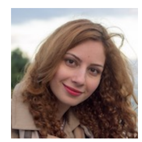
Discerning Canonical User Representation for Cross-Domain Recommendation

Siqian Zhao

szhao2@albany.edu



Sherry Sahebi

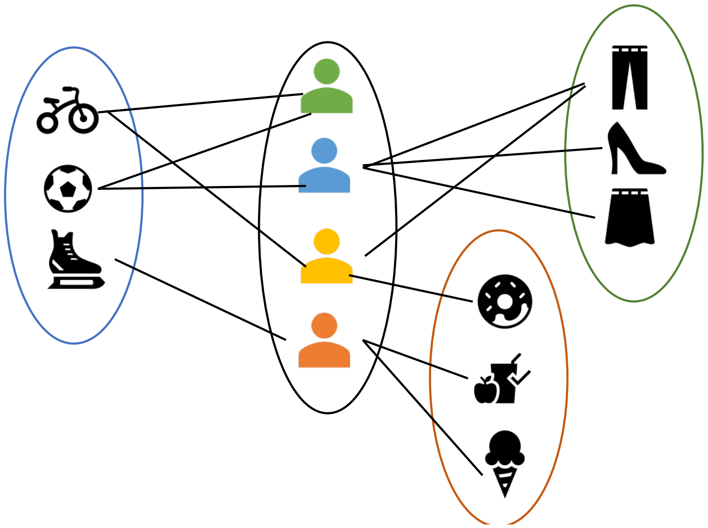
szhao2@albany.edu

RecSys
acm.org

Cross-domain recommender systems (CDRs)

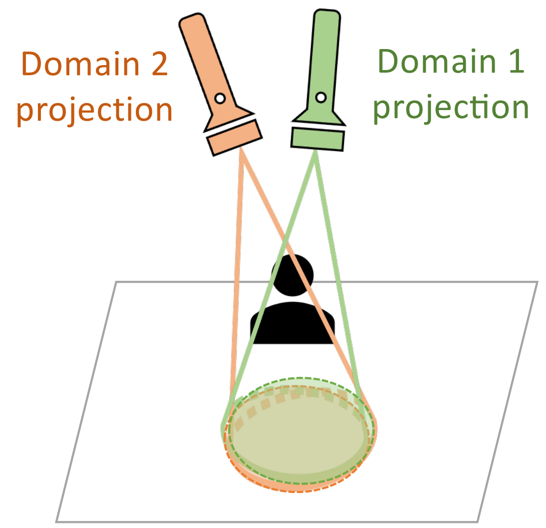
- Transferring information across domains [Fernández-Tobías et. al. 2012]
  - Addressing data sparsity and the cold-start issues
  - Improving the quality of recommendations

- Assuming some user interest similarity across domains



User representation similarity across domains

- Complete overlap
  - Same or similar user representation shared across domains
  - Too restricted

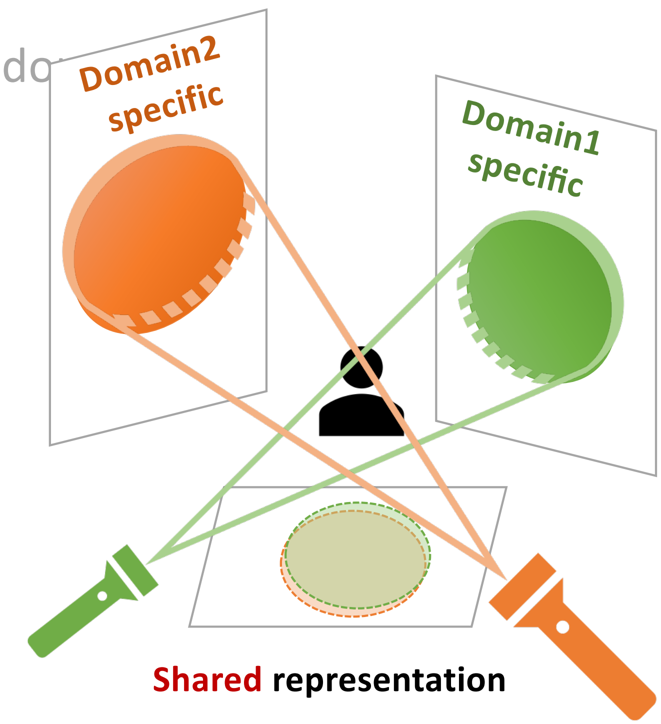


Shared representation space

User representation similarity across domains

- Complete overlap
  - Same or similar user representation shared across do

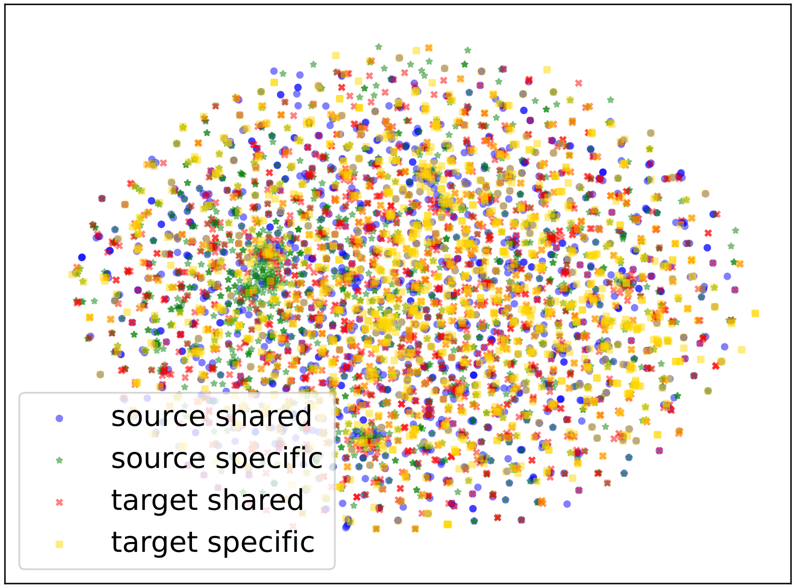
- Some overlap
  - Domain-shared representation
    - Information sharing between domains
  - Domain-specific representation
    - Representing unique user interests within each domain
  - Unbalanced freedom



