Federated Recommendation Systems

Ben Tan,

AI Group, WeBank, China

Recommender Systems Have Been Widely Used

E-commerce

Online Video

Social Network

News Feeds

Online Advertising

Recommender Systems Improve User Engagement

personized services

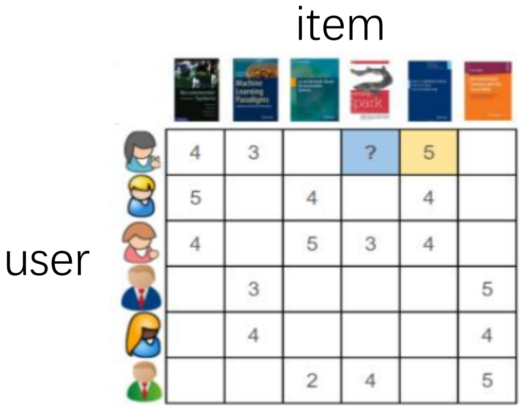
precision marketing

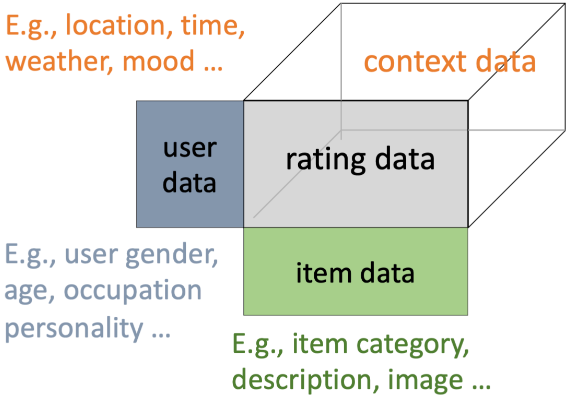
YouTube Homepage: 60%+ more clicks [Davidson et al. 2010]

Netflix: 80%+ more movie watches [Gomze-Uribe et al 2016]

Amazon: 30%+ more page views [Smith and Linden, 2017]

Overview of Recommender Systems





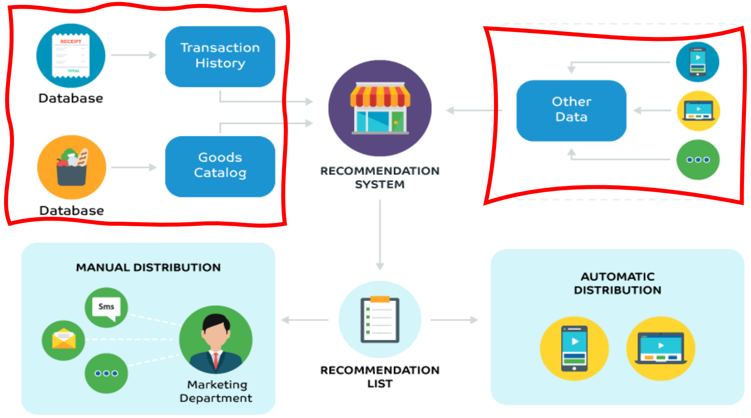
Input: historical user-item interactions, and optionally additional side information (e.g., user demographic, item attributes)

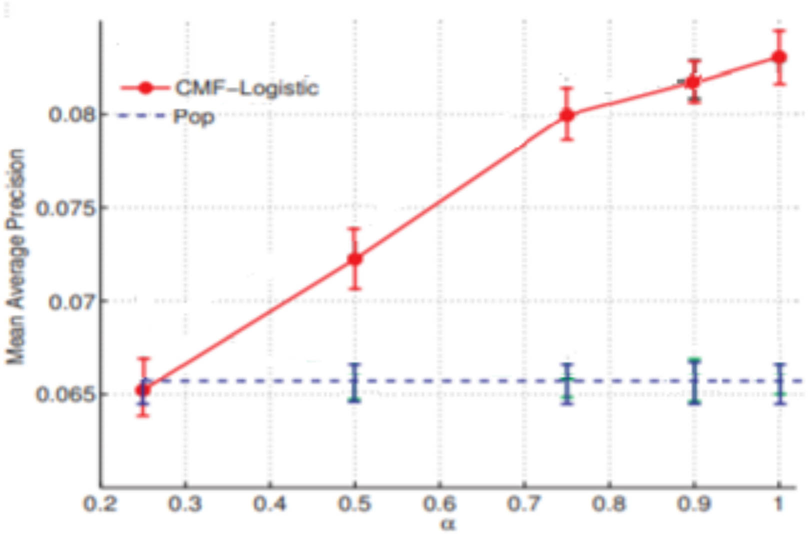
Output: how likely a user would interact with an item (e.g., a movie, a song, a product)

More Data Used in Recommender Systems, Better Performance

Party A

Party B





Auxiliary data used

- Singh and Gordon 2008. Relational learning via collective matrix factorization. ACM KDD 2008.
- Pan 2016. A survey of transfer learning for collaborative recommendation with auxiliary data. Neurocomputing.

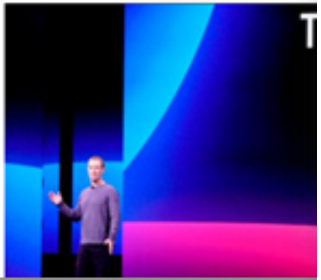
Reality in Recommender Systems: Data Silos

CNN BUSINESS Markets Tech Media Success Perspectives Videos Edition

Facebook finally rolls out privacy tool for your browsing history

By Kaya Yurieff, CNN Business

Updated 1839 GMT (0239 HKT) August 2



Google strengthens Chrome's privacy controls

Frederic Lardinois @fredericl / 7

Google today announced that will, in the long run, intr cookies and enhance its us

With this move, Google is n anti-fingerprinting technolog happening in the Chrome b change and adapt their coo

