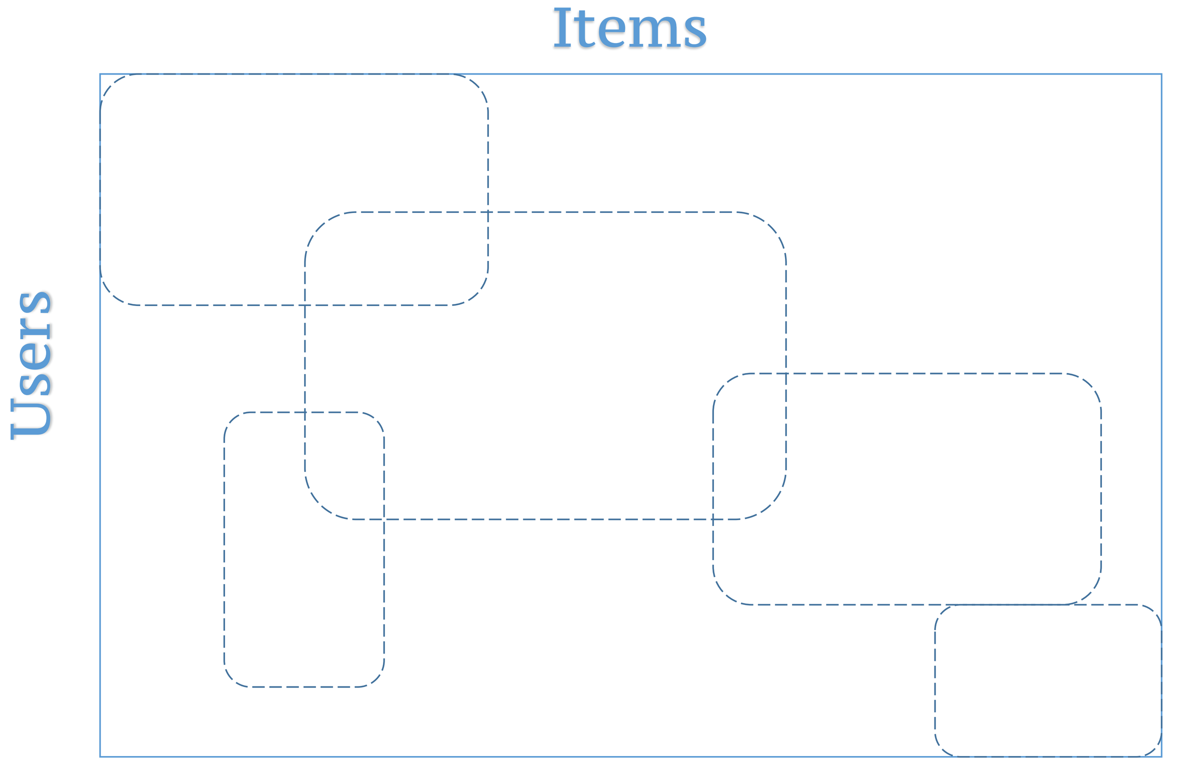
CCCF: Improving Collaborative Filtering via Scalable User-Item Co-Clustering

Yao Wu[1⋆], Xudong Liu[2⋆], Min Xie[3], Martin Ester[1] and Qing Yang[2]

⋆ Equal Contribution [1]Simon Fraser University [2]Chinese Academy of Sciences [3]@WalmartLabs

Motivation



Examples:

- People living in Phoenix are more likely to visit local restaurants rather than a restaurant in New York.
- An Amazon user shares similar tastes on books with a certain group of users, while having similar preferences with another group of users on movies.

Background:

Co-clustering methods can be used to utilize the general structure of the user-item utility matrix to improve recommender systems. However, previous co-clustering methods fail to consider several important properties of recommendation. In this paper, we propose a novel co-clustering method, called CCCF (Co-Clustering for Collaborative Filtering) to address these problems.