Imbalance between domain specific & shared

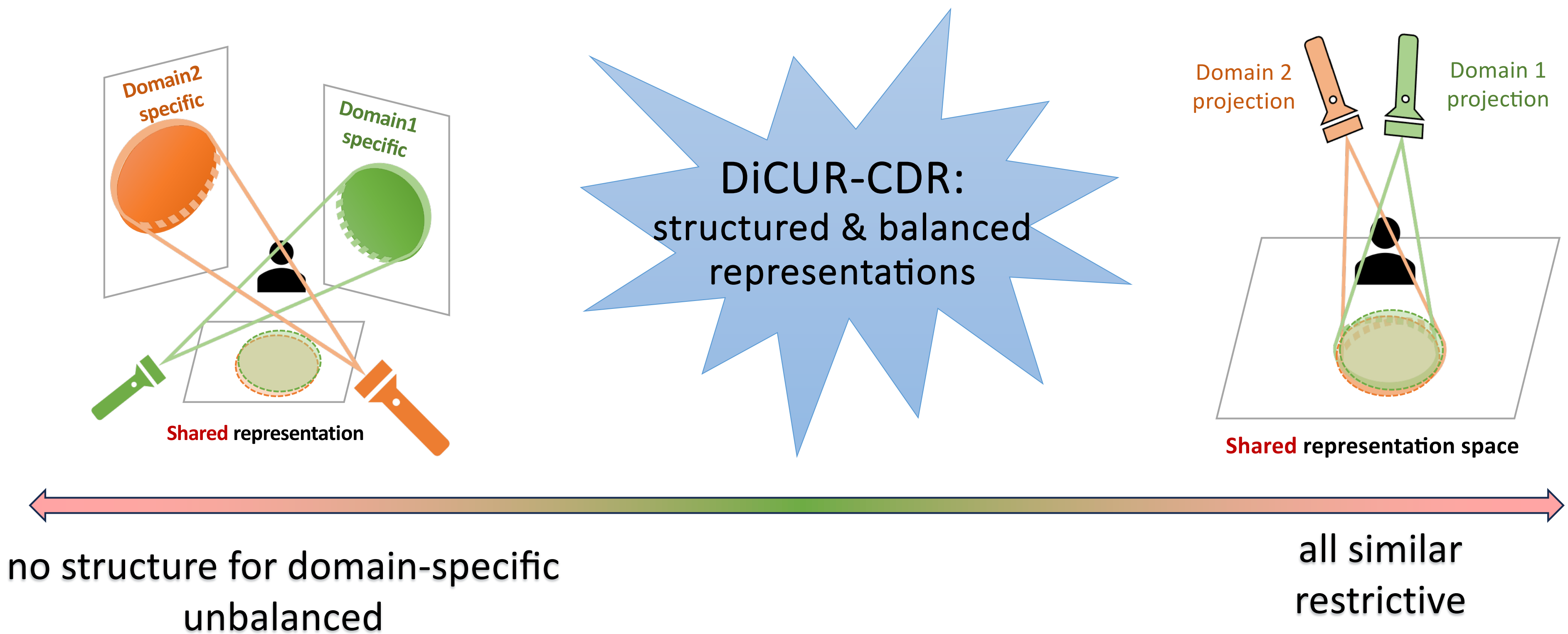
- Over-restriction of domain-shared representations
- Free representations of domain disparities

- Pushing most of the information into domain-specific representations
- Can't differentiate Domain-specific and domain-shared representations



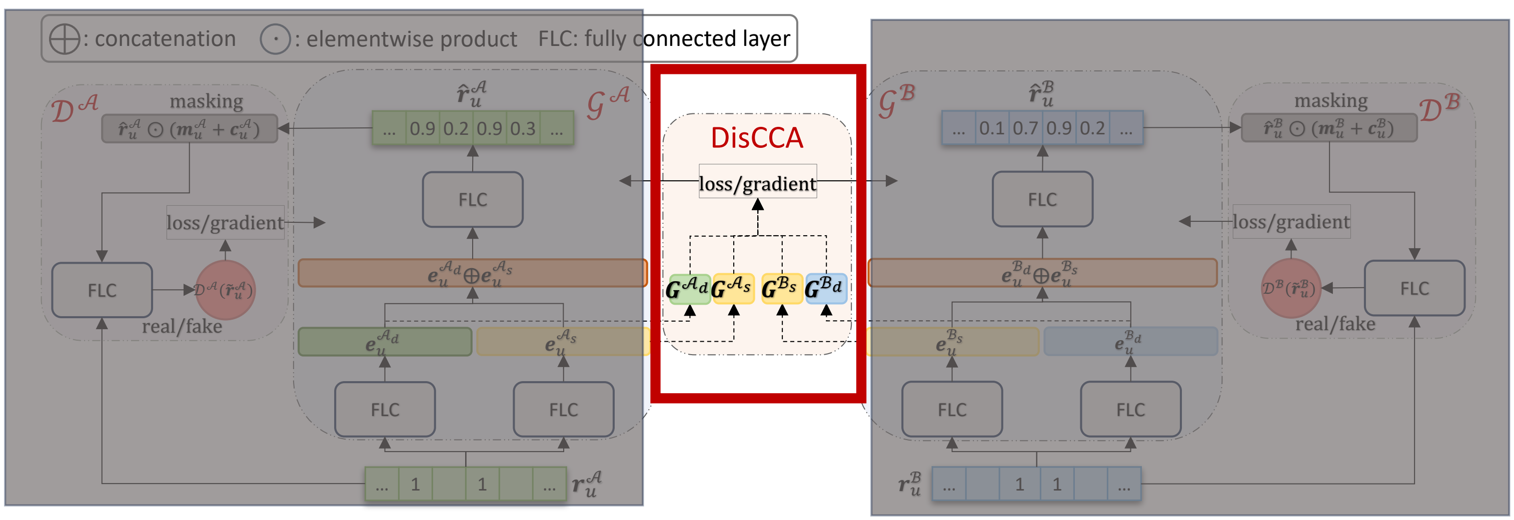
T-SNE of DisenCDR [Cao et al., 22] User Representations