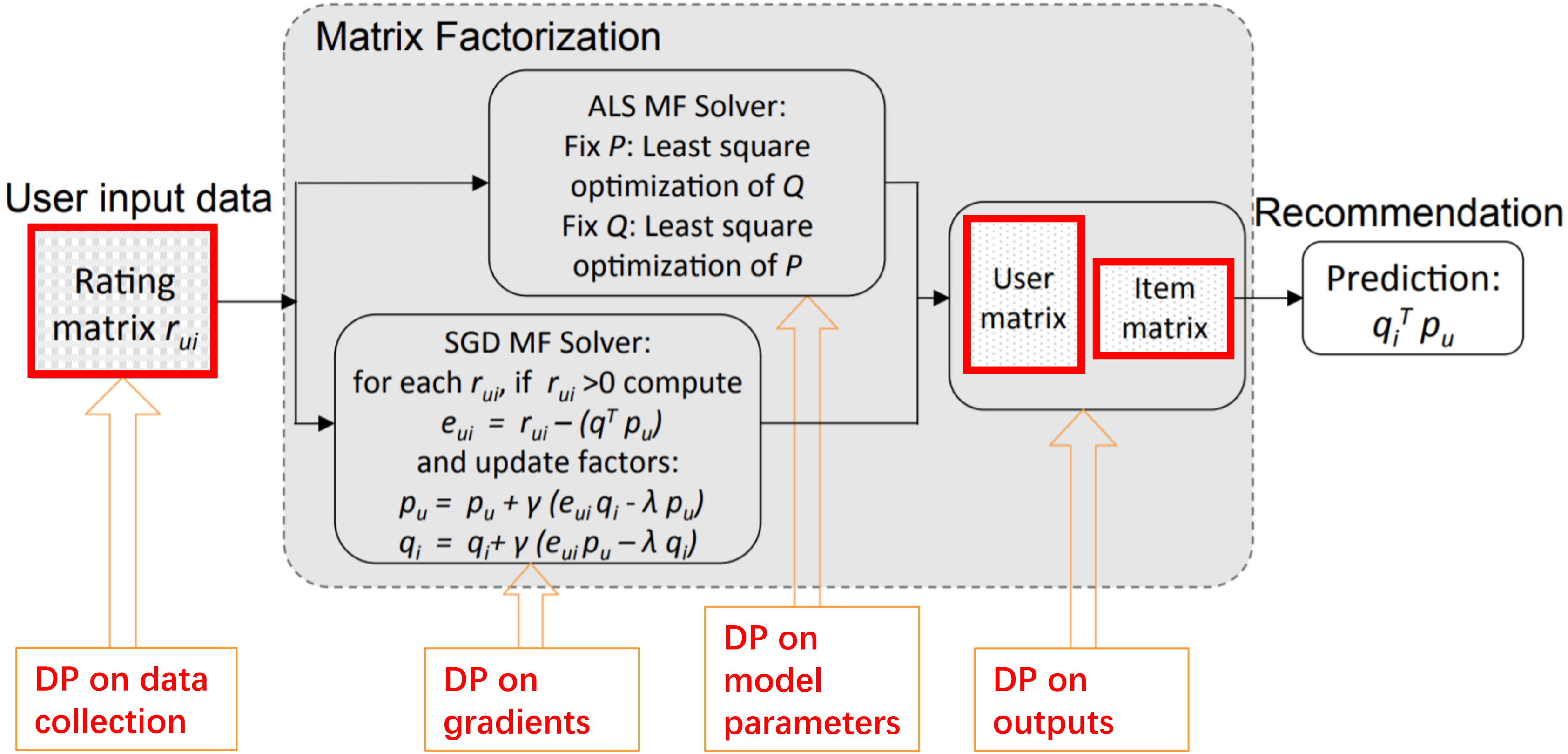
CNN BUSINESS Markets Tech Media Success Perspectives Videos Edition

Top Microsoft exec says online privacy has reached 'a crisis point'

By Clare Duffy, CNN Business

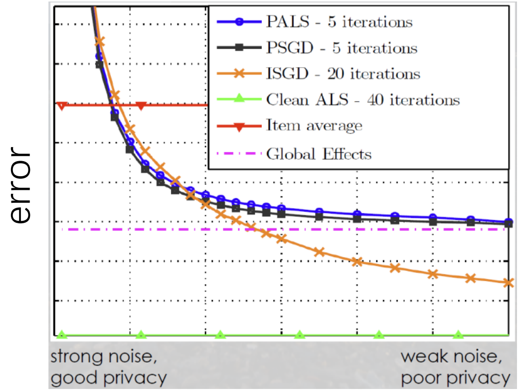
Updated 1749 GMT (0149 HKT) October 14, 2019

Differentially Private Matrix Factorization [Knijnenburg and Berkovsky, 2017]



Bart P. Knijnenburg and Shlomo Berkovsky, 2017. Privacy for Recommender Systems. RecSys 2017 Tutorial.

We Need New Technology for RecSys with Decentralized Data

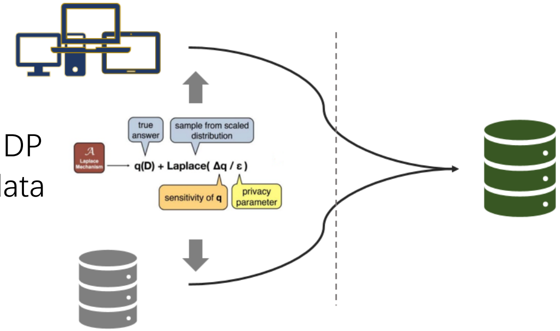


Increasing noise, decreasing performance

Desired properties for new technology:

Lossless performance in decentralized setting, compared with centralized setting.

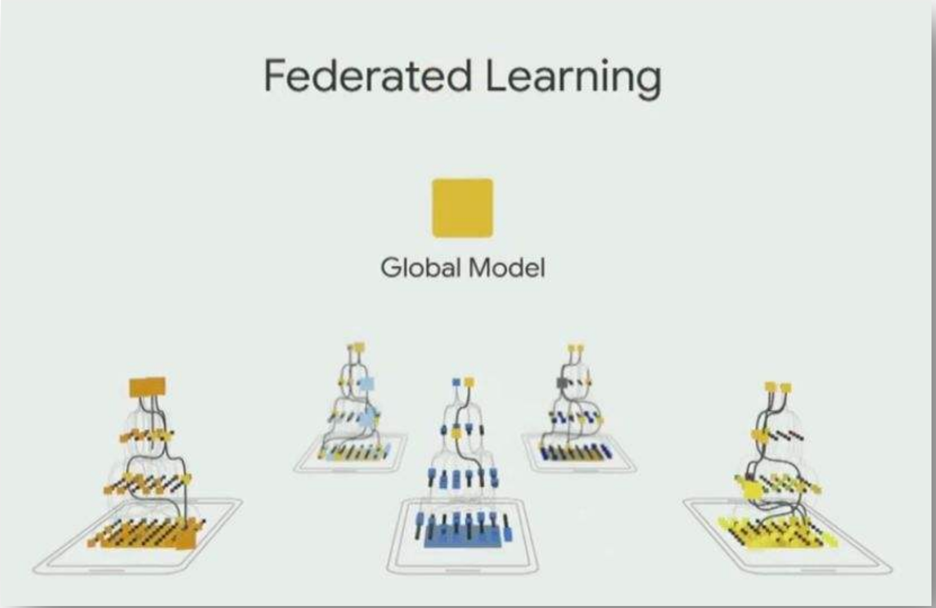
Apply DP over data



Raw data after DP is transmitted between parties.

Data protected in decentralized setting, with raw data staying locally.

Federated Learning to Bridge Decentralized Data



Lossless performance

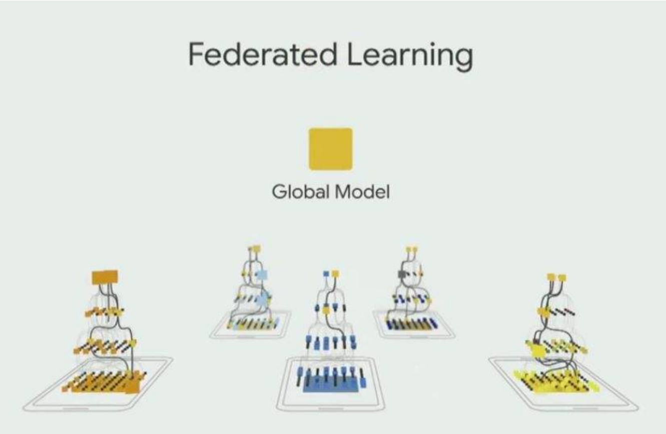
- Performance of 'A fed B' is close to 'A+B'

Data protected

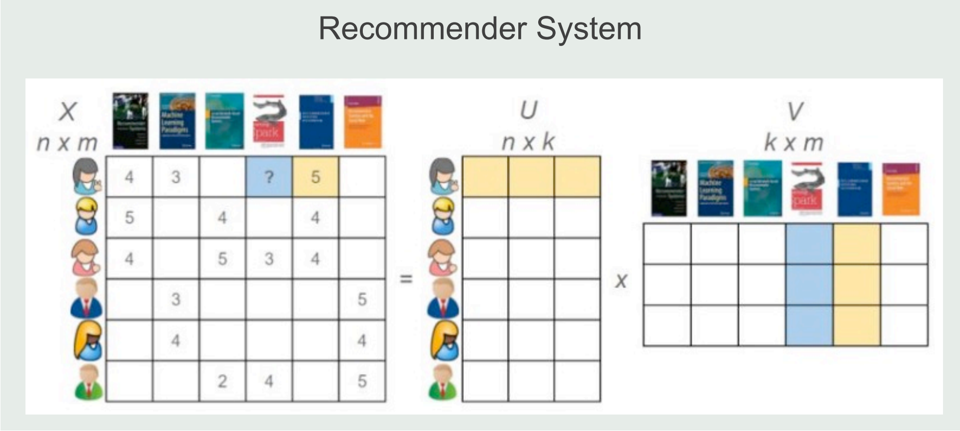
- Raw data stays locally
- Only parameters and gradients are securely transmitted

Yang et al. 2018, Federated machine learning: concepts and applications. ACM TIST.

