Recommendation for New Users and New Items via Randomized Training and Mixture-of-Experts Transformation

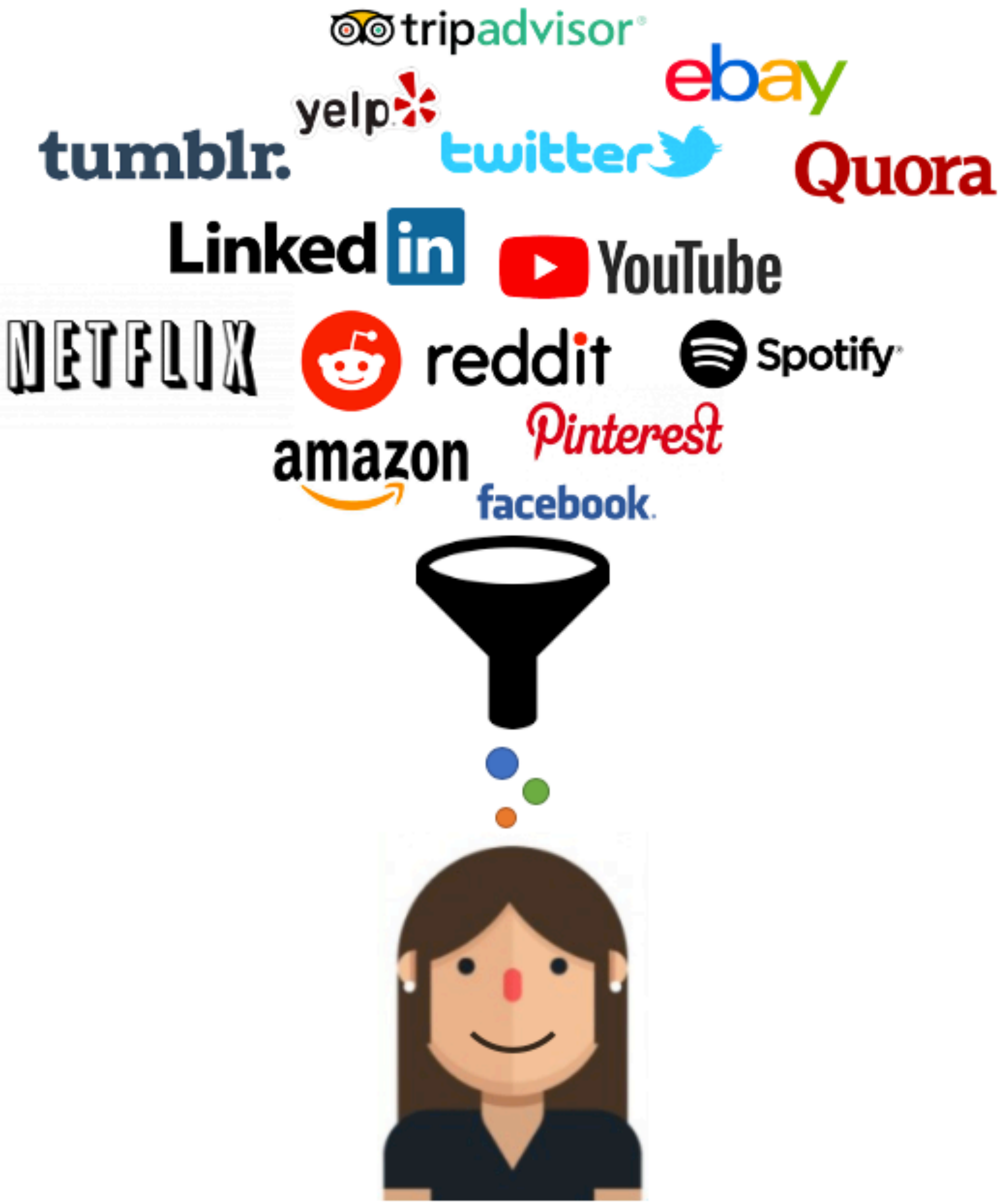
Ziwei Zhu, Shahin Sefati*, Parsa Saadatpanah*, and James Caverlee