Our solution: DiCUR-CDR



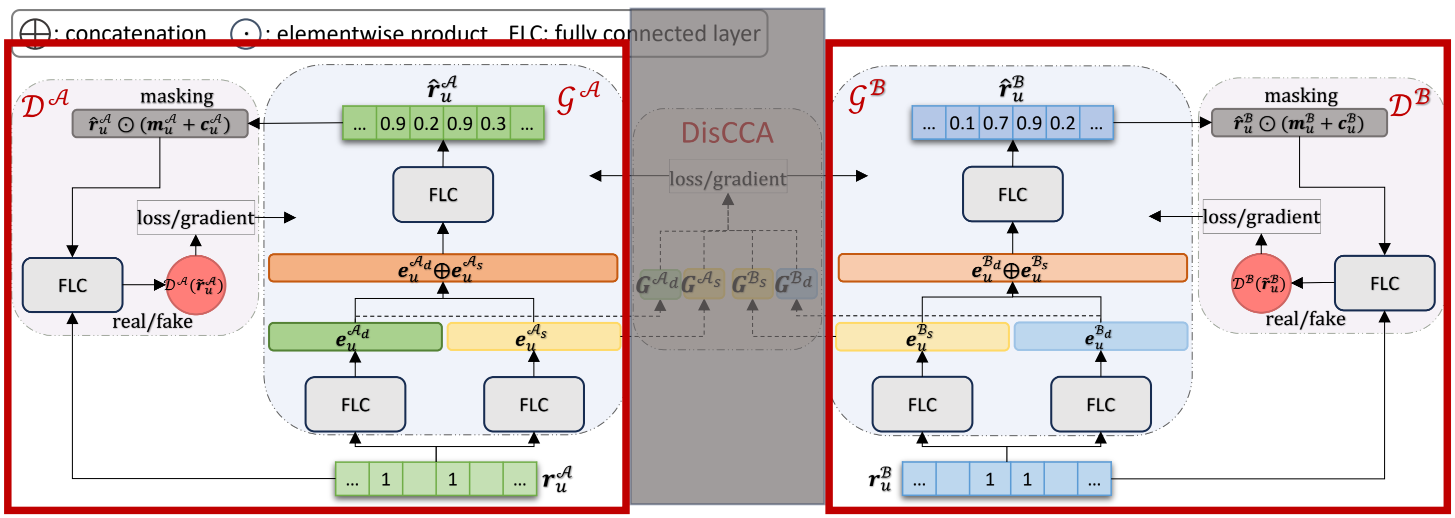
Discerning Canonical User Representation Learning for Cross-Domain Recommendation (DiCUR-CDR)

- Discerning Canonical Correlation (DisCCA) user representation learning
- Generative adversarial learning to model user preferences and generate recommendations

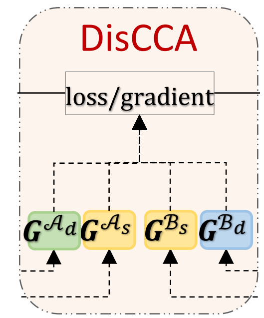


Discerning Canonical User Representation Learning for Cross-Domain Recommendation (DiCUR-CDR)

- Discerning Canonical Correlation (DisCCA) user representation learning
- Generative adversarial learning to model user preferences and generate recommendations

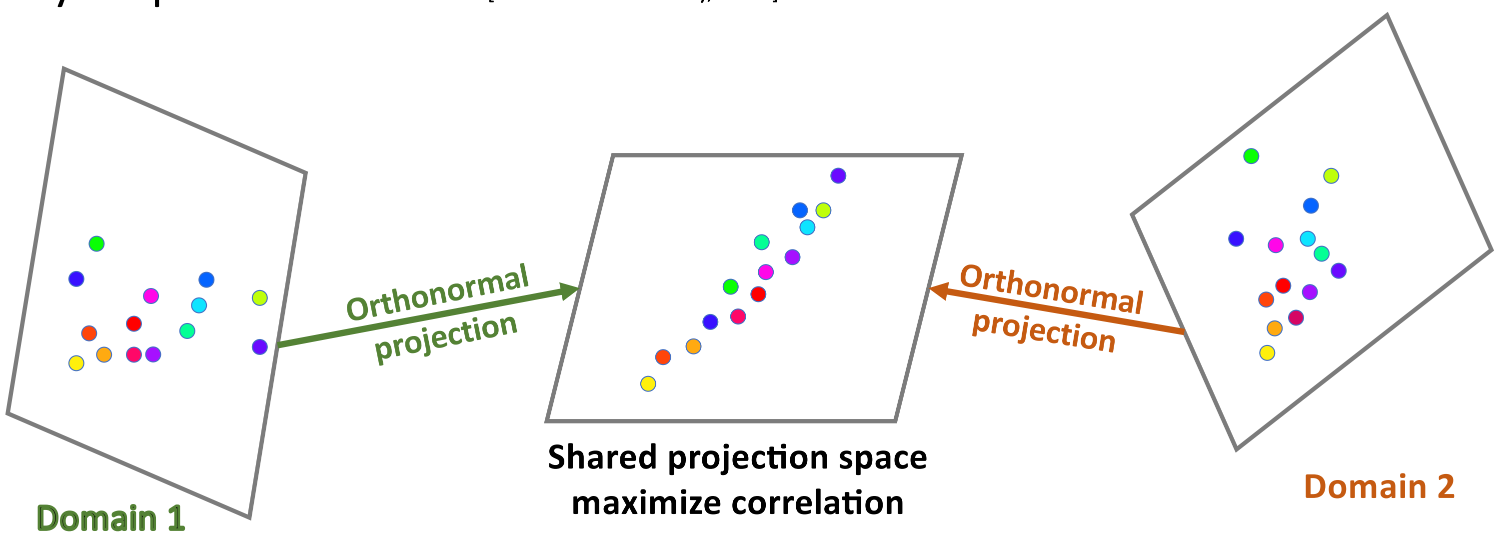


Discerning Canonical Correlation - Background



Canonical Correlation Analysis (CCA) [Michel van de Velden, 2011]