Overview

Overall Procedure:

1. Cluster users and items into overlapping subgroups. We present a probabilistic co-clustering method and a scalable inference algorithm based on stochastic Monte Carlo Markov Chain (MCMC).
2. In each subgroup, we can apply any collaborative filtering methods to learn users' preferences over the items within this subgroup. Since the learning of CF models in each subgroup is independent from each other, this step can be easily done in parallel.
3. For each user, we aggregate the recommendation results from all the subgroups that she/he is involved in. We discuss several aggregation strategies and study their performance.

CCCF (Co-Clustering for Collaborative Filtering)

Goal:

- Users/items should belong to multiple subgroups.
- Links can be explained by the subgroup affiliations.
- The more overlapping subgroups a user shares with an item, and the larger their weights are, the higher the probability that the user likes the item.
- A link between a user and an item can be explained by a dominant reason.

Co-Clustering

Generative Process:

Input: $\mathcal{O}$, $\mathcal{O}'$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$

for all $k \in \{1,\dots,K\}$ **do**
  for all $u \in \mathcal{U}$ **do**
    Draw $\phi_{uk} \sim \text{Beta}(\alpha_{k1}, \alpha_{k2})$
  end for
  for all $i \in \mathcal{I}$ **do**
    Draw $\phi_{ik} \sim \text{Beta}(\alpha_{k1}, \alpha_{k2})$
  end for
  Draw $\theta_k \sim \text{Beta}(\beta_1, \beta_2)$
end for
for all $(u, i, y_{ui}) \in \mathcal{O} \cup \mathcal{O}'$ **do**
  for all $k \in \{1,\dots,K\}$ **do**
    Draw $z_{u\to i,k} \sim \text{Bernoulli}(\phi_{uk})$
    Draw $z_{i\to u,k} \sim \text{Bernoulli}(\phi_{ik})$
    if $z_{u\to i,k} = 1$ and $z_{i\to u,k} = 1$ **then**
      Set $p_{uik} = \theta_k$
    else
      Set $p_{uik} = 0$
    end if
  end for
  Set $p_{ui} = 1 - \prod_k (1 - p_{uik})$
  Draw $y_{ui} \sim \text{Bernoulli}(p_{ui})$
end for

Parameter Estimation:

We present a stochastic MCMC method to estimate the parameters in the model.

Properties:

- **Scalability:** The complexity of training CCCF is linear in the number of observed links.
- **Flexibility:** The structure of the subgroups in CCCF is flexible to model any kind of overlapping subgroups (*e.g.*, densely overlapping, hierarchically nested or non-overlapping).
- **Interpretability:** CCCF is a probabilistic model where the resulting affiliation strengths have semantic meanings.
- **Extensibility:** When there are some informative features for users and items, it is easy to extend CCCF to leverage these features in the learning procedure.

Collaborative Filtering in Subgroups

The choice of the underlying CF methods depends on a lot of factors, such as data sparsity, complexity, model interpretability and latency of recommendation. One benefit of the framework is that the method does not rely on a specific CF algorithm. Any CF method of interest can be used and the framework will likely improve their performances. In our paper, we study four popular methods:

- **POP:** Popularity.
- **ITEMCF:** Item-base Collaborative Filtering.
- **MF:** Matrix Factorization with negative sampling.
- **WARP:** Latent Factor Model with Weighted Approximately Ranked Pairwise loss.

Recommendation

Remark:

- The prediction has a large value if the user and the item both have large affiliation strengths with this subgroup.
- The more subgroups the user and the item share, the larger their predicted score is.

Four strategies:

- **CCCF:** $\hat{y}_{ui} = \sum_k \phi_{uk} \cdot \phi_{ik} \cdot \theta_k \cdot \hat{y}^k_{ui}$
- **CCCF-PR:** $\hat{y}_{ui} = \sum_k \phi_{uk} \cdot \phi_{ik} \cdot \theta_k$
- **CCCF-AVG:** $\hat{y}_{ui} = \sum_k \hat{y}^k_{ui} / K$
- **CCCF-MAX:** $\hat{y}_{ui} = \max_k \hat{y}^k_{ui}$