Federated Recommendation



+

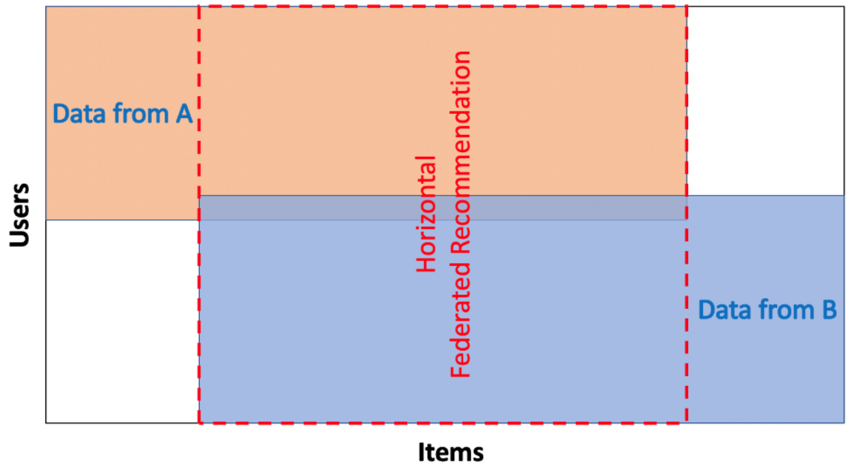


Assumption: for easier understanding and system efficiency, we assume the existence of a trustworthy 3rd-party server in the following federated recommendation solution discussion.

In general, such 3rd-party servers can be removed to strengthen the data security.

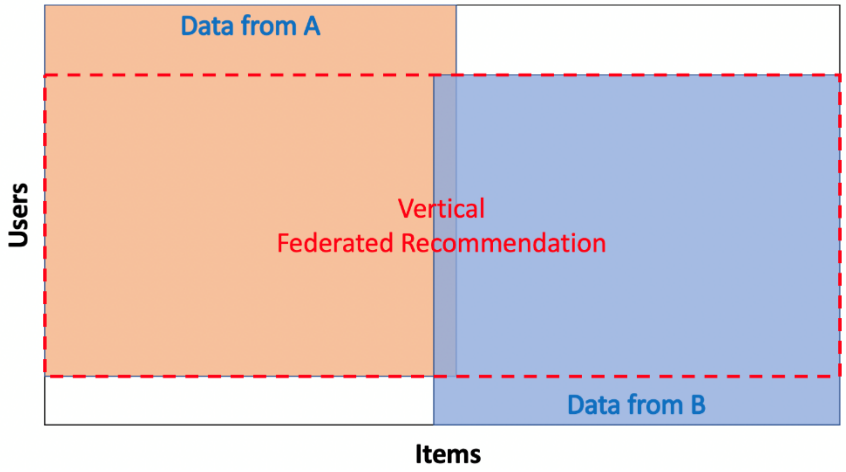
Categorization of Federated Recommendation

Horizontal Federated Recommendation
(a.k.a. Item-based FedRec)



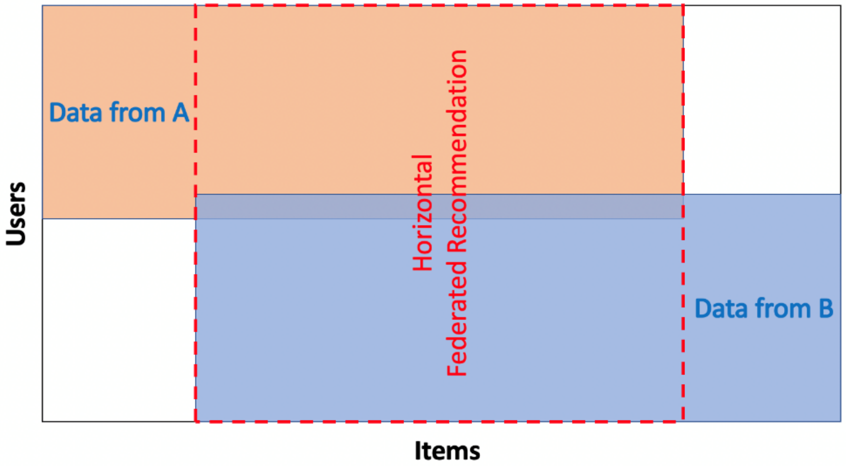
Large overlap of items of the two rating matrices

Vertical Federated Recommendation
(a.k.a. User-based FedRec)



Large overlap of users of the two rating matrices

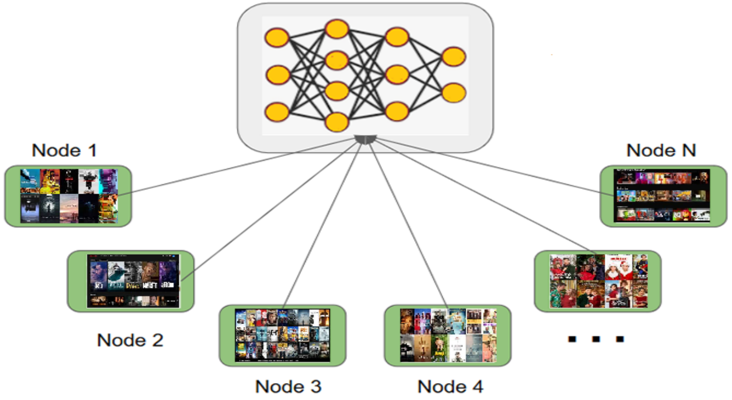
Category 1: Horizontal Federated Recommendation



Large overlap of items of the two rating matrices

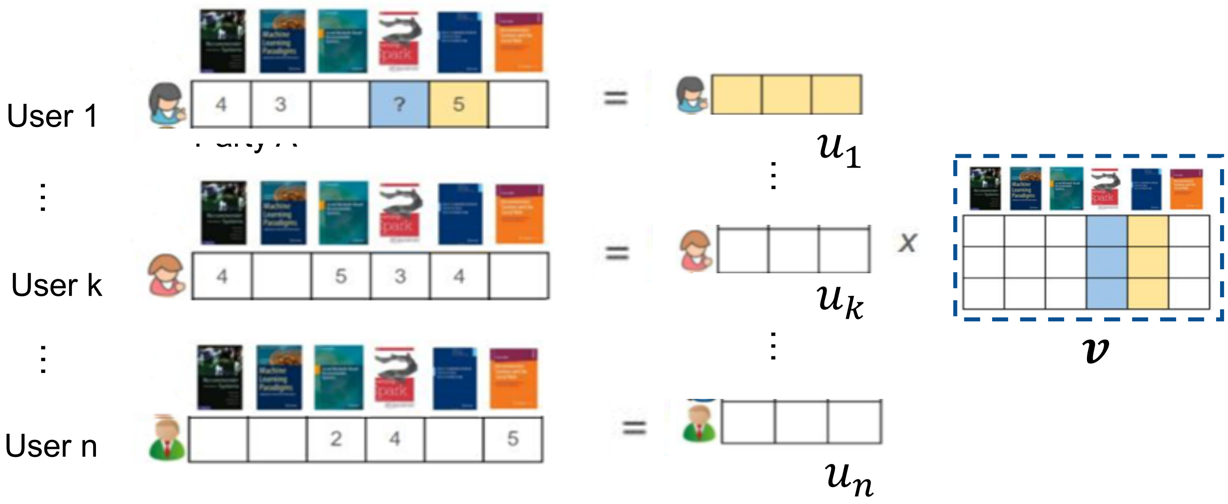
Horizontal Federated Recommendation: Case 1

Example: movie recommendation with data from individual users



Federated Collaborative Filtering [Ammad et al. 2019]

Intuition: decentralized matrix factorization, each user profile is updated locally, item profiles are aggregated and updated by server.



