How to Put Users in Control of their Data in Federated Top-N Recommendation with Learning to Rank

Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, Antonio Ferrara, Fedelucio Narducci*
Politecnico di Bari
Bari, Italy
firstname.lastname@poliba.it

ABSTRACT

Recommendation services are extensively adopted in several user-centered applications as a tool to alleviate the information overload problem and help users in orienteering in a vast space of possible choices. In such scenarios, data ownership is a crucial concern since users may not be willing to share their *sensitive* preferences (e.g., visited locations, read books, bought items) with a central server. Unfortunately, data harvesting and collection is at the basis of modern, state-of-the-art approaches to recommendation. To address this issue, we present Federated Pair-wise Learning (FPL), an architecture in which users collaborate in training a central factorization model while controlling the amount of sensitive data leaving their devices. The proposed approach implements pair-wise learning-to-rank optimization by following the *Federated Learning* principles, conceived originally to mitigate the privacy risks of traditional machine learning.

CCS CONCEPTS

• **Information systems** → **Recommender systems**; **Personalization**;

KEYWORDS

federated learning, recommender systems, BPR, privacy control

ACM Reference Format:
Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, Antonio Ferrara, Fedelucio Narducci. 2021. How to Put Users in Control of their Data in Federated Top-N Recommendation with Learning to Rank. In *Proceedings of ACM SAC Conference (SAC'21).* ACM, New York, NY, USA, Article 4, 5 pages. https://doi.org/xx.xxx/xxx_x

1 INTRODUCTION

Collaborative filtering (CF) models have been mainstream research in the recommender system (RS) community over the last two decades thanks to their performance accuracy [11]. Among them, a prominent class uses the matrix factorization (MF) approach as the inference model. The MF model's main aim is to uncover user and item latent representations whose linear interaction explains observed feedback. To date, the majority of existing MF models are trained in a *centralized* fashion causing several concerns about the privacy of user data. The consequent data scarcity dilemma can thereby jeopardize the training of MF models. Training high-quality MF models strongly relies on sufficient in-domain[1] interaction data to ensure that enough co-occurrence information exists to shape similar behavioral/preference patterns in a user community. In recent years, federated learning (FL) was proposed by Google as a mean to offer a *privacy-by-design* solution [3, 13] for machine-learned models. Federated learning aims to meet ML privacy shortcomings by horizontally distributing the model's training over user devices; thus, clients exploit private data without sharing them [13]. Weiss *et al.* [15] state that privacy can be preserved by limiting data collection, which is one of the main privacy concerns [7]. Indeed, the accuracy of RS based on the CF paradigm is strictly dependent on the amount of user preferences available. Our idea is to put users in control of their sensitive data by allowing them to choose the amount of information to share with the server. Hence, if data collection from the server side is reduced, other threats related to retention, sales, and unauthorized data browsing are limited. The proposed system, FPL (short for Federated Pair-wise Learning), is a *federated* factorization model for collaborative recommendation[2]. It extends state-of-the-art factorization approaches to build a RS that puts users in control of their sensitive data. Users participating in the federation process can decide *if* and *to which extent* they are willing to disclose their *sensitive* private data (i.e., what they liked/consumed). FPL mainly leverages not-sensitive information (e.g., places the user has not visited) – which can be large and non-sensitive – to reach a competitive accuracy and, at the same time, respect a satisfactory balance between accuracy and privacy. We have carried out extensive experiments on real-world datasets [16] in the Point of Interest (PoI) domain by considering the accuracy of recommendation and diversity metrics. The experimental evaluation shows that FPL can provide high-quality recommendations, putting the user in control of the amount of sensitive data to share.

2 BACKGROUND

Federated Learning. Federated learning (FL) is a paradigm initially envisioned by Google [9, 13] to train a machine-learning model from data distributed among a loose federation of users'

*The authors are listed in alphabetical order. Corresponding author: Antonio Ferrara (antonio.ferrara@poliba.it).

SAC'21, March 22-March 26, 2021, Gwangju, South Korea

ACM ISBN 978-1-4503-8104-8/21/03…$15.00
https://doi.org/xx.xxx/xxx_x

[1]Although cross-domain recommendation approaches allow combating data scarcity, their applicability depends upon the availability of data providers that can collect cross-domain in their platform. These approaches remain out of focus in this work.
[2]A public implementation of FPL is available at https://anonymous.4open.science/r/10ba8e5b-cc43-4cb9-b6c6-f08876b0449a/.

devices (e.g., personal mobile phones). The rationale is to face the increasing issues of ownership and locality of data to mitigate the privacy risks resulting from centralized machine learning [8]. In particular, given Θ denoting the parameters of a machine learning model, we consider a learning scenario where the objective is to minimize a generic loss function $G(\Theta)$. FL is a learning paradigm in which the users $u \in \mathcal{U}$ of a federation collaborate to solve the learning problem under the coordination of a central server S without sharing or exchanging their raw data with S. From an algorithmic point of view, we start with S sharing Θ with the federation of devices. Then, specific methods solve a local optimization problem on the single device, i.e., using its data, and exploiting Θ. Afterwards, the client shares the parameters of its local model with S. The parameters provided by the clients are then used to update Θ, which is sent back to the devices in a new iteration step.