- Maximize the linear correlation between multivariate variables
- Keep an orthonormal projection space
- Only maps “similarities” [Sahebi & Brusilovsky, 2016]

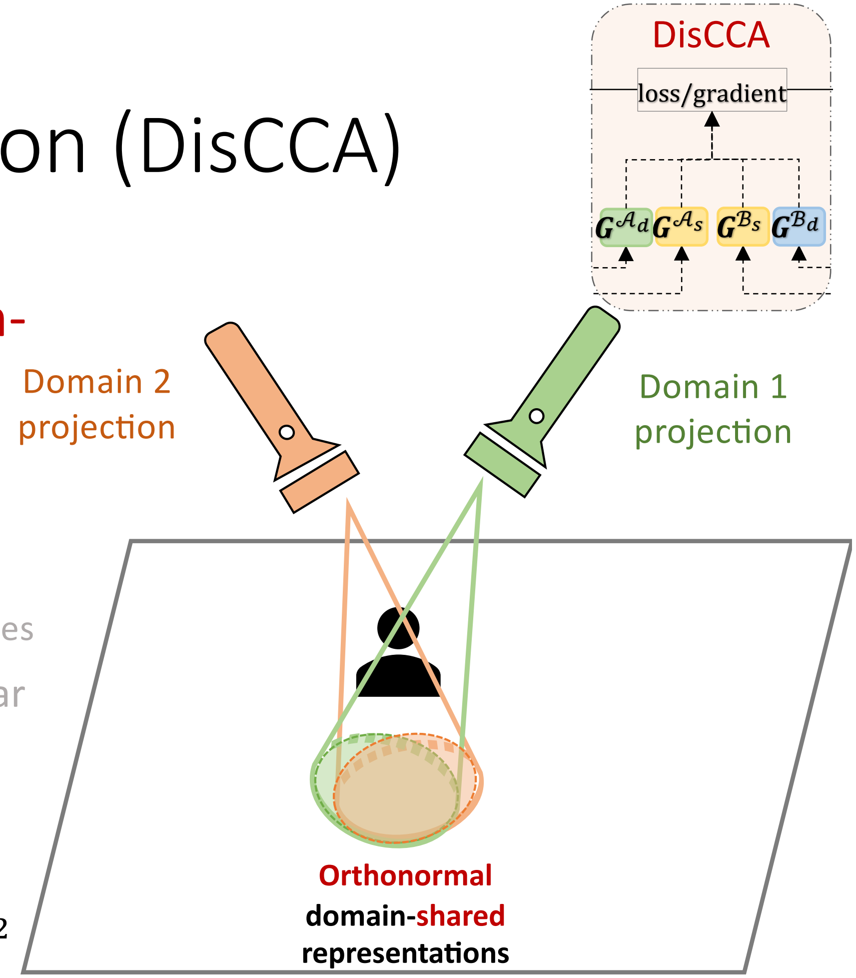


Discerning Canonical Correlation (DisCCA)

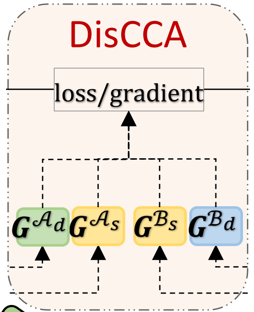
- Maximize the correlation between domain-shared user representations
    - Keep an orthonormal projection space
- Maximize the disparities between domain-specific representations
    - Structured: Using the same projections as domain-shared ones
- Keep the domain-shared and domain-specific projections far apart

$$G^{\mathcal{A}_s} = LRelu\left(\boldsymbol{M}^{\mathcal{A}}\boldsymbol{e}^{\mathcal{A}_s}\right) \qquad G^{\mathcal{B}_s} = LRelu\left(\boldsymbol{M}^{\mathcal{B}}\boldsymbol{e}^{\mathcal{B}_s}\right)$$

$$\mathcal{L}_s = \left\|G^{\mathcal{A}_s} - G^{\mathcal{B}_s}\right\|^2 + \left\|{G^{\mathcal{A}_s}}^T G^{\mathcal{A}_s} - I\right\|^2 + \left\|{G^{\mathcal{B}_s}}^T G^{\mathcal{B}_s} - I\right\|^2$$



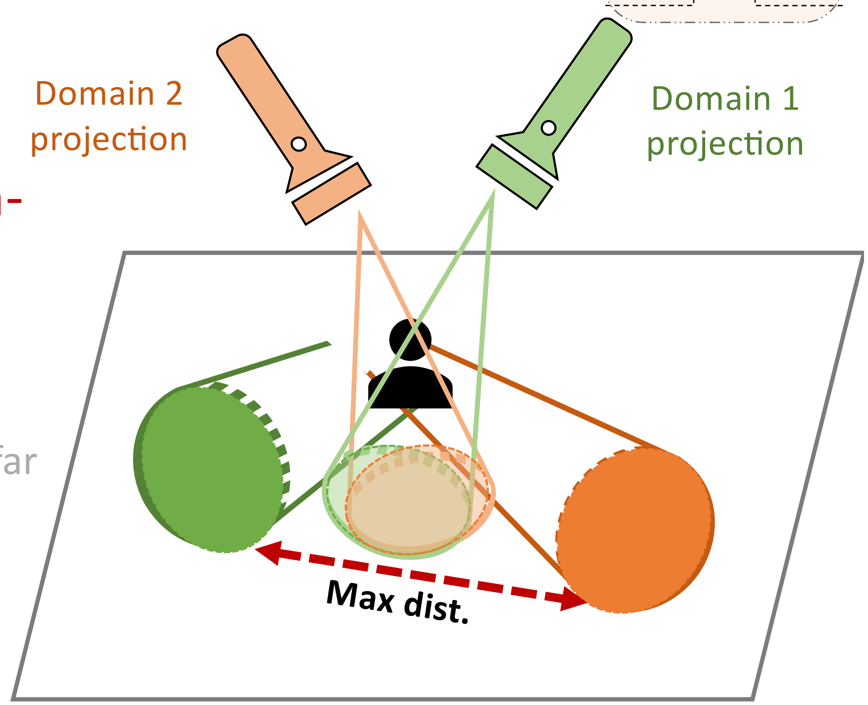
Discerning Canonical Correlation (DisCCA)