Loss function $$\min_{U,V} \frac{1}{M}(r_{i,j} - \langle u_i, v_j \rangle)^2 + \lambda||U||_2^2 + \mu||V||_2^2$$

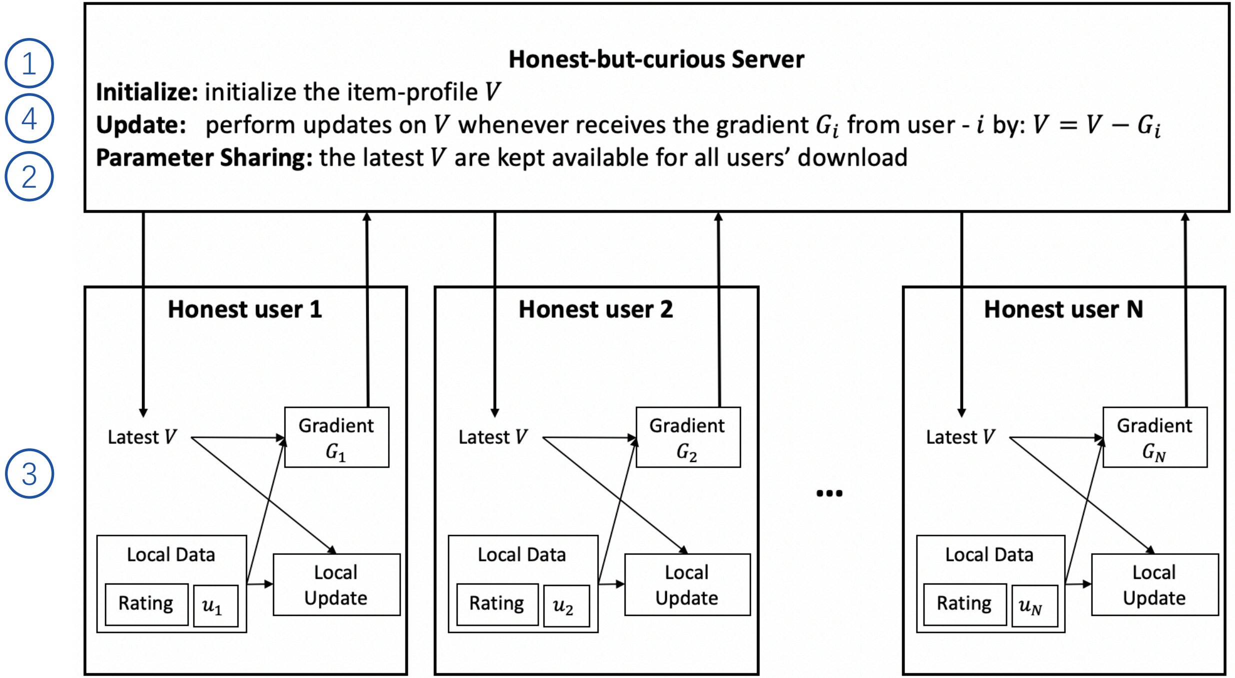
Update function $$\boxed{u_i^t = u_i^{t-1} - \gamma \nabla_{u_i} F(U^{t-1}, V^{t-1})}$$ User local updates

Server updates $$v_i^t = v_i^{t-1} - \boxed{\gamma \nabla_{v_i} F(U^{t-1}, V^{t-1})}$$ Gradients from users

Ammad-ud-din, et al. 2019. Federated Collaborative Filtering for Privacy-Preserving Personalized Recommendation System. arXiv:1901.09888.

Federated Collaborative Filtering [Ammad et al. 2019]

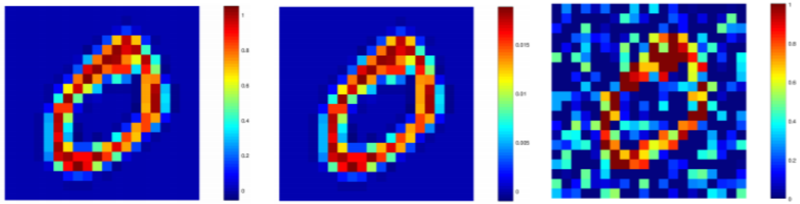
Pros: user data is decentralized.
Cons: no MPC (plaintext gradients).



Training Process:

1. Server initializes item profiles, parties initializes user profiles;
2. Sever distributes item profiles to parties;
3. Parties locally update user profiles with item profiles; Parties send item profile gradient updates to server;
4. Server updates item profile.

Gradient leaks information

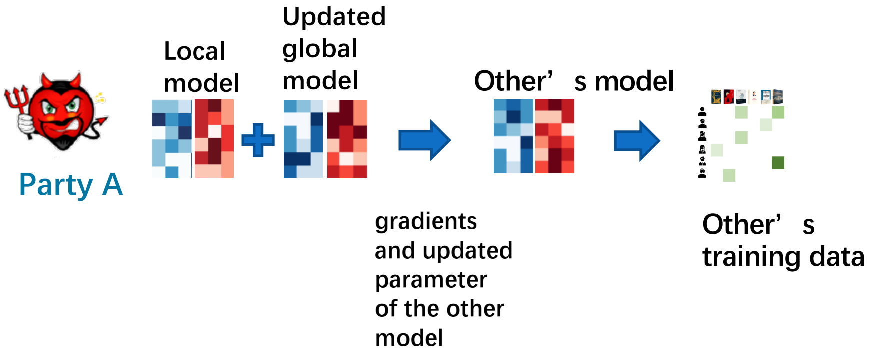


(a) Original 20x20 image of hand-written number 0, seen as a vector over $\mathbb{R}^{400}$ fed to a neural network.

(b) Recovered image using 400/10285 (3.89%) gradients (see Sect.3, Example 2). The difference with the original (a) is only at the value bar.

(c) Recovered image using 400/10285 (3.89%) gradients (see Sect.3, Example 3). There are noises but the truth label 0 can still be seen.

Fig. 3. Original data (**a**) vs. leakage information (**b**), (**c**) from a small part of gradients in a neural network.