Texas A&M University

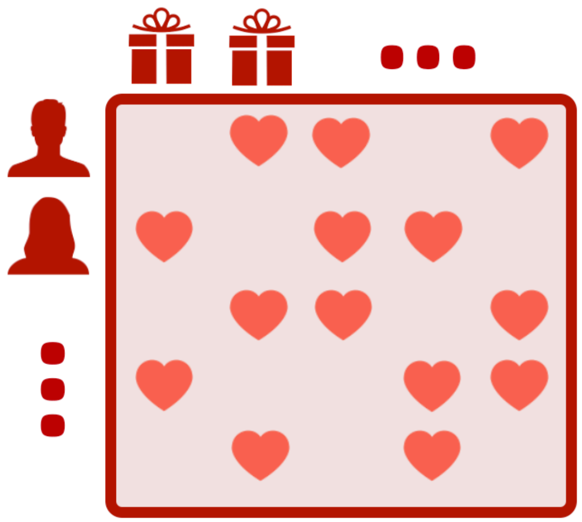
*Comcast Applied AI Lab

SIGIR 2020
XI'AN·CHINA

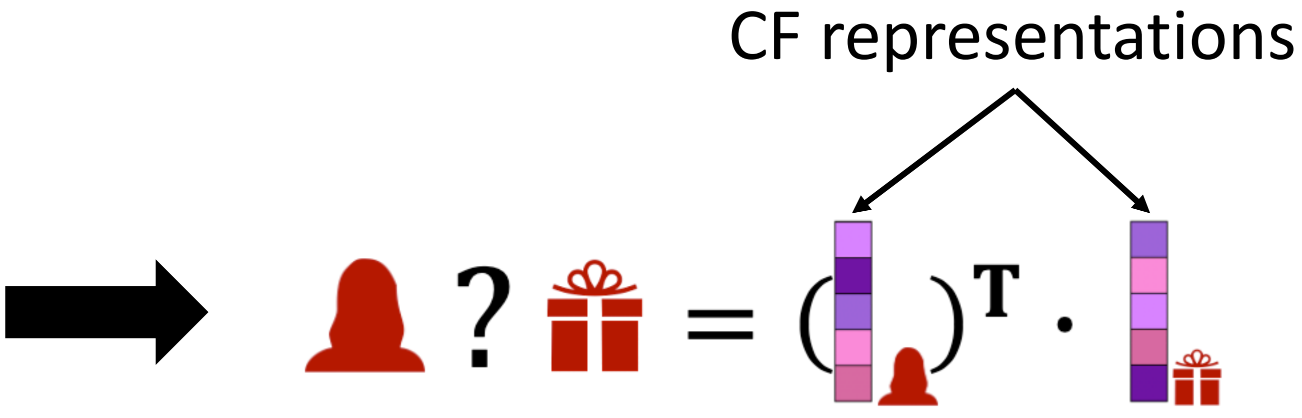
Recommenders – essential conduits



Recommenders with warm start users and items

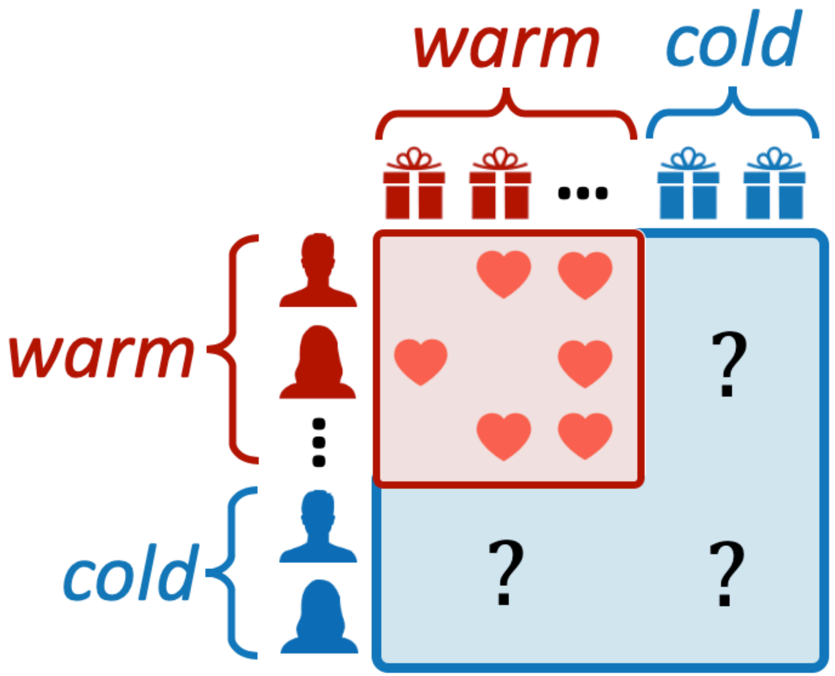


user-item interactions



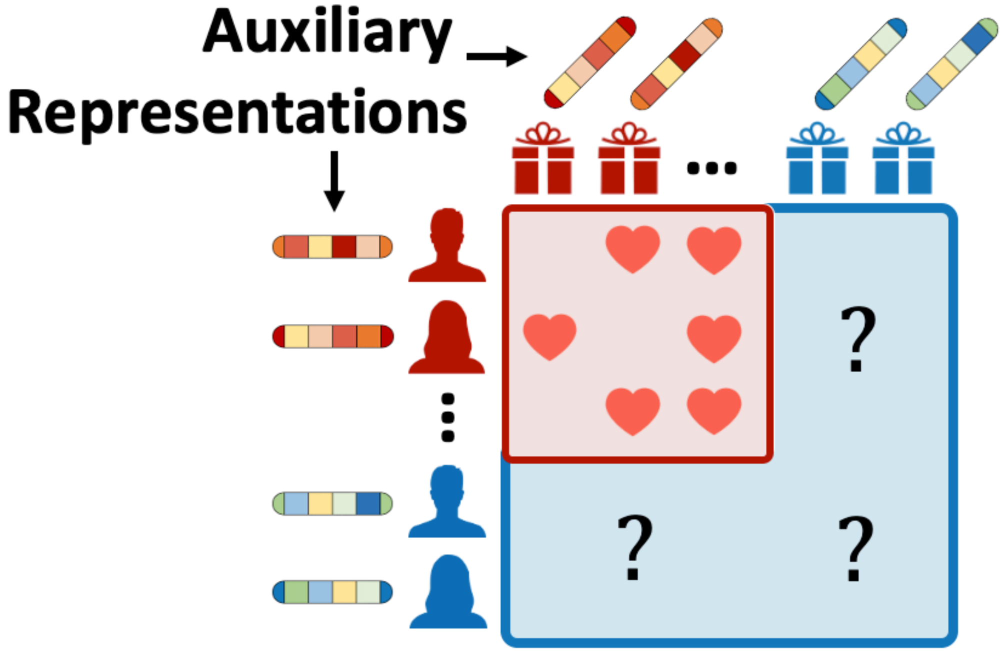
collaborative filtering (CF)

Cannot work for cold start users and items

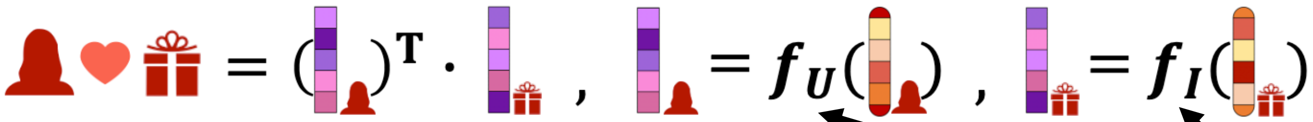


No historical record for and , thus cannot learn CF representation and to predict ? .

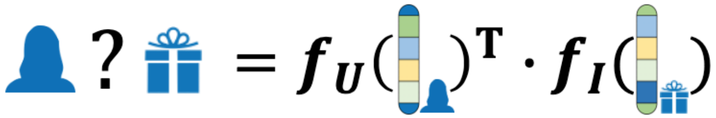
With the help of auxiliary information



Train with warm start users and items:



Infer for cold start users and items:

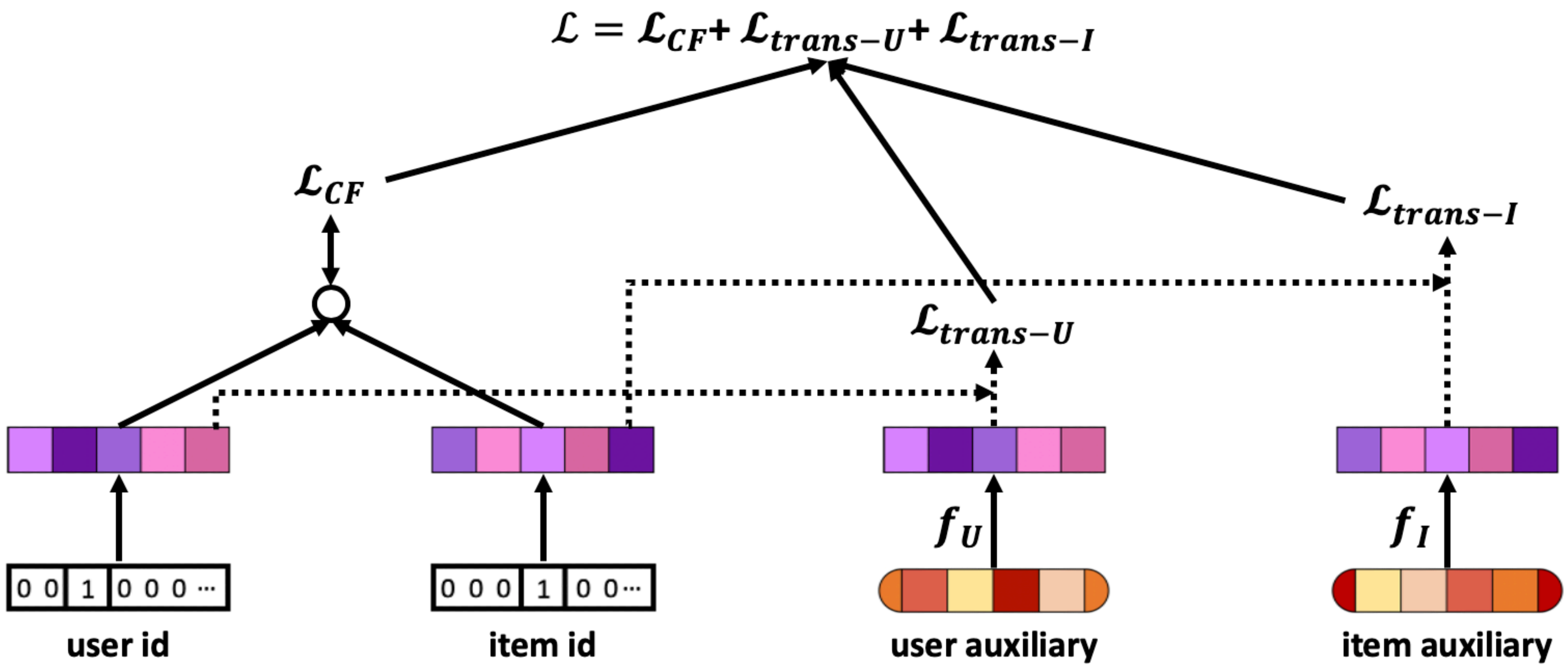


transformation functions

Two categories of methods

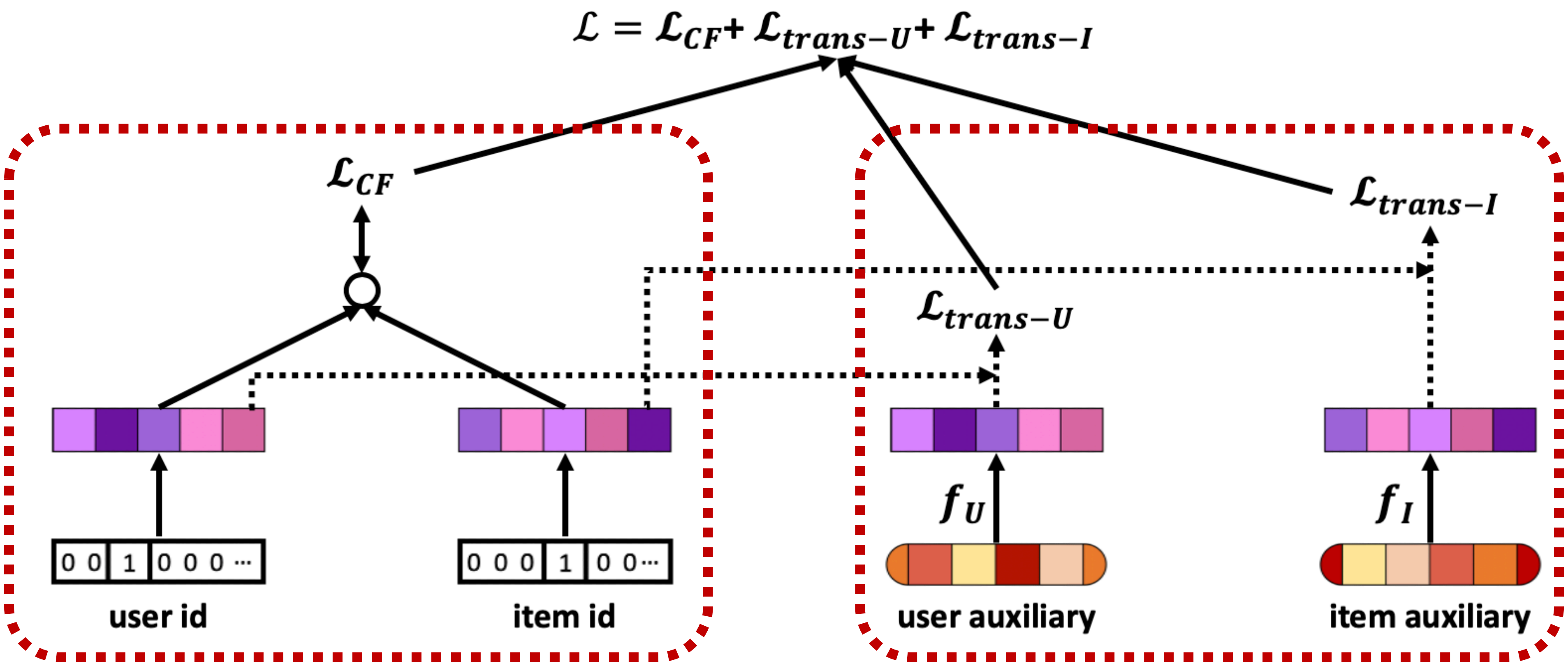
- Separate-training method
- Joint-training method

Separate-training method



Separately train a CF component and an 'auxiliary to CF' transformation component.

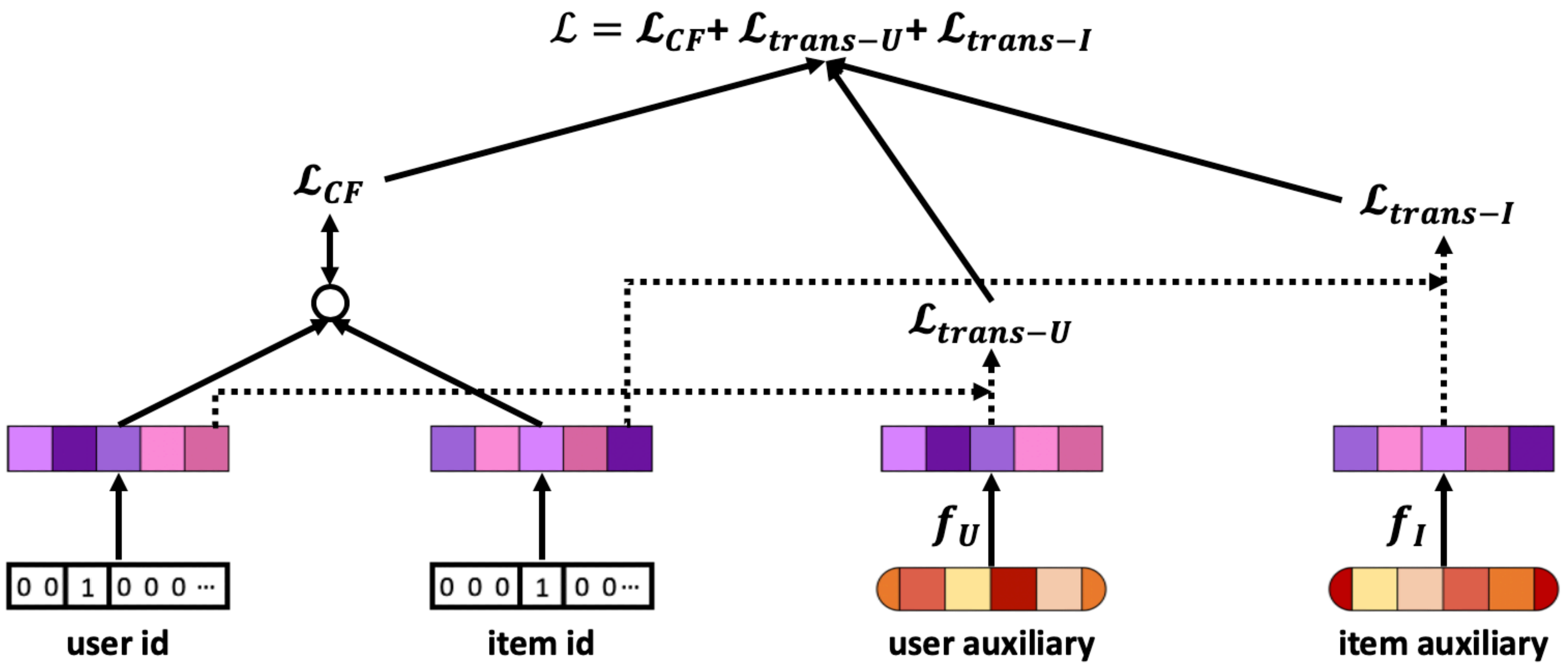
Separate-training method



Learn CF representations for users and items from user-item interactions

Learn 'auxiliary to CF representation' transformation

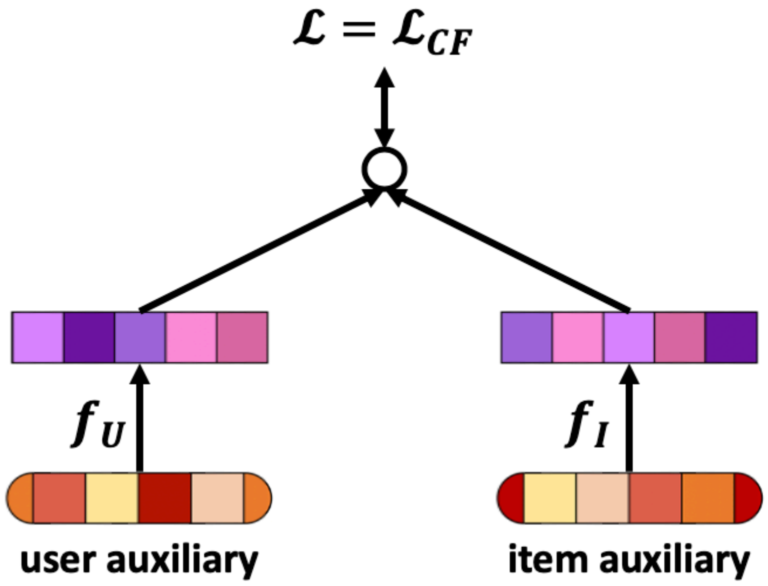
Separate-training method -- error superimposition problem



Pros: Learn high-quality CF representations.

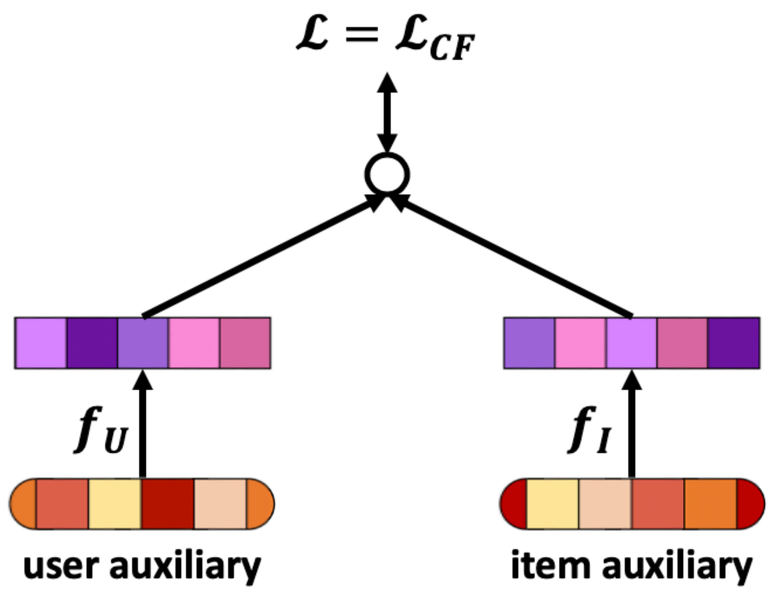
Cons: The final cold start recommendation error is $\mathcal{L}_{CF} + \mathcal{L}_{trans}$ (***error superimposition***).