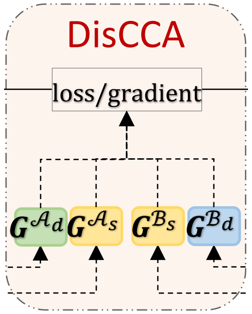


- Maximize the correlation between domain-shared user representations
  - Keep an orthonormal projection space
- Maximize the disparities between domain-specific representations
  - Structured: Using the same correlation projections as domain-shared ones
- Keep the domain-shared and domain-specific projections far apart

$$G^{\mathcal{A}_d} = LRelu\left(\boldsymbol{M}^{\mathcal{A}}\boldsymbol{e}^{\mathcal{A}_d}\right) \qquad G^{\mathcal{B}_d} = LRelu(\boldsymbol{M}^{\mathcal{B}}\boldsymbol{e}^{\mathcal{B}_d})$$

$$\mathcal{L}_d = \cdots + \left\|G^{\mathcal{A}_d} - G^{\mathcal{B}_d}\right\|^2$$

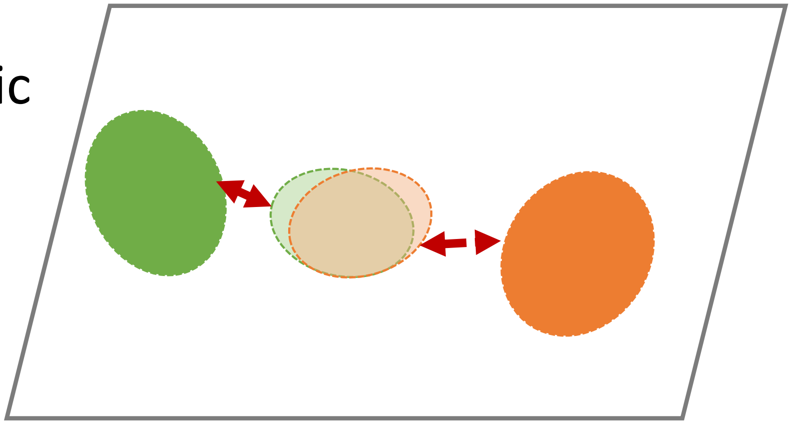




Discerning Canonical Correlation (DisCCA)

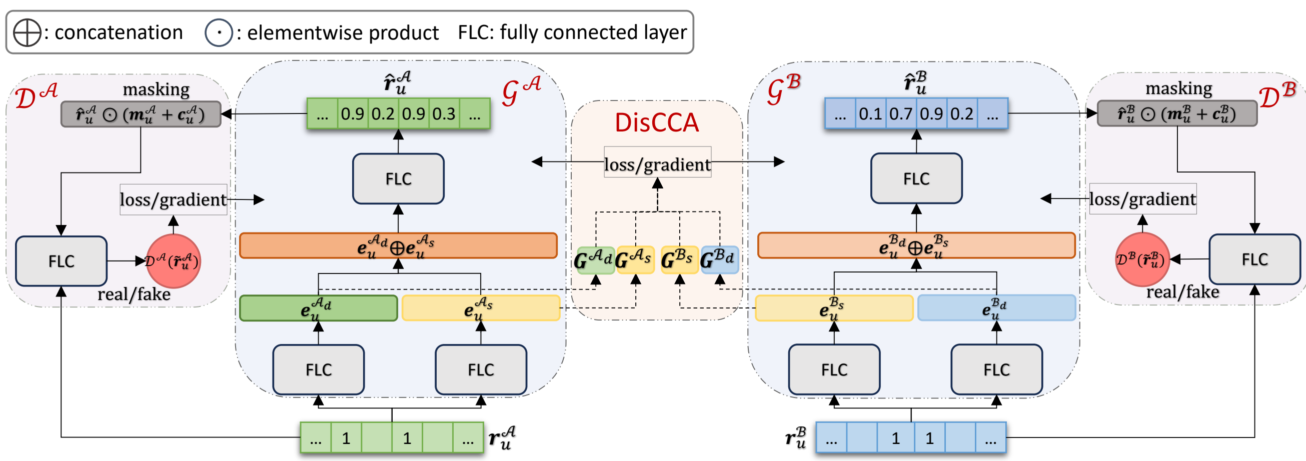
- Maximize the non-linear correlation between domain-shared user representations
  - Keep an orthonormal projection space
- Maximize the non-linear disparities between domain-specific representations
  - Structured: Using the same projections as domain-shared ones
- Keep the domain-shared and domain-specific projections far apart

$$\mathcal{L}_d = \left\|G^{\mathcal{A}_s} - G^{\mathcal{A}_d}\right\|^2 + \left\|G^{\mathcal{B}_s} - G^{\mathcal{B}_d}\right\|^2 + \dots$$

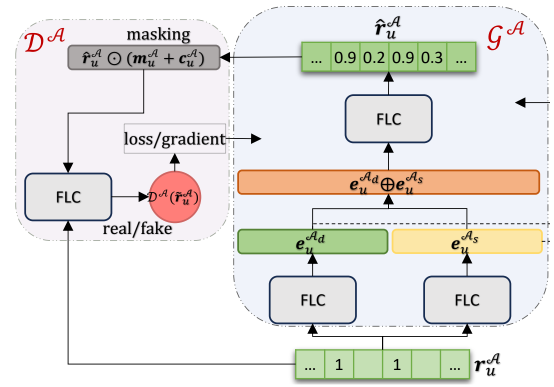


DiCUR-CDR

- GAN for each domain
  - Generator ($\mathcal{G}$): generates implicit feedback vectors
  - Discriminator ($\mathfrak{D}$): differentiates between the actual & generated feedback vectors



DiCUR-CDR Model – GAN



- Generator:
  - Learning Domain-Shared & domain-Specific Representation
    - Fully-connected layer:

$$\boldsymbol{e}_u^{\mathcal{A}_d} = \delta\left(\boldsymbol{W}_d^{\mathcal{A}} \boldsymbol{r}_u^{\mathcal{A}} + \boldsymbol{b}_d^{\mathcal{A}}\right) \qquad \boldsymbol{e}_u^{\mathcal{B}_d} = \delta\left(\boldsymbol{W}_d^{\mathcal{B}} \boldsymbol{r}_u^{\mathcal{B}} + \boldsymbol{b}_d^{\mathcal{B}}\right)$$