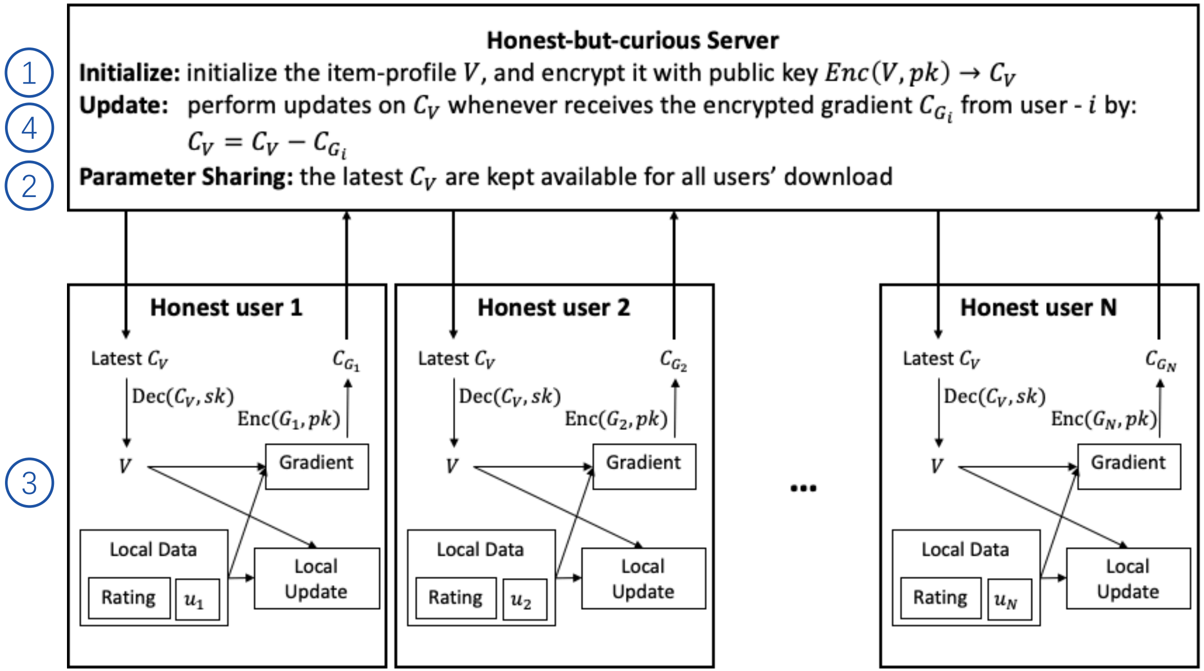


- Phong, et al. 2018. Privacy-Preserving Deep Learning via Additively Homomorphic Encryption. IEEE Trans. Information Forensics and Security , 13, 5 (2018),1333–1345

- Gao, et al. 2020. Privacy Threats against Federated Matrix Factorization, International Workshop on Federated Learning for User Privacy and Data Confidentiality in Conjunction with IJCAI 2020, (FL-IJCAI'20), Kyoto, Japan

Horizontal Federated Matrix Factorization [Chai et al. 2019]

Intuition: Item profile gradients are encrypted by HE. Semi-honest server securely aggregates encrypted item profiles gradients, and knows nothing about the profile content.



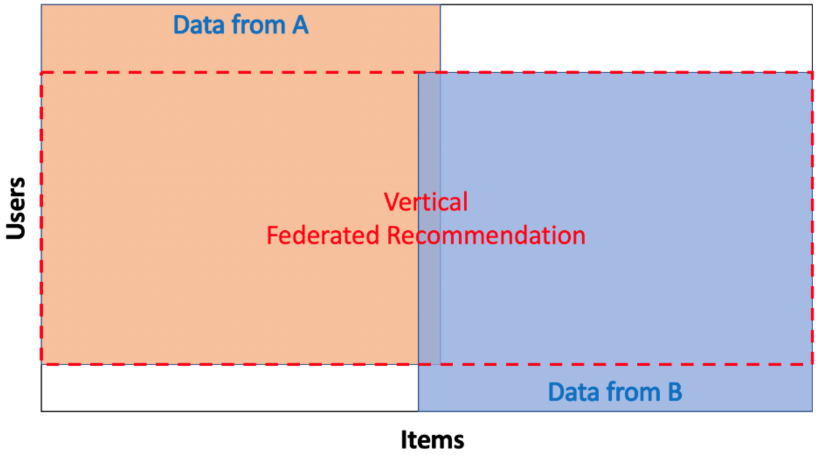
Training Process:

1. Server initializes and encrypts item profiles;
2. Sever distributes encrypted item profiles to parties;
3. Parties locally update user profiles with encrypted item profiles; Parties send encrypted item profile gradient updates to server;
4. Server securely aggregates item profile gradients and updates item profiles.

Security of secure aggregation protocol is guaranteed [Bonawitz et al. 2017].

- Chai, et al. 2019. Secure Federated Matrix Factorization. arXiv:1906.05108.
- Bonawitz et al. 2017. Practical Secure Aggregation for Privacy-Preserving Machine Learning. CCS, pages 1175–1191.

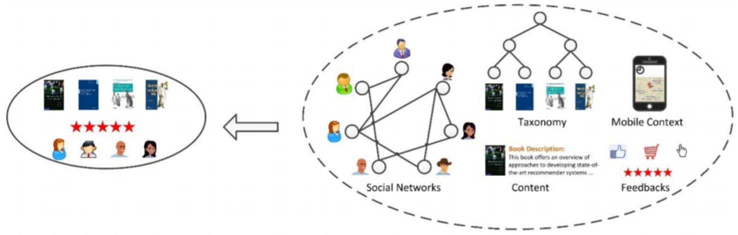
Category 2: Vertical Federated Recommendation



Large overlap of users of the two rating matrices

Vertical Federated Recommendation: Case

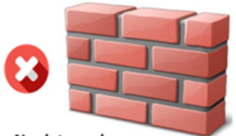
Example:
Shared users
different features



book recommendation auxiliary data from third-parties

						Location	Time
4	3		?	5		Georgia	2018.5
5		4		4		Florida	2019.1
4		5	3	4		Hawaii	2017.3
	3				5	Kansas	2018.5
	4				4	Georgia	2018.10
		2	4		5	Florida	2019.9

Party A



Sports	Photography	Movie	Food
Y	N	N	N
N	N	N	N
Y	N	N	Y
Y	Y	N	N
N	Y	Y	N
N	Y	Y	Y

Party B

Federated Factorization Machine [Zheng et al. 2019]

Intuition: cross-features between A and B are useful, but features are sensitive. Federated factorization machine computes these cross-party cross-features and their gradients under encryption.

Party A

4	3		?	5	
5		4		4	
4		5	3	4	
	3				5
	4				4
		2	4		5

Location	Time
Georgia	2018.5
Florida	2019.1
Hawaii	2017.3
Kansas	2018.5
Georgia	2018.10
Florida	2019.9