Joint-training method



Train the CF component and the 'auxiliary to CF' component in the same back-propagation flow.

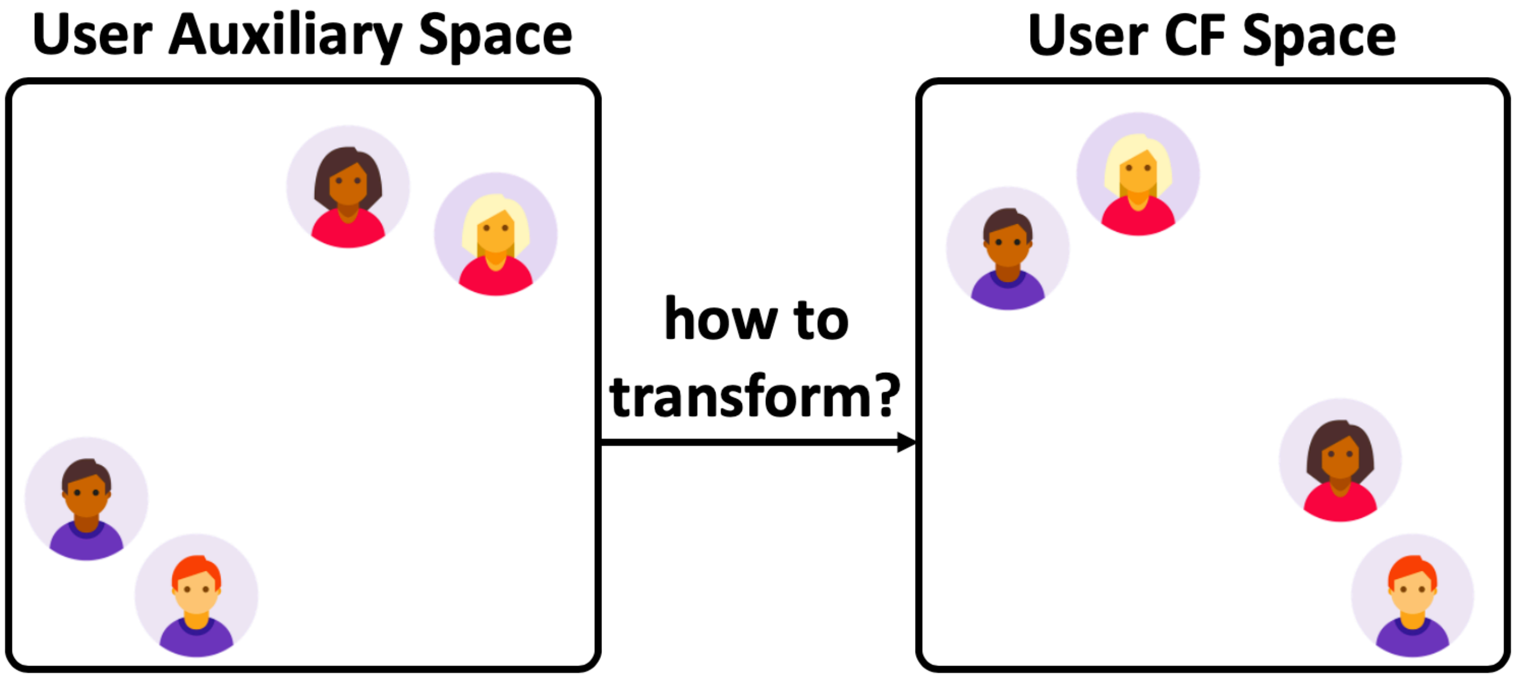
Joint-training method -- ineffective learning problem



Pros: No error superimposition problem.

Cons: the first few layers of f_U and f_I are far from output layer, leading to ***ineffective learning*** of the transformation process.

Unified transformation problem



A ***unified transformation*** function will hold the relationship between users (items) in the auxiliary space to the transformed CF space, which is not always true in practice.

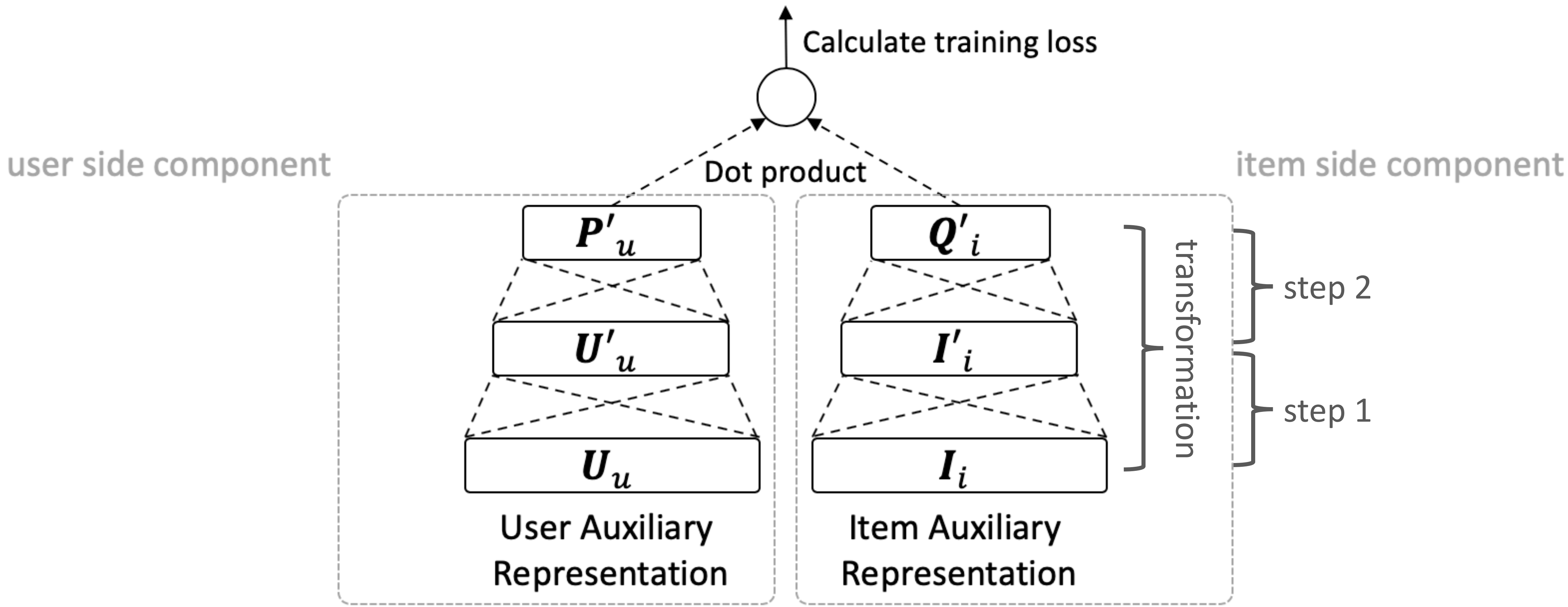
Motivation – three challenges

- Error superimposition problem
- Ineffective learning problem
- Unified transformation problem