CCCF vs. No Co-Clustering

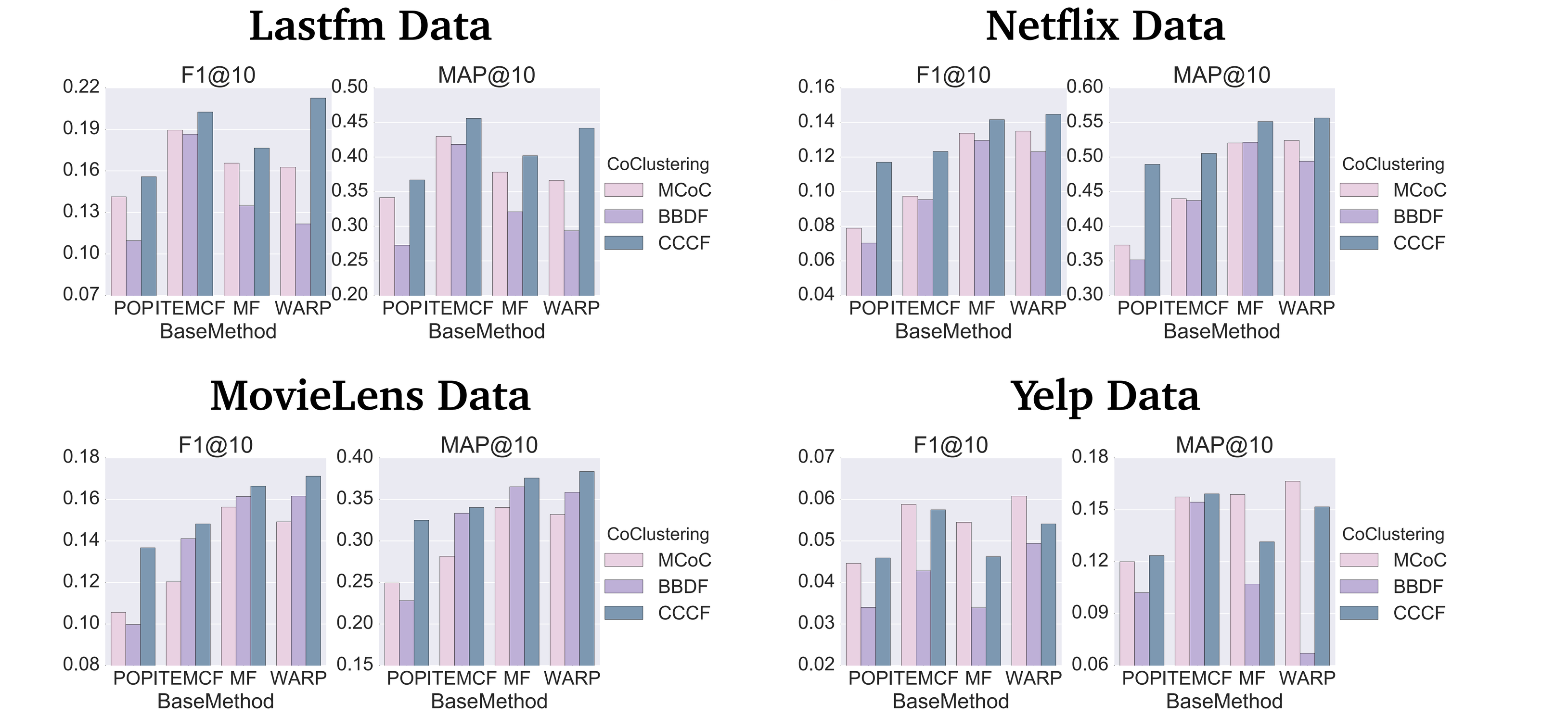
Yelp Data

	POP		ITEMCF		MF		WARP	
	NONE	CCCF	NONE	CCCF	NONE	CCCF	NONE	CCCF
P@10	0.018	0.046	0.048	0.057	0.043	0.046	0.047	0.056
R@10	0.024	0.058	0.062	0.072	0.055	0.058	0.058	0.068
F1@10	0.018	0.046	0.048	0.058	0.044	0.046	0.047	0.055
MAP@10	0.050	0.124	0.142	0.159	0.124	0.131	0.137	0.152