Party B

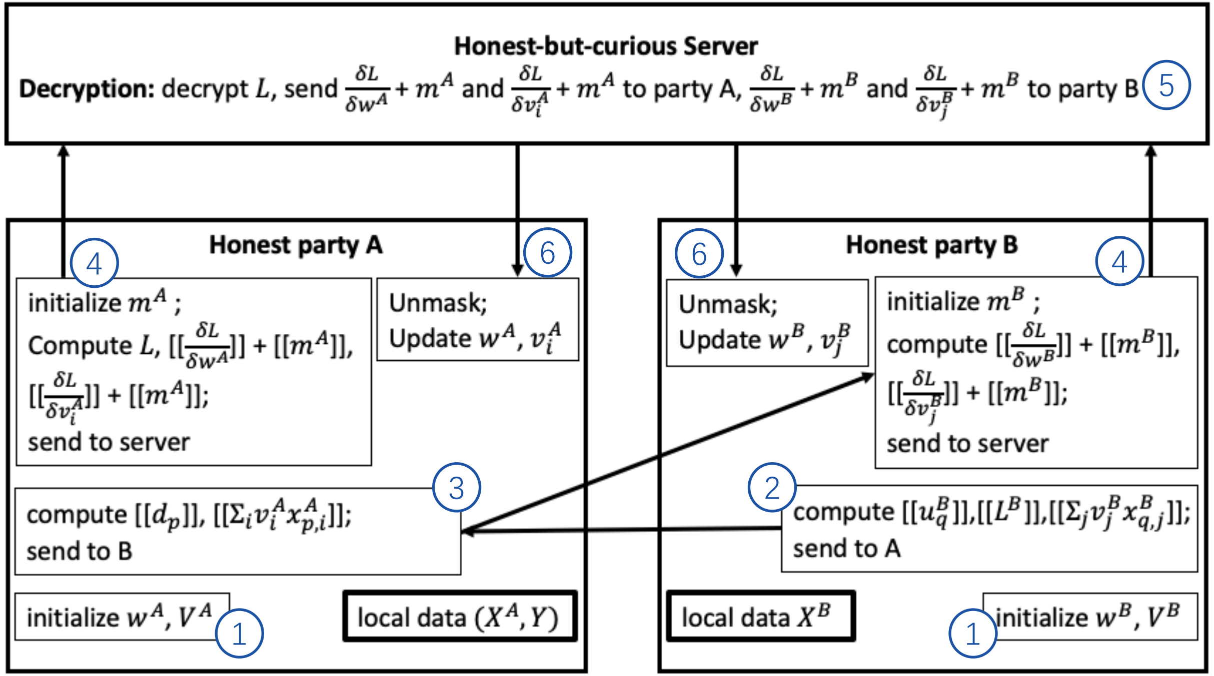
Sports	Photography	Movie	Food
Y	N	N	N
N	N	N	N
Y	N	N	Y
Y	Y	N	N
N	Y	Y	N
N	Y	Y	Y

Cross features between A and B are useful;
e.g., "location x sports" can be a strong indicator for predicting Georgia user's preference to sports movies.

Prediction function $$f([\mathbf{x}_p^{(A)}; \mathbf{x}_q^{(B)}]) = \underbrace{f(\mathbf{x}_p^{(A)})}_{\text{Cross features in A}} + \underbrace{f(\mathbf{x}_q^{(B)})}_{\text{Cross features in B}} + \underbrace{\sum \langle \mathbf{v}_i^{(A)}, \mathbf{v}_j^{(B)} \rangle x_{p,i}^{(A)} x_{q,j}^{(B)}}_{\text{Cross features between A and B}}$$

- Rendle 2012: Factorization Machines with libFM, in ACM Trans. Intell. Syst. Technol., 3(3), May.
- Zheng. 2019. Federated factorization machine. Tech Report WeBank.

Federated Factorization Machine [Zheng et al. 2019]



Training Process

1 Parties initialize models

2 Party B sends encrypted partial prediction, partial loss and partial feature gradients to party A

3 Party A sends encrypted error and partial feature gradients to party B

4 Parties send encrypted and masked gradients to server

5 Server decrypts and sends back

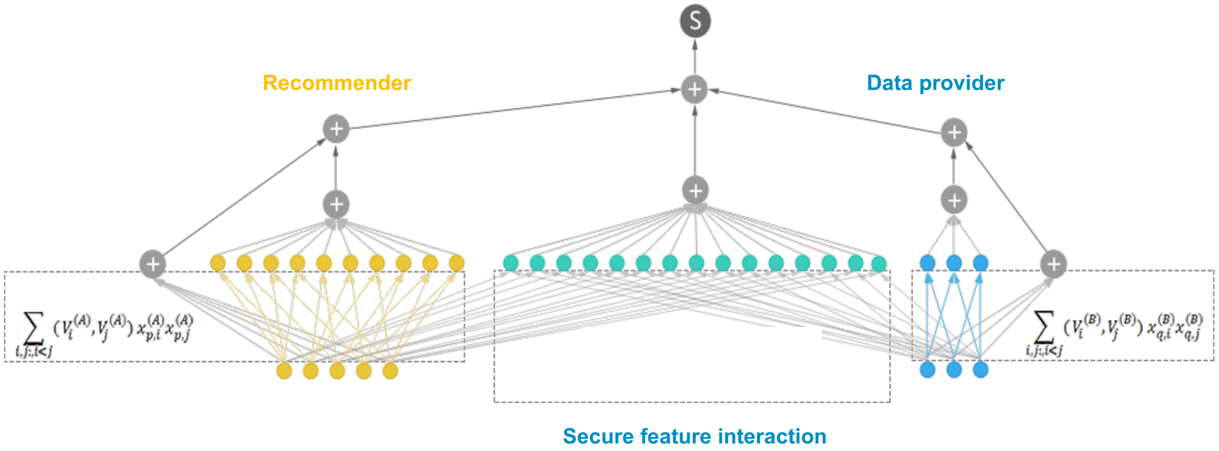
6 Parties unmask and update models

Security of semi-honest MPC protocol is guaranteed [Goldreich et al. 1987].

Goldreich et al. 1987. How to Play ANY Mental Game. STOC, pages 218–229.

Federated Factorization Machine [Zheng et al. 2019]

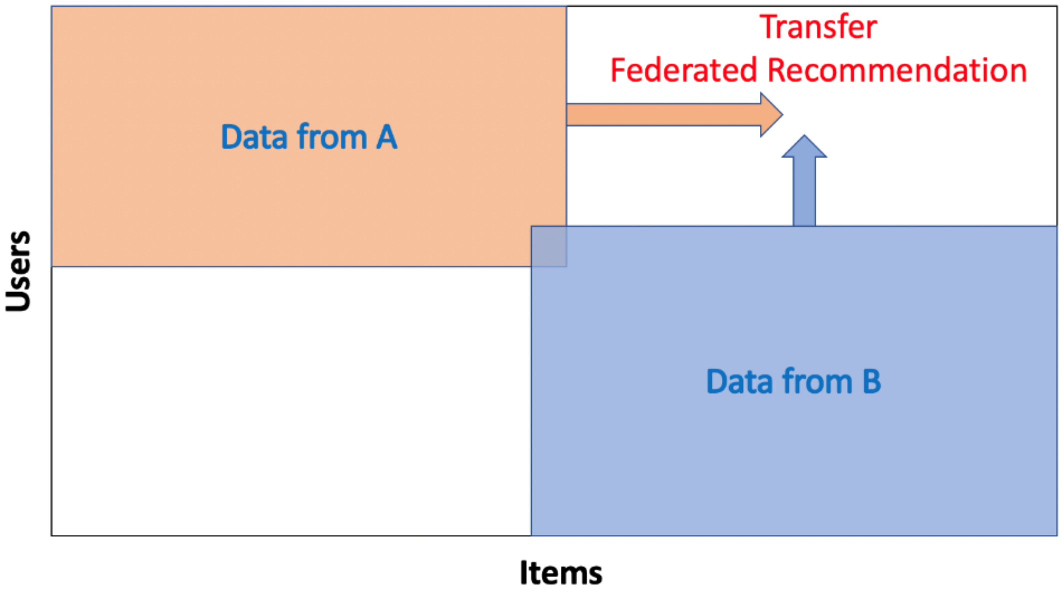
Inference Process: encrypted prediction on party A’ s features + encrypted prediction on A&B features + encrypted prediction on party B’ s features.



1. Party A and B compute encrypted intermediate results
2. Server aggregates the encrypted intermediate results and decrypts
3. Sever sends plain-text prediction to party A

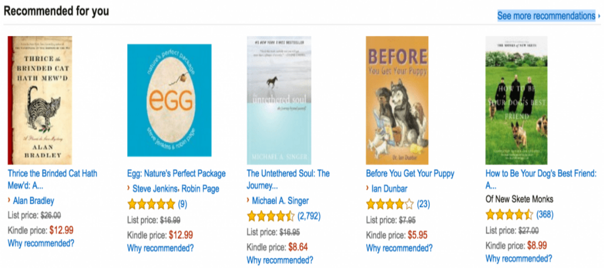
What If Different Users and Items at the Same Time?