Our proposal -- Heater

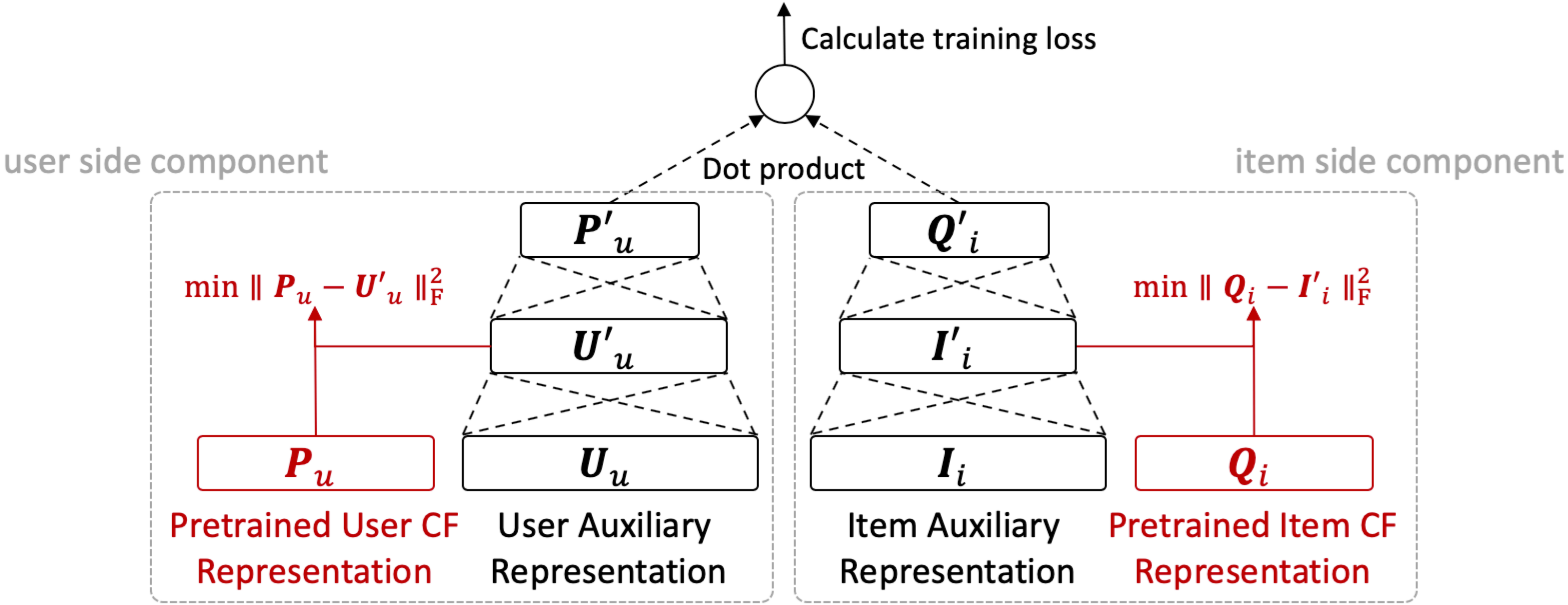
- Error superimposition problem – **a joint training based framework**
- Ineffective learning problem – **similarity constraint, randomized training**
- Unified transformation problem – **mixture-of-expert transformation**

Heater -- framework



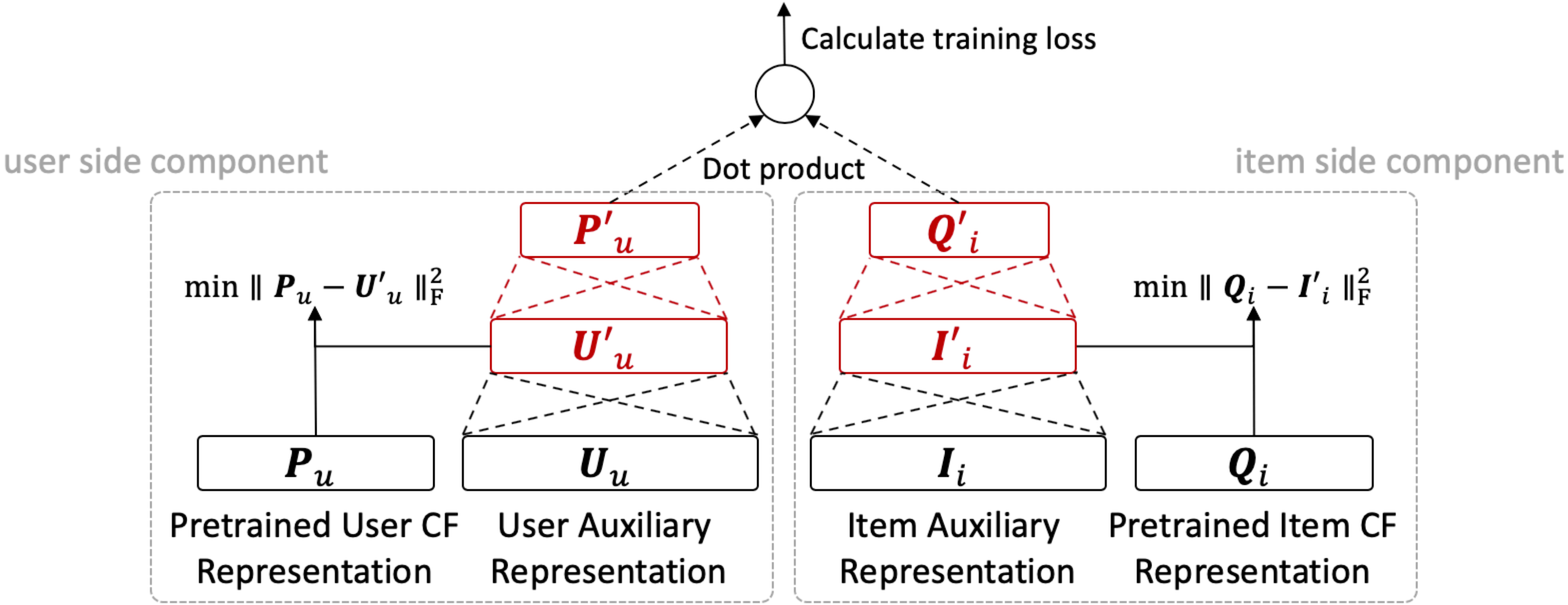
Joint-training method as the main framework to **avoid error superimposition problem**.

Heater – similarity constraint



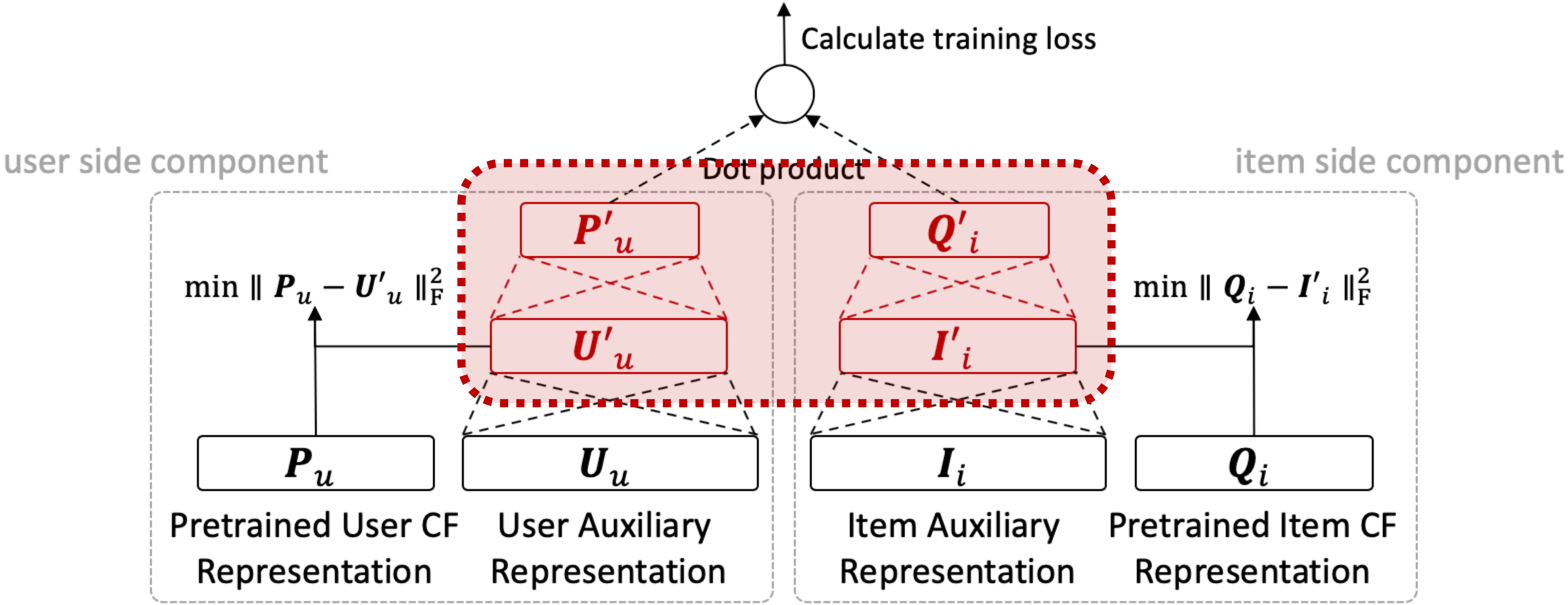
Add a similarity constraint in the middle to **improve the learning effectiveness for transformation functions**.