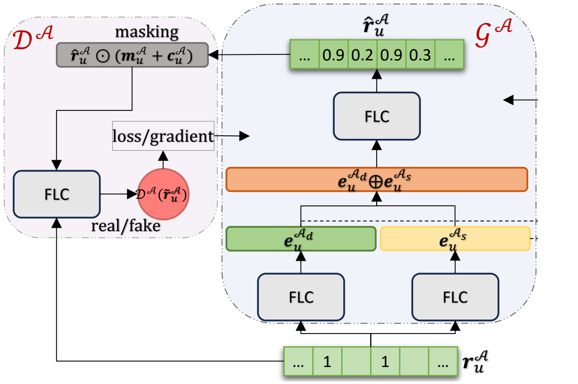
$$\boldsymbol{e}_u^{\mathcal{A}_s} = \delta\left(\boldsymbol{W}_s^{\mathcal{A}} \boldsymbol{r}_u^{\mathcal{A}} + \boldsymbol{b}_s^{\mathcal{A}}\right) \qquad \boldsymbol{e}_u^{\mathcal{B}_s} = \delta\left(\boldsymbol{W}_s^{\mathcal{B}} \boldsymbol{r}_u^{\mathcal{B}} + \boldsymbol{b}_s^{\mathcal{B}}\right)$$

  - Recommendation Prediction
    - Fully-connected layer:

$$\hat{\boldsymbol{r}}_u^{\mathcal{A}} = \mathcal{G}^{\mathcal{A}}\left(\boldsymbol{r}_u^{\mathcal{A}}\right) = \boldsymbol{W}_o^{\mathcal{A}}[\boldsymbol{e}_u^{\mathcal{A}_d} \oplus \boldsymbol{e}_u^{\mathcal{A}_s}] + \boldsymbol{b}_o^{\mathcal{A}}$$

$$\hat{\boldsymbol{r}}_u^{\mathcal{B}} = \mathcal{G}^{\mathcal{B}}\left(\boldsymbol{r}_u^{\mathcal{B}}\right) = \boldsymbol{W}_o^{\mathcal{B}}[\boldsymbol{e}_u^{\mathcal{B}_d} \oplus \boldsymbol{e}_u^{\mathcal{B}_s}] + \boldsymbol{b}_o^{\mathcal{B}}$$

DiCUR-CDR Model – GAN



- Discriminator:
  - Estimating the probability of $\tilde{\boldsymbol{r}}_u^{\mathcal{A}}$ amd $\tilde{\boldsymbol{r}}_u^{\mathcal{B}}$ being real

$$\mathfrak{D}^{\mathcal{A}}\left(\tilde{\boldsymbol{r}}_u^{\mathcal{A}}\right) = \delta(\boldsymbol{W}_n^{\mathcal{A}}[\boldsymbol{r}_u^{\mathcal{A}} \oplus (\tilde{\boldsymbol{r}}_u^{\mathcal{A}} \odot \boldsymbol{m}_u^{\mathcal{A}})] + \boldsymbol{b}_n^{\mathcal{A}})$$

$$\mathfrak{D}^{\mathcal{B}}\left(\tilde{\boldsymbol{r}}_u^{\mathcal{B}}\right) = \delta(\boldsymbol{W}_n^{\mathcal{B}}[\boldsymbol{r}_u^{\mathcal{B}} \oplus (\tilde{\boldsymbol{r}}_u^{\mathcal{B}} \odot \boldsymbol{m}_u^{\mathcal{B}})] + \boldsymbol{b}_n^{\mathcal{B}})$$

DiCUR-CDR Model – Learning

- Negative Sampling
  - Randomly selecting ω (hyperparameter) portion of the unobserved items as the negative items ($N_u^{\mathcal{A}}$ and $N_u^{\mathcal{B}}$)
  - Generating values close to 1 for the observed items
  - Producing low values for the negative ones.

- Generators loss:

$$\tilde{\mathcal{L}}_{\mathcal{G}^{\mathcal{A}}} = \sum_u \log\left(1 - \mathfrak{D}^{\mathcal{A}}\left(\tilde{\boldsymbol{r}}_u^{\mathcal{A}}\right)\right) + \lambda_s \mathcal{L}_s + \lambda_d \mathcal{L}_d + \lambda_\theta \left\|\theta_{\mathcal{G}^{\mathcal{A}}}\right\|^2 + \lambda_n \left\|(\boldsymbol{r}_u^{\mathcal{A}} - \hat{\boldsymbol{r}}_u^{\mathcal{A}}) \odot (\boldsymbol{m}_u^{\mathcal{A}} + \boldsymbol{c}_u^{\mathcal{A}})\right\|^2$$

- Discriminators loss:

$$\mathcal{L}_{\mathfrak{D}^{\mathcal{A}}} = -\sum_u ((\log\left(\mathfrak{D}^{\mathcal{A}}\left(\boldsymbol{r}_u^{\mathcal{A}}\right)\right) + \log\left(1 - \mathfrak{D}^{\mathcal{A}}\left(\tilde{\boldsymbol{r}}_u^{\mathcal{A}}\right)\right)) + \lambda_\theta \left\|\theta_{\mathfrak{D}^{\mathcal{A}}}\right\|^2$$

Experiments

- Five sets of experiments
  - Recommendation Prediction
  - Ablation Studies
  - Sensitivity Analysis
  - Cold-start Analysis
  - Learned User Representations Visualization

