Collaborative Filtering

Group 2 - Joaquim Cavalheiro, Tiantian Chen, Stephanie Park, Yufei Zhao, Xiaoyu Zhou

Agenda

Introduction

Datasets

Assignment

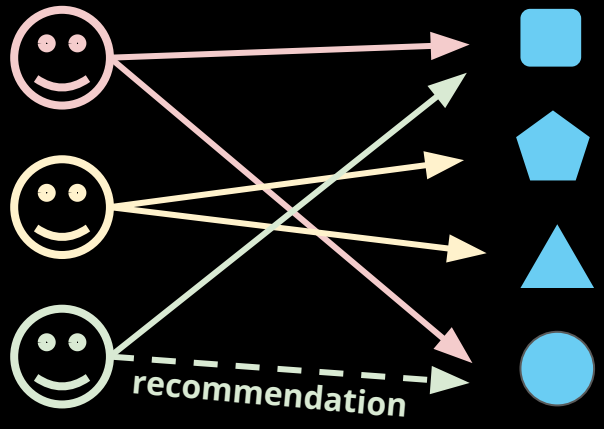
Methods

Results

Introduction: Collaborative Filtering

In **collaborative filtering**, algorithms are used to make automatic predictions about a user's interests by compiling preferences from several users. It is commonly used to build personalized recommendations.

Memory-Based filtering calculates the likeness between the users or items to make recommendations, while **Model-Based** filtering uses the user database to estimate or learn a model to make predictions.



Datasets

1. Anonymous Microsoft Web Data - An implicit voting data, with each vroot is either characterized as being visited or not, which is represented respectively as one vote or no vote.

2. Compaq System Research Center's EachMovie Data - An explicit voting data, with votes ranging in value from 0 to 5. The dataset comes from 61,265 users entering a total of 2,811,983 numeric ratings on 1,623 movies.

Assignment

	Similarity Weight				Signific. Weight		Variance Weight		Selecting Neighbors			Rating Normalization	Cluster Model
	S	V	E	SR	V	NV	V	NV	W	B	C	Dev. Mean	
Dataset 1	●	●	●	●		●	●	●	●	●	●	●	
Dataset 2	●	●	●			●	●	●	●	●	●	●	●

Methods

Memory-Based Algorithms

Preprocessing

↓

User correlation & Similarity Weight Matrix

↓

Evaluation

Model-Based Algorithms

Preprocessing

↓

EM Algorithm to learn parameters & cross-validation to decide cluster numbers

↓

Evaluation

Evaluation

Cluster Model	Spearman Corr.	Vector Similarity	Entropy
	Threshold 0.65	Threshold 0.68	Threshold 0.69
MAE 1.14	Best-n 0.69	Best-n 0.70	Best-n 0.70
ROC 0.75	Combined 0.74	Combined 0.75	Combined 0.76

Algorithm		**MAE**	**ROC**					
Model-based Algorithm	Clustering Model	1.138387	0.7578769					
	MAE	Weight Threshold(0.3)		Best-n-estimator(20)		combined(Threshold:0.1, Top:20)		
Algorithm	Spearman Correlation	0.6542189		0.6973992		0.6973922		
Memory-based Algorithm	Vector Similarity	0.6876153		0.6976481		0.6976861		
for MS	Entropy	0.6983197		0.7017892		0.701569		
	MAE	Weight Threshold(0.3)		Best-n-estimator(20)		combined(Threshold:0.1, Top:20)		
Algorithm	Spearman Correlation	0.7548987		0.7543198		0.7536121		
Memory-based Algorithm	Vector Similarity	0.7494321		0.74891		0.7389159		
for movie	Entropy	0.7457895		0.7567332		0.761001		

Thank you!