Heater – randomized training



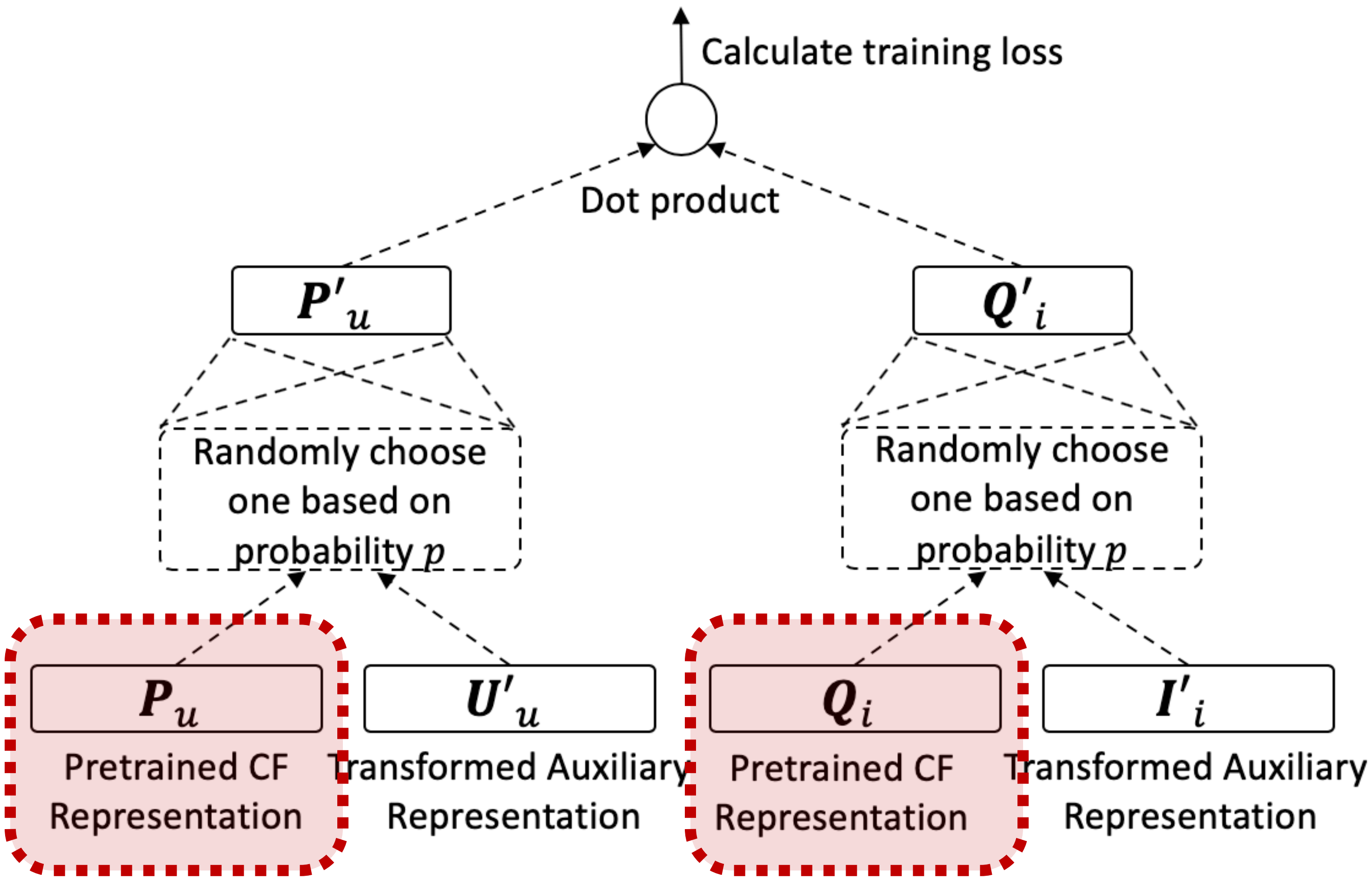
Still cannot guarantee the quality of U'_u and I'_i, especially for initial epochs of training, leading to ineffective training.

Heater – randomized training



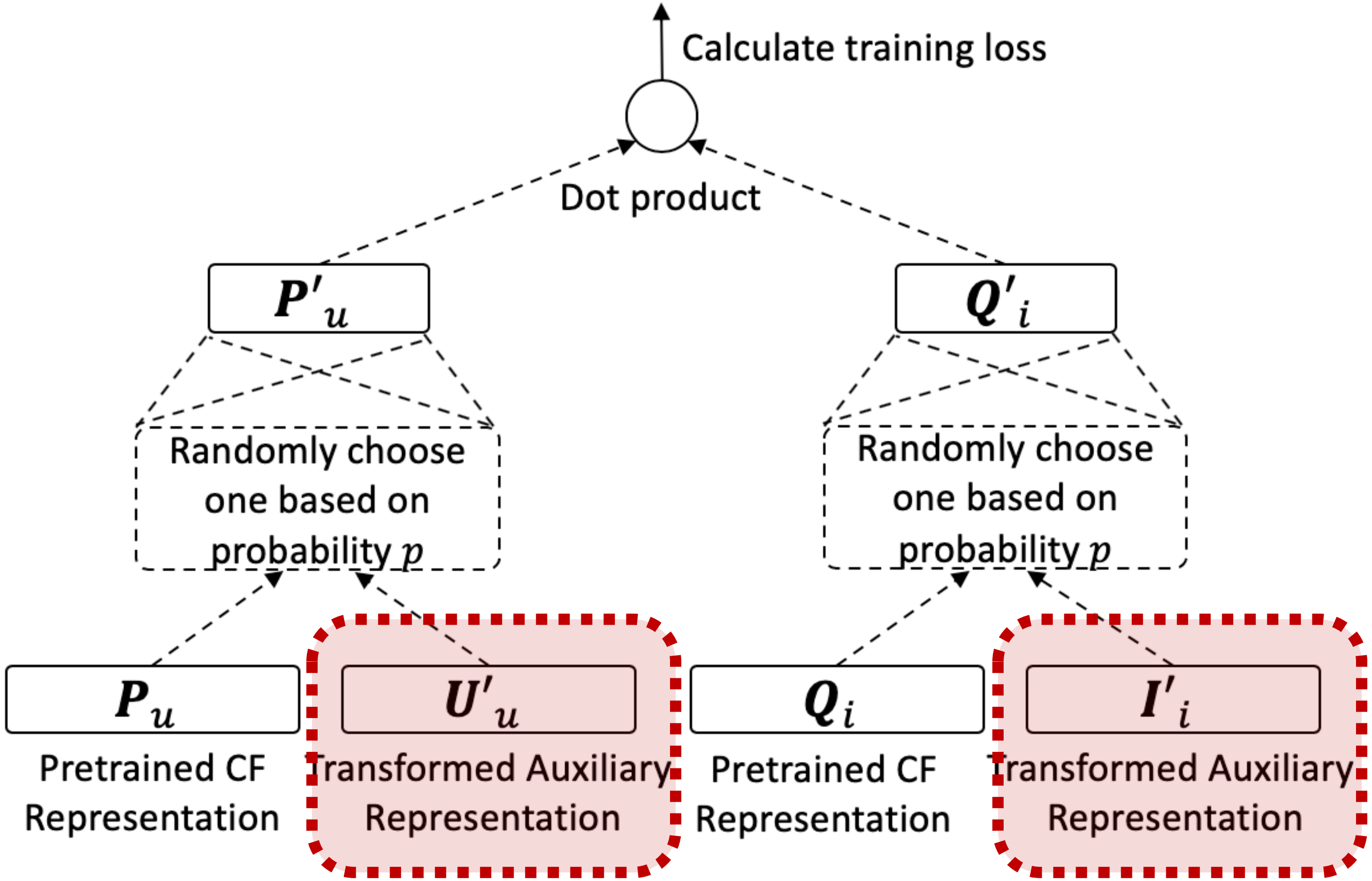
Still cannot promise the quality of U'_u and I'_i, especially for initial epochs of training, leading to ineffective training.