Comparison with Baselines

Dataset	Amazon						Yelp					
Domain	Movies_and_TV			Video_Games			Restaurants			Shopping		
Metrics	HR	NDCG	MRR	HR	NDCG	MRR	HR	NDCG	MRR	HR	NDCG	MRR
CDAE	0.33568**	0.24340**	0.20778**	0.27396**	0.17902**	0.14791**	0.42771**	0.27756**	0.22832**	0.25556**	0.17131**	0.14367**
IRGAN	0.33157**	0.23693**	0.23222**	0.31791**	0.22461**	0.20295**	0.50491**	0.34964**	0.18901**	0.26976**	0.17966**	0.08962**
DASO	0.33214**	0.23944**	0.20377**	0.27104**	0.17747**	0.14681**	0.42239**	0.28516**	0.24051**	0.26000**	0.17447**	0.14645**
CDAE w/ M	0.33682**	0.24455**	0.20892**	0.27380**	0.17975**	0.14895**	0.41352**	0.26338**	0.21457**	0.27420**	0.17848**	0.14711**
IRGAN w/ M	0.31626**	0.22674**	0.22717**	0.30294**	0.22136**	0.07760**	0.50211**	0.28131**	0.14198**	0.26799**	0.17900**	0.08785**
DASO w/ M	0.32930**	0.23862**	0.20363**	0.26440**	0.17027**	0.13954**	0.42061**	0.28853**	0.24503**	0.18191**	0.11756**	0.09665**
CoNet	0.37387**	0.32448**	0.27704**	0.37507**	0.24989**	0.20206**	0.49308**	0.26150**	0.17489**	0.24183**	0.09057**	0.07780**
DDTCDR	0.41748**	0.34188**	0.29904**	0.38964**	0.27384**	0.23580**	0.54020**	0.41304**	0.36528**	0.32191**	0.20618**	0.16825**
ETL	0.43038**	0.30931**	0.37091**	0.40373**	0.27032**	0.35928**	0.56406**	0.43037*	0.35102**	0.37265**	0.26387**	0.23093**
DisenCDR	0.44062**	0.33783**	0.18056**	0.47047**	0.32741**	0.21987**	0.55993**	0.38873**	0.33760**	0.44191**	0.30617**	0.29475**
CAT-ART	0.50733*	0.40527**	0.37043**	0.49988*	0.38892**	0.35217**	0.59800*	0.42238**	0.36446**	0.51718*	0.37004*	0.32146*
DiCUR-CDR	**0.51235**	**0.42119**	**0.39360**	**0.51006**	**0.40646**	**0.37303**	**0.61548**	**0.43752**	**0.37968**	**0.52833**	**0.37740**	**0.32785**

Recommendation Performance Prediction results, $**$ and $*$ indicate t-test $p - \text{value} < 0.05$ and $p - \text{value} < 0.1$

- DiCUR-CDR outperforms all baselines across both datasets and domains
  - Demonstrating the feasibility of incorporating DisCCA

Ablation Studies: all discerning representations matter

Dataset	Amazon						Yelp					
Domain	Movies_and_TV			Video_Games			Restaurants			Shopping		
Metrics	HR	NDCG	MRR	HR	NDCG	MRR	HR	NDCG	MRR	HR	NDCG	MRR
DiCUR-CDR w/o S	0.47833	0.40129	0.37589	0.50196	0.40057	0.36784	0.60478	0.43682	0.38292	0.51604	0.37253	0.32563
DiCUR-CDR w/o D	0.47914	0.40216	0.37679	0.50344	0.40116	0.36771	0.60205	0.42718	0.37165	0.51877	0.37640	0.32631
DiCUR-CDR w/o S&D	0.47578	0.39918	0.37381	0.49627	0.39939	0.36730	0.60000	0.42964	0.37513	0.51809	0.37242	0.32579
DiCUR-CDR w/o GAN (MLP w/ DisCCA)	0.47253	0.39982	0.37671	0.44026	0.36168	0.33622	0.5372	0.39011	0.34272	0.4942	0.36145	0.31926
DiCUR-CDR	**0.51235**	**0.42119**	**0.39360**	**0.51006**	**0.40646**	**0.37303**	**0.61548**	**0.43752**	**0.37968**	**0.52833**	**0.37740**	**0.32785**

Ablation study results

- Removing either $\mathcal{L}_S$ or $\mathcal{L}_D$, or both, leads to a performance decline
  - Distinguishing between domain-shared & domain-specific representation, ensuring distinct domain-specific representations, are all crucial

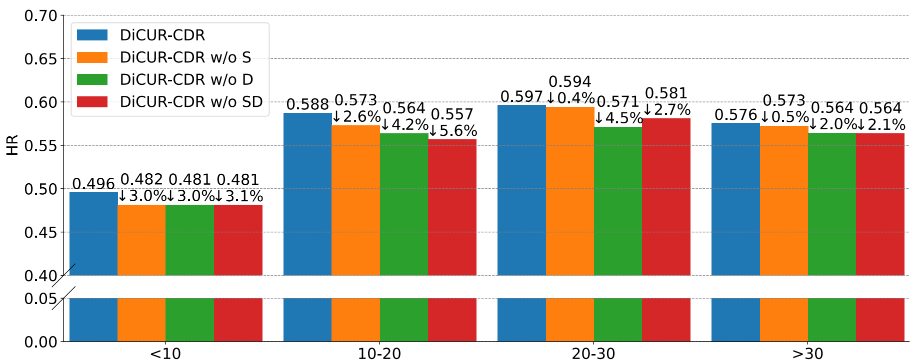
Ablation Studies: GAN matters

Dataset	Amazon						Yelp					
Domain	Movies_and_TV			Video_Games			Restaurants			Shopping		
Metrics	HR	NDCG	MRR	HR	NDCG	MRR	HR	NDCG	MRR	HR	NDCG	MRR
DiCUR-CDR w/o S	0.47833	0.40129	0.37589	0.50196	0.40057	0.36784	0.60478	0.43682	0.38292	0.51604	0.37253	0.32563
DiCUR-CDR w/o D	0.47914	0.40216	0.37679	0.50344	0.40116	0.36771	0.60205	0.42718	0.37165	0.51877	0.37640	0.32631
DiCUR-CDR w/o S&D	0.47578	0.39918	0.37381	0.49627	0.39939	0.36730	0.60000	0.42964	0.37513	0.51809	0.37242	0.32579
DiCUR-CDR w/o GAN (MLP w/ DisCCA)	0.47253	0.39982	0.37671	0.44026	0.36168	0.33622	0.5372	0.39011	0.34272	0.4942	0.36145	0.31926
DiCUR-CDR	**0.51235**	**0.42119**	**0.39360**	**0.51006**	**0.40646**	**0.37303**	**0.61548**	**0.43752**	**0.37968**	**0.52833**	**0.37740**	**0.32785**

