	5	5	5		1									
U3	5	5		5	5											
U4						5	5	5	5	5						
U5						5	5	5	5	5			3			
U6						5	4	5	4	5						
U7											5	5	5	5	5	
U8											5	5	5	5	5	
U9											5	5	5	5	5	
U10	5	5	5	5	5											
U11						5	5	5	5	5						

Thank You