Heater – randomized training



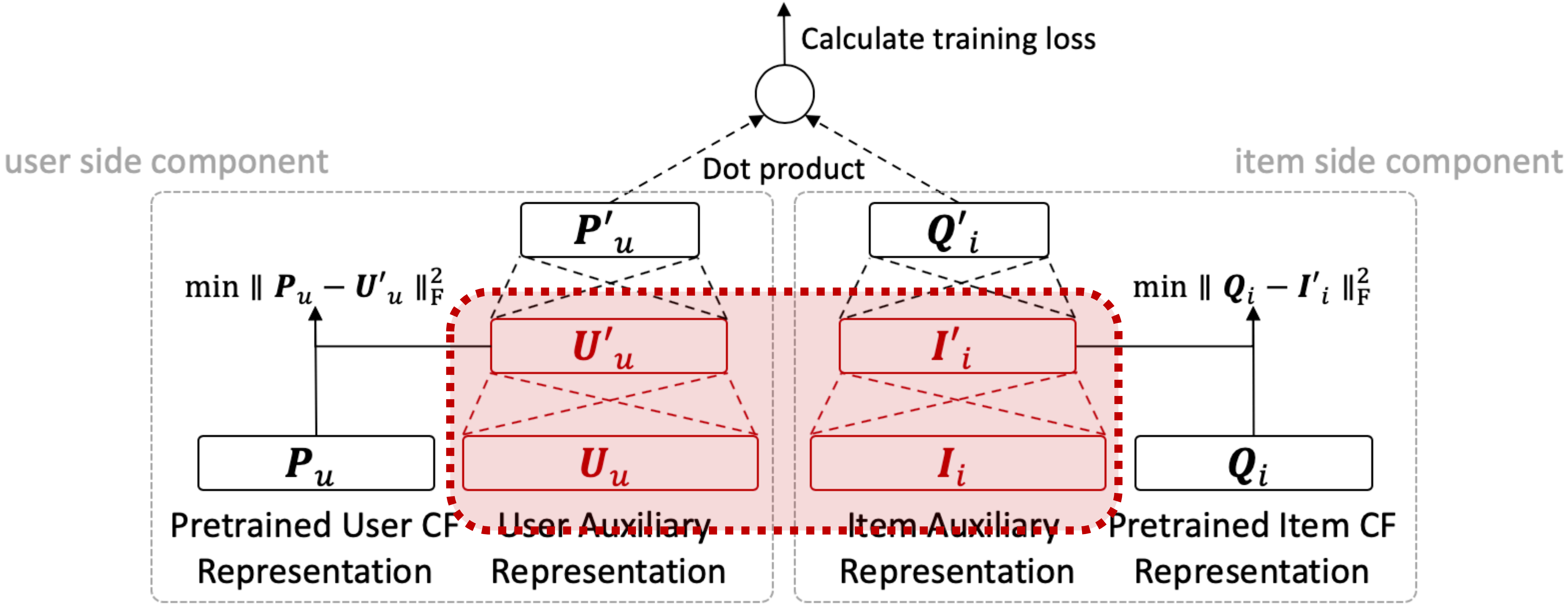
Randomly input high-quality pretrained CF representation or transformed auxiliary representation to following layers to further **improve the effectiveness of training**.

Heater – randomized training



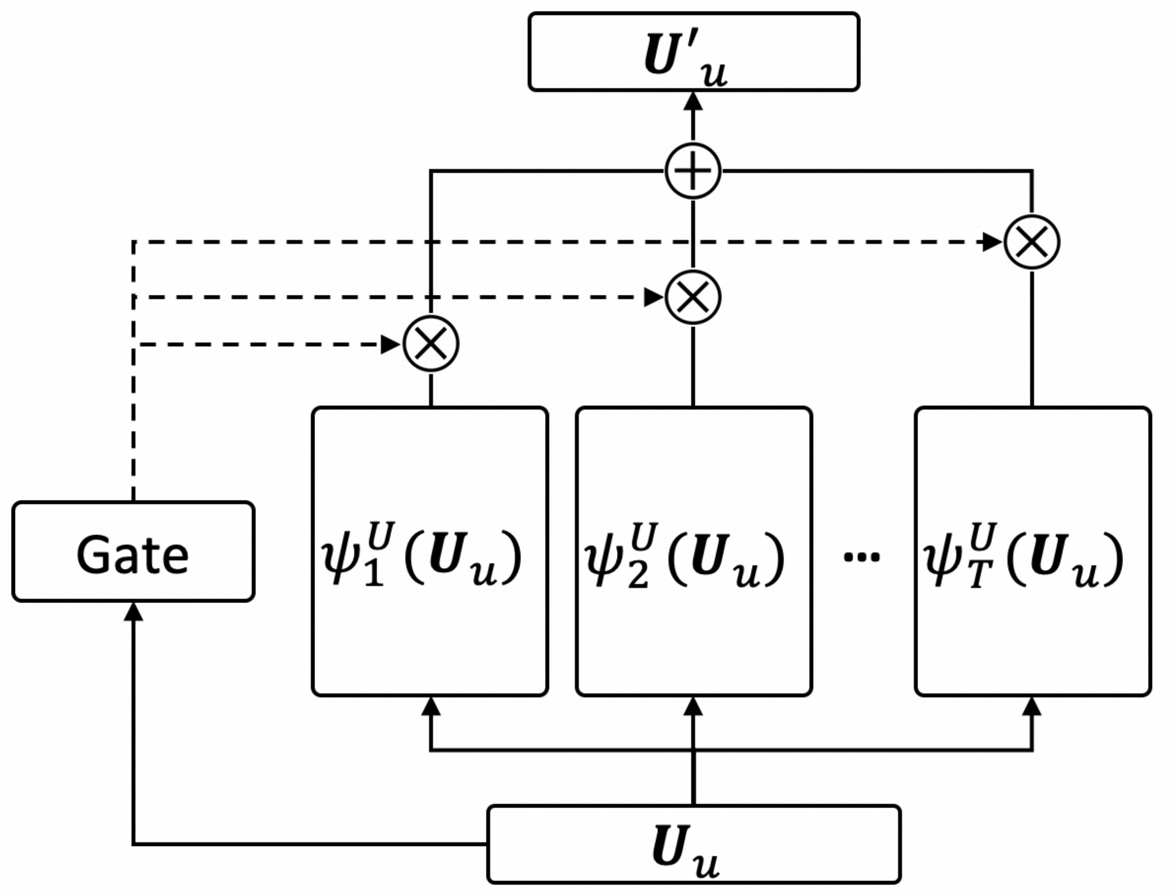
Randomly input high-quality pretrained CF representation or transformed auxiliary representation to following layers to further **improve the effectiveness of training**.

Heater – mixture-of-expert transformation



To address the **unified transformation problem**, replace the MLP with a mixture-of-expert layer.

Heater – mixture-of-expert transformation



To address the **unified transformation problem**, replace the MLP with a mixture-of-expert layer.

Experiments – research questions

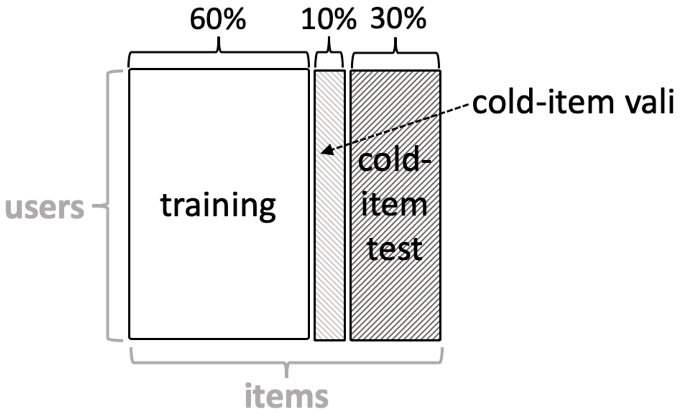
- RQ1: How does Heater perform compared with SOTA baselines?
- RQ2: How effective are the proposed similarity constraint, Randomized Training, and Mixture-of-Experts Transformation mechanisms?
- RQ3: What are the impact of three key hyper-parameters: similarity constraint weight α, Randomized Training probability p, and number of experts T in Mixture-of-Experts Transformation?
- RQ4: What is the impact of the quality of pretrained CF representations on Heater compared with other models that also take pretrained representations as input?