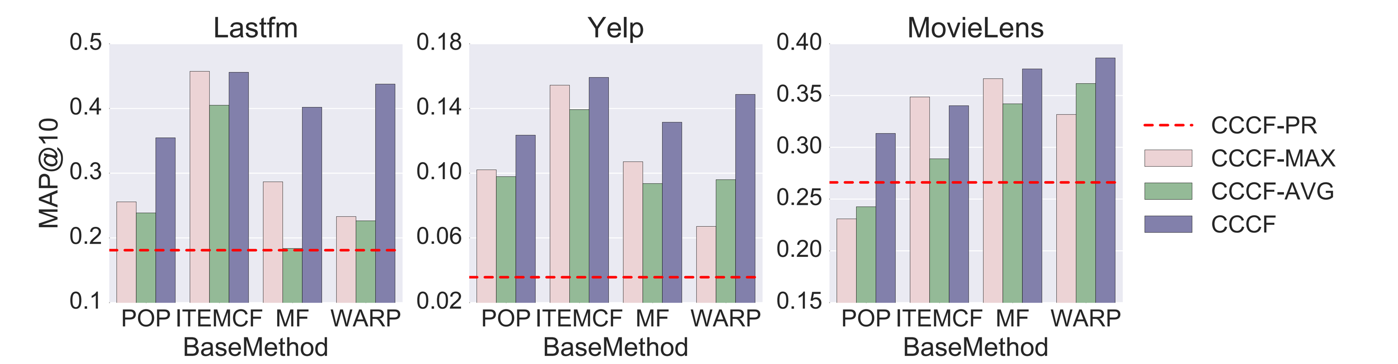
Netflix Data

	POP		ITEMCF		MF		WARP	
	NONE	CCCF	NONE	CCCF	NONE	CCCF	NONE	CCCF
P@10	0.189	0.283	0.308	0.312	0.330	0.354	0.347	0.364
R@10	0.047	0.075	0.085	0.085	0.091	0.098	0.095	0.100
F1@10	0.070	0.109	0.123	0.123	0.132	0.142	0.138	0.145
MAP@10	0.348	0.471	0.493	0.505	0.528	0.551	0.538	0.556

CCCF vs. Other Co-Clustering Methods



Recommendation Strategies



Qualitative Analysis on Yelp data

Items with largest weights in subgroups:

Cluster 1			Cluster 2			Cluster 3		
Business Name	City	Description	Business Name	City	Description	Business Name	City	Description
Gallo Blanco	Phoenix	Restaurants	Bread and Butter	Henderson	Restaurants	Las Vegas North Premium Outlets	Las Vegas	Shopping
Lux	Phoenix	Restaurants	Snow Ono Shave Ice	Las Vegas	Restaurants	Orleans Hotel & Casino	Las Vegas	Casinos & Hotels
Postino Central	Phoenix	Restaurants	Buldogis Gourmet Hot Dogs	Las Vegas	Restaurants	Fremont Street Experience	Las Vegas	Entertainment
Maizie's Cafe & Bistro	Phoenix	Restaurants	Sweet Tomatoes	Las Vegas	Restaurants	Conservatory & Botanical Garden	Las Vegas	Entertainment
Cherryblossom Noodle Cafe	Phoenix	Restaurants	Strip N Dip Chicken Strips	Las Vegas	Restaurants	Planet Hollywood Las Vegas Resort & Casino	Las Vegas	Casinos & Hotels
Pizza a Metro	Phoenix	Restaurants	Patisserie Manon	Las Vegas	Restaurants	Flamingo Las Vegas Hotel & Casino	Las Vegas	Casinos & Hotels
SideBar	Phoenix	Bars	Island Flavor	Las Vegas	Restaurants	New York - New York	Las Vegas	Casinos & Hotels
Churn	Phoenix	Coffee & Tea	Japanese Curry Zen	Las Vegas	Restaurants	Miracle Mile Shops	Las Vegas	Shopping
Carly's Bistro	Phoenix	Restaurants	Slidin' Thru	Las Vegas	Restaurants	Palms Casino Resort	Las Vegas	Casinos & Hotels
Federal Pizza	Phoenix	Restaurants	Art of Flavors	Las Vegas	Food	The Forum Shops	Las Vegas	Shopping