Transfer Federated Recommendation



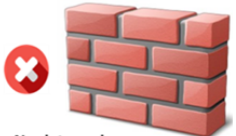
Category 3: Transfer Federated Recommendation

Example: movie and book recommenders with different groups of users



⟺

4	3		?	5	
5		4		4	
4		5	3	4	
	3				5
	4				4
		2	4		5

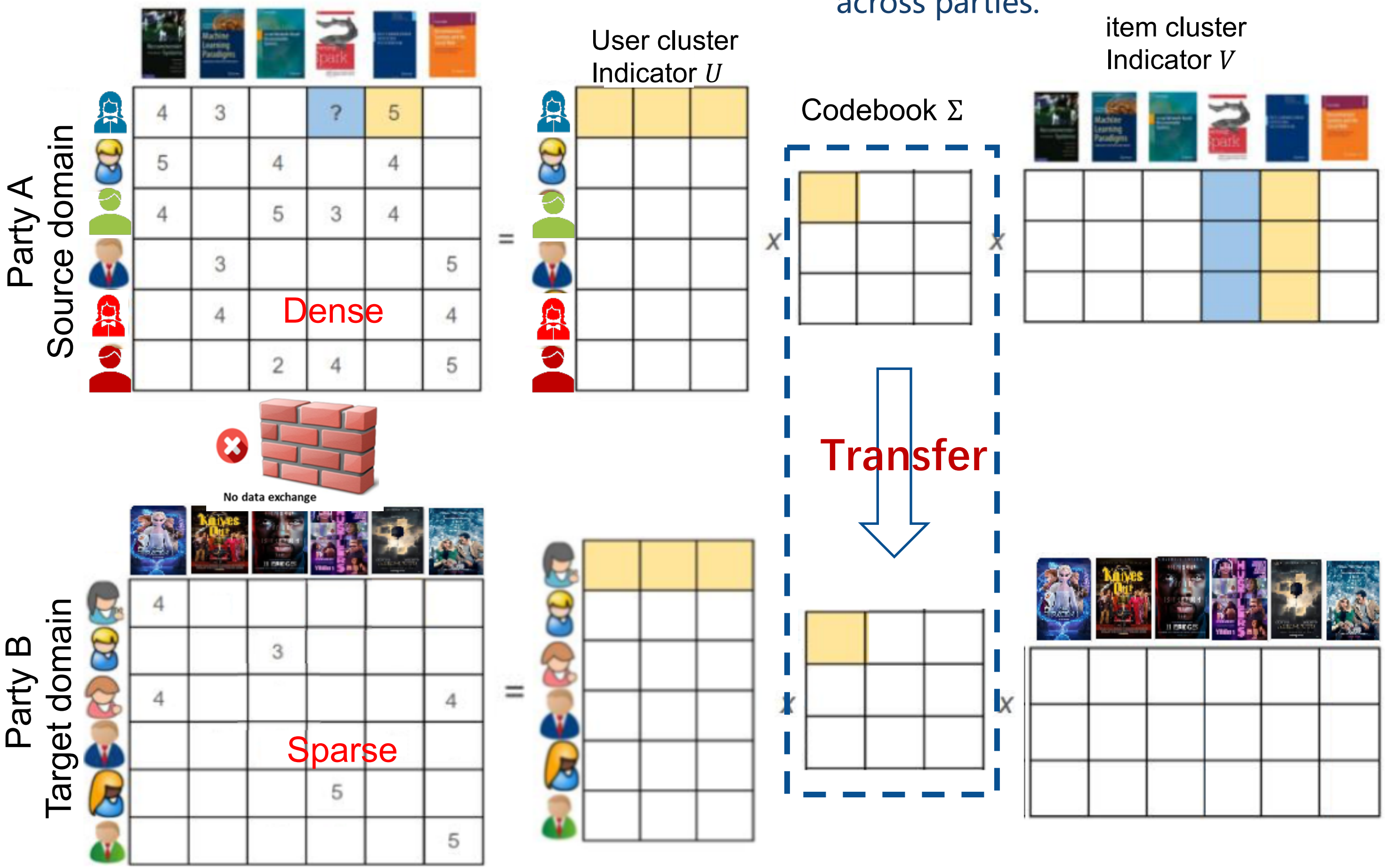


No data exchange

4			4	3	
5		3		4	
4		5	3	4	4
	3	4			5
	4		5		4
3		2	4		5

Matrix Tri-factorization [Li et al. 2009]

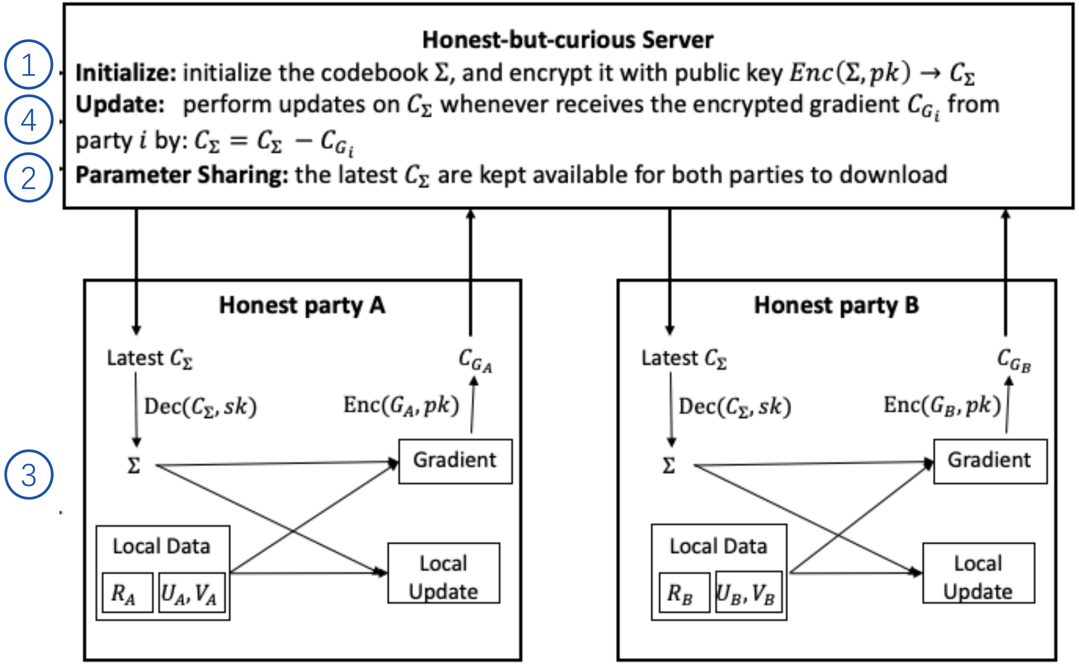
Intuition: similar users/items can be clustered into groups, and there exist group correspondences across parties.



Li et al. Transfer Learning for Collaborative Filtering via a Rating-Matrix Generative Model, ICML, pp.617-624.

Federated Matrix Tri-factorization [Tan et al. 2019]

Intuition: codebooks as group correspondences are used for transfer, they are encrypted and securely aggregated by semi-honest server, and user/item profiles are updated by parties.