Experiments – three cold start recommendation tasks

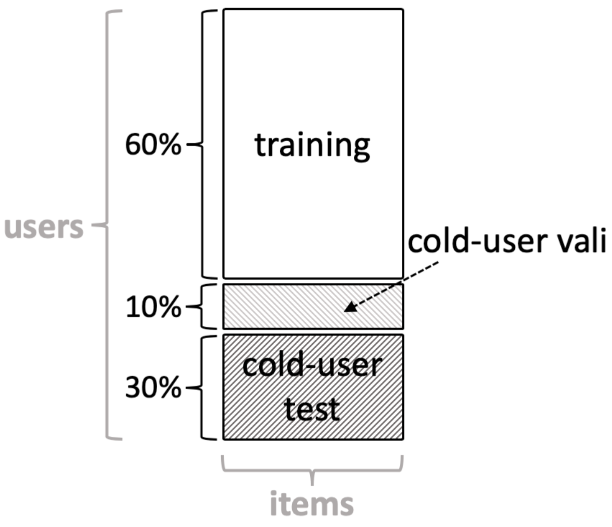
- Task 1: recommend warm items to cold users;
- Task 2: recommend cold items to warm users;
- Task 3: recommend cold items to cold users.

Experiments – dataset

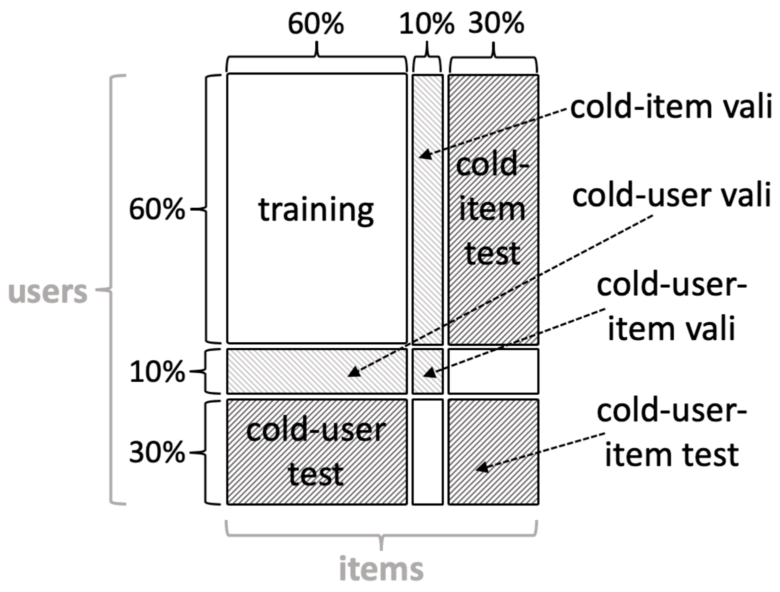
	Training				Validation			Test		
	#user	#item	#record	density	#user	#item	#record	#user	#item	#record
LastFM (Task 1)	1,136	12,850	55,810	0.38%	189	12,850	9,209	567	12,850	27,815
CiteULike (Task 2)	5,551	13,584	164,210	0.22%	5,551	1,018	13,037	5,551	2,378	27,739
XING-U (Task 1)	64,129	12,312	1,549,242	0.20%	10,688	12,312	258,497	32,064	12,312	775,837
XING-I (Task 2)	64,129	12,312	1,549,242	0.20%	64,129	2,051	275,782	64,129	6,156	756,638
XING-UI (Task 3)	64,129	12,312	1,549,242	0.20%	10,688	2,051	45,807	32,064	6,156	379,730



LastFM



CiteULike



XING-UI

With corresponding auxiliary representations of users and/or items.

Experiments – baselines

- **KNN**: from Sedhain et al. 2014, can work for Task 1 and Task 2;
- **CMF**: from Singh et al. 2008, can work for Task 1 and Task 2;
- **LinMap**: from Gantner et al. 2010, **can work for all three tasks**;
- **NLinMap**: from Ooed et al. 2013, **can work for all three tasks**;
- **LoCo**: from Sedhain et al. 2017, can work for Task 1 and Task 2;
- **LWA**: from Vartak et al. 2017, can only work for Task 2;
- **DropoutNet**: from Volkovs et al. 2017, **can work for all three tasks**;
- **LLAE**: from Li et al. 2019, can work for Task 1 and Task 2.

Experiments – RQ1 compare with baselines

NDCG@20

	LastFM (Task 1)	CiteULike (Task 2)	XING-U (Task 1)	XING-I (Task 2)	XING-UI (Task 3)
KNN	0.3537	0.1500	0.1722	0.0740	-
LinMap	0.2880	0.2150	0.3933	0.1605	0.1095
CMF	0.3332	0.2289	0.3488	0.0628	-
LoCo	0.3586	0.2503	0.3538	0.2230	-
NLinMap	0.3535	0.2641	<u>0.4001</u>	0.2118	0.1418
LWA	-	0.2960	-	0.2008	-
DropoutNet	0.3439	0.3089	0.2761	<u>0.2236</u>	<u>0.1454</u>
LLAE	<u>0.3658</u>	<u>0.3249</u>	-	-	-
Heater	**0.3705**	**0.3731**	**0.4150**	**0.2372**	**0.1566**
Δ	1.3%*	14.8%**	3.7%**	6.1%**	7.7%**

'-' represents unavailable result: KNN, CMF, LoCo, LWA and LLAE cannot work for Task 3; LWA cannot work for Task 1; LLAE run into out-of-memory error on XING dataset.

Experiments – Heater vs. baselines

NDCG@20

	LastFM (Task 1)	CiteULike (Task 2)	XING-U (Task 1)	XING-I (Task 2)	XING-UI (Task 3)
KNN	0.3537	0.1500	0.1722	0.0740	-
LinMap	0.2880	0.2150	0.3933	0.1605	0.1095
CMF	0.3332	0.2289	0.3488	0.0628	-
LoCo	0.3586	0.2503	0.3538	0.2230	-
NLinMap	0.3535	0.2641	0.4001	0.2118	0.1418
LWA	-	0.2960	-	0.2008	-
DropoutNet	0.3439	0.3089	0.2761	0.2236	0.1454
LLAE	0.3658	0.3249	-	-	-
Heater	**0.3705**	**0.3731**	**0.4150**	**0.2372**	**0.1566**
Δ	1.3%*	14.8%**	3.7%**	6.1%**	7.7%**

Significant improvement over the best baseline models.

Experiments – ablation study

NDCG@20

	LastFM (Task 1)	CiteULike (Task 2)	XING-U (Task 1)	XING-I (Task 2)	XING-UI (Task 3)
Heater	**0.3705**	**0.3731**	**0.4150**	**0.2372**	**0.1566**
w/o SC	0.2387	0.3437	0.3595	0.2053	0.1263
w/o RT	0.3532	0.3672	0.3145	0.1833	0.1511
w/o MoET	0.3689	0.3382	0.3753	0.2132	0.1434

SC: similarity constraint

RT: randomized training

MoET: mixture-of-expert transformation

Experiments – ablation study

NDCG@20

	LastFM (Task 1)	CiteULike (Task 2)	XING-U (Task 1)	XING-I (Task 2)	XING-UI (Task 3)
Heater	**0.3705**	**0.3731**	**0.4150**	**0.2372**	**0.1566**
w/o SC	0.2387	0.3437	0.3595	0.2053	0.1263
w/o RT	0.3532	0.3672	0.3145	0.1833	0.1511
w/o MoET	0.3689	0.3382	0.3753	0.2132	0.1434

Without any one of the three components, the performance decreased.

Experiments – more

Welcome to read our paper to find more experimental results.

Conclusions

- Propose a novel **cold start recommendation algorithm** that can provide recommendation for both new users and new items;
- Propose the **similarity constraint**, **randomized training**, and **mixture-of-expert transformation** to address three remaining challenges of existing cold start recommendation algorithms;
- Extensive experiments on three public datasets show the **effectiveness** of the proposed model and three components.

Thank you!

Ziwei Zhu, Shahin Sefati*, Parsa Saadatpanah*, and James Caverlee

Texas A&M University

*Comcast Applied AI Lab

TEXAS A&M UNIVERSITY

COMCAST

SIGIR 2020
XI'AN·CHINA