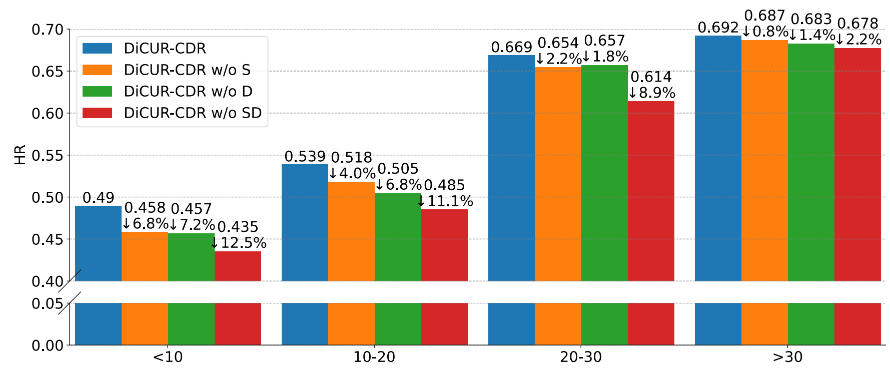
Ablation study results

- DiCUR-CDR w/o GAN exhibits the lowest performance
  - Highlighting the effectiveness of using GAN

Cold-start Analysis



Restaurants domain

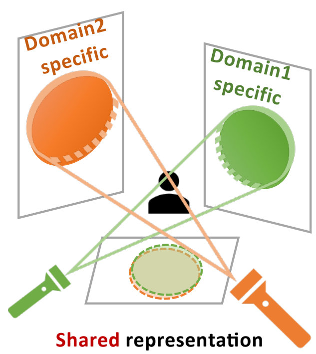


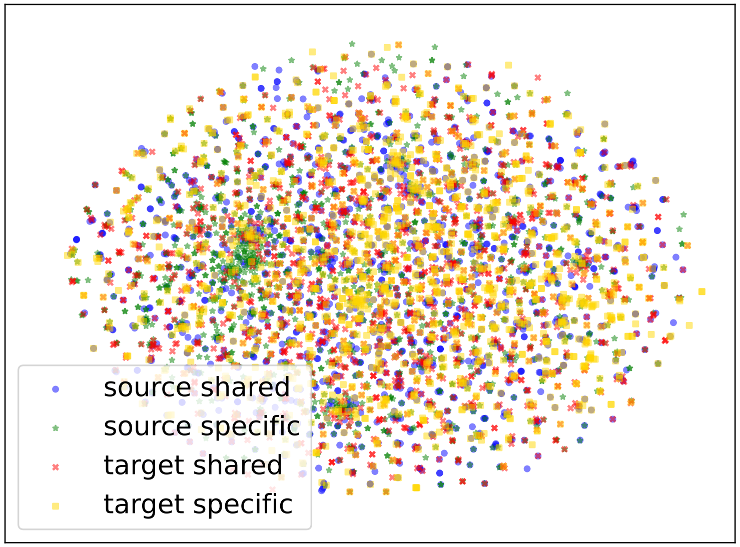
Shopping domain

Recommendation results (HR) for users with different numbers of interactions on the Yelp dataset

- The largest recommendation prediction improvements happen for
  - users with fewer interactions (groups < 10 and 10 – 20)
  - sparser domain

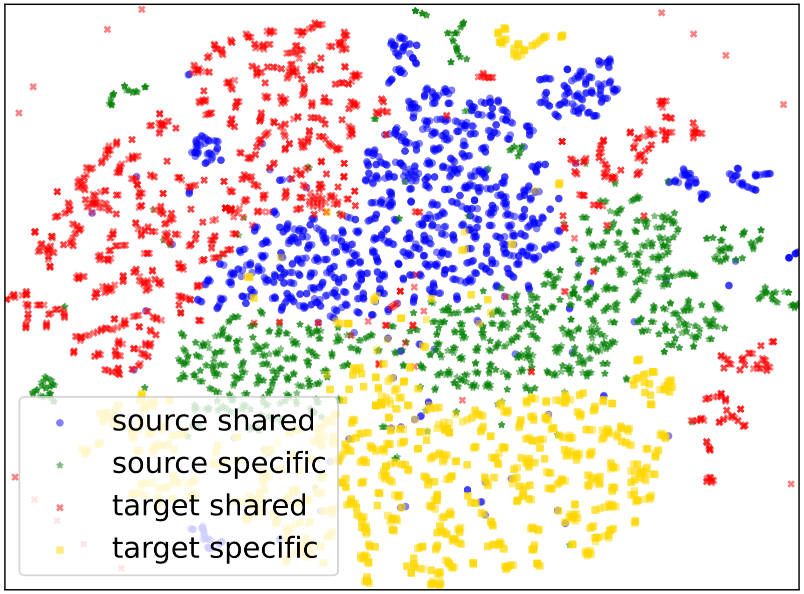
Baseline Learned User Representations





T-SNE of DisenCDR [Cao et al., 22] User Representations

DiCUR-CDR Learned User Representations



T-SNE of DiCUR-CDR User Representations

Pushing the domain-shared representations close to each other
Separating and differentiating between the domain-specific ones

Conclusions

- Proposed Discerning Canonical User Representation Learning for Cross-Domain Recommendation (DiCUR-CDR)
- Introduced Discerning Canonical Correlation (DisCCA) user representation learning
- Emphasized the effectiveness of
  - Capturing both cross-domain and within-domain user preferences
  - Discerning the domain-specific representations in a structured way

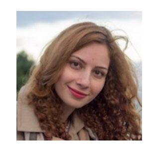
UNIVERSITY AT ALBANY

State University of New York

PersAI onalized

Thank you!

Q & A



szhao2@albany.edu

ssahebi@albany.edu

Our code and sample data are available at GitHub:
https://github.com/persai-lab/2024-RecSys-DiCUR-CDR

This paper is based upon work supported by the National Science Foundation under Grant No. 2047500.