Training Process

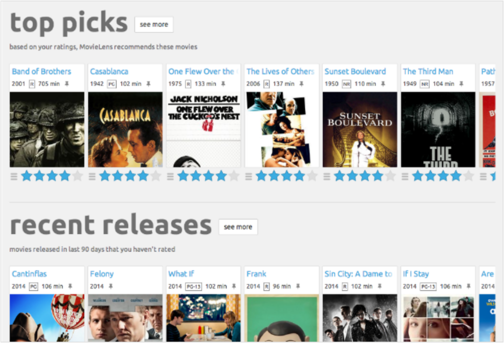
1. Server initializes and encrypts codebook; Parties initializes user and item profiles;
2. Server distributes encrypted codebook to parties;
3. Parties update user and item factors by decrypted codebook; Parties compute codebook gradients and send encrypted gradients to server;
4. Server securely aggregates encrypted codebook gradients and updates codebook.

Security of secure aggregation protocol is guaranteed [Bonawitz et al. 2017].

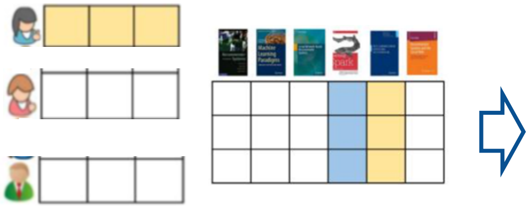
- Tan et al. 2019. Federated matrix tri-factorization. Tech Report, WeBank.
- Bonawitz et al. 2017. Practical Secure Aggregation for Privacy-Preserving Machine Learning. CCS, pages 1175–1191.

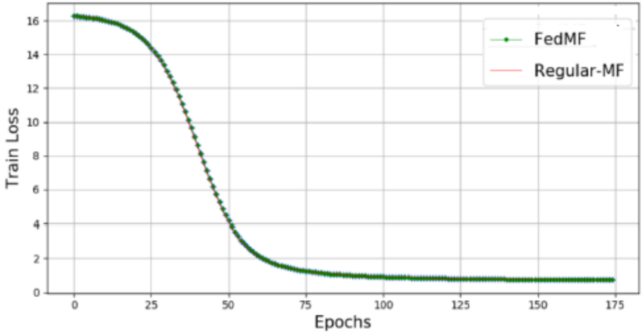
Application 1: Horizontal Federated Movie Recommendation

Recommender keeps user data on local devices, protects privacy while achieving lossless performance.



MovieLens





	RegularMF	FedMF
RMSE	1.3969165	1.3965372

Chai, et al. 2019. Secure federated matrix factorization. arXiv:1906.05108.

FedRec: Open-sourced Project

https://github.com/FederatedAI/FedRec

3. Algorithms list:

1. Hetero FM(factorization machine)

Build a hetero factorization machine model through multiple parties.

- Corresponding module name: HeteroFM
- Data Input: Input DTable.
- Model Output: Factorization Machine model.

2. Homo-FM

Build a homo factorization machine model through multiple parties.

- Corresponding module name: HomoFM
- Data Input: Input DTable.
- Model Output: Factorization Machine model.

3. Hetero MF(matrix factorization)

Build a hetero matrix factorization model through multiple parties.

- Corresponding module name: HeteroMF
- Data Input: Input DTable of user-item rating matrix data.
- Model Output: Matrix Factorization model.

4. Hetero SVD

Build a hetero SVD model through multiple parties.

- Corresponding module name: HeteroSVD
- Data Input: Input DTable of user-item rating matrix data.
- Model Output: Hetero SVD model.

5. Hetero SVD++

Build a hetero SVD++ model through multiple parties.

- Corresponding module name: HeteroSVDPP
- Data Input: Input DTable of user-item rating matrix data.
- Model Output: Hetero SVD++ model.

6. Hetero GMF

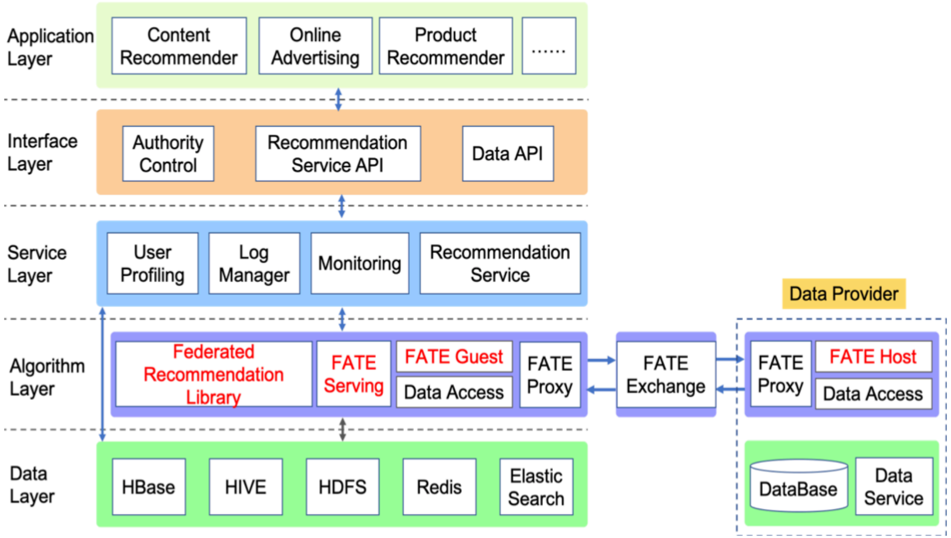
Build a hetero GMF model through multiple parties.

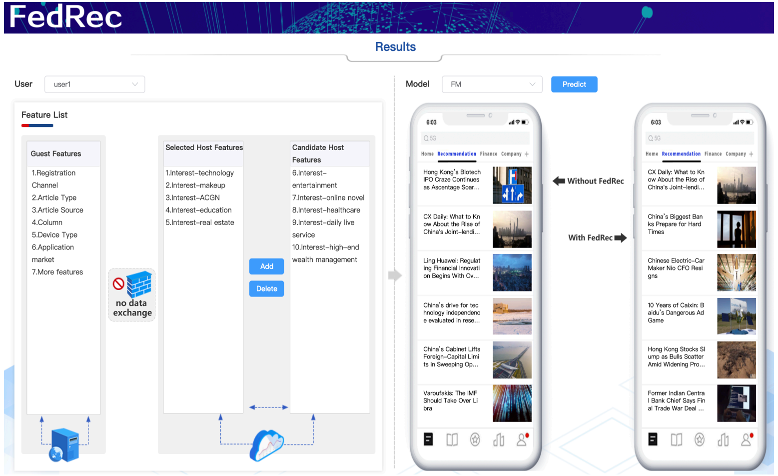
- Corresponding module name: HeteroGMF
- Data Input: Input DTable of user-item rating matrix data(using positive data only).
- Model Output: Hetero GMF model.

More available algorithms are coming soon.

Application 2: Vertical Federated News Feeds Recommendation

https://ad.webank.com/fedrecdemo/index.html?type=en





Tan et al, 2020, A Federated Recommender System for Online Services. RecSys ’20, Virtual Event, Brazil, September 21–26, 2020

Application 2: Vertical Federated News Feeds Recommendation

Recommender leverages auxiliary user data to address cold start and improve performance.

User's Internet browsing behaviors from 3rd-party

Finance News Feeds Recommendation

PV	21%
UV	22%
CTR	11%

Summary

- Recommender systems can be improved with more data
- Yet privacy and security needs to be addressed
- Federated learning to bridge decentralized data in recommendation
  - Vertical Federated Recommendation (a.k.a. user-based FedRec)
  - Horizontal Federated Recommendation (a.k.a. item-based FedRec)
  - Transfer Federated Recommendation
- FedRec is an underexplored area with a lot of opportunities

Contact us