Factorization Models and Pair-Wise Recommendation. A recommendation problem over a set of users $\mathcal{U}$ and a set of items $\mathcal{I}$ is defined as the activity of finding for each user $u \in \mathcal{U}$ an item $i \in \mathcal{I}$ that maximizes a utility function $g : \mathcal{U} \times \mathcal{I} \rightarrow \mathbb{R}$. In this context, $\mathbf{X} \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{I}|}$ is the user-item matrix containing for each x_{ui} an explicit or implicit feedback (e.g., rating or check-in, respectively) of user $u \in \mathcal{U}$ for item $i \in \mathcal{I}$. In the work at hand, an implicit feedback scenario is considered — i.e., feedback is, e.g., purchases, visits, clicks, views, check-ins —, with $\mathbf{X}$ containing binary values. Therefore, $x_{ui} = 1$ and $x_{ui} = 0$ denote either user u has consumed or not item i, respectively. In FPL, the underlying data model is a Factorization model, inspired by MF [10], a recommendation model that became popular in the last decade thanks to its state-of-the-art recommendation accuracy [11]. This technique aims to build a model Θ in which each user u and each item i is represented by the embedding vectors $\mathbf{p}_u$ and $\mathbf{q}_i$, respectively, in the shared latent space $\mathbb{R}^F$. The algorithm relies on the assumption that $\mathbf{X}$ can be factorized such that the dot product between $\mathbf{p}_u$ and $\mathbf{q}_i$ can explain any observed user-item interaction x_{ui}, and that any non-observed interaction can be estimated as $\hat{x}_{ui}(\Theta) = b_i(\Theta) + \mathbf{p}_u^T(\Theta) \cdot \mathbf{q}_i(\Theta)$ where b_i is a term denoting the bias of the item i. Among pair-wise approaches for learning-to-rank the items of a catalog, Bayesian Personalized Ranking (BPR) [14] is one of the most broadly adopted, thanks to its capabilities to correctly rank with *acceptable* computational complexity. In detail, given a training set defined by $\mathcal{K} = \{(u, i, j) \mid x_{ui} = 1 \wedge x_{uj} = 0\}$, BPR solves the optimization problem via the criterion $\max_{\Theta} \sum_{(u,i,j) \in \mathcal{K}} \ln\ \sigma(\hat{x}_{uij}(\Theta)) - \lambda \|\Theta\|^2$, where $\hat{x}_{uij}(\Theta) = \hat{x}_{ui}(\Theta) - \hat{x}_{uj}(\Theta)$ is a real value modeling the relation between user u, item i and item j, σ is the sigmoid function, and λ is a regularization parameter to prevent overfitting. Pair-wise optimization can be applied to a wide range of recommendation models, included factorization. Hereafter, we denote the model $\Theta = \langle \mathbf{P}, \mathbf{Q}, \mathbf{b} \rangle$, where $\mathbf{P} \in \mathbb{R}^{|\mathcal{U}| \times F}$ is a matrix whose u-th row corresponds to the vector $\mathbf{p}_u$, and $\mathbf{Q} \in \mathbb{R}^{|\mathcal{I}| \times F}$ is a matrix in which the i-th row corresponds to the vector $\mathbf{q}_i$. Finally, $\mathbf{b} \in \mathbb{R}^{|\mathcal{I}|}$ is a vector whose i-th element corresponds to the value b_i.

3 APPROACH

Following the aforementioned federated learning principles, let $\mathcal{U}$ be the set of users (clients) with a server S coordinating them. Let us assume that users consume items from a catalog $\mathcal{I}$ and give feedback about them (as in the recommendation problem of Section 2). S is aware of the catalog $\mathcal{I}$, while exclusively user u knows her own set of consumed items.

To setup the federation for FPL, a global model is built on S such that $\Theta = \langle \mathbf{Q}, \mathbf{b} \rangle$, where $\mathbf{Q} \in \mathbb{R}^{|\mathcal{I}| \times F}$ and $\mathbf{b} \in \mathbb{R}^{|\mathcal{I}|}$ are the item-factor matrix and the bias vector (introduced in Section 2). On the other hand, on each user u's device FPL builds a model $\Theta_u = \langle \mathbf{p}_u \rangle$, which corresponds to the representation of user u in the latent space of dimensionality F. Each user u holds her own private dataset $\mathbf{x}_u \in \mathbb{R}^{\mathcal{I}}$, which — analogously to a centralized recommender system — corresponds to the u-th row of matrix $\mathbf{X}$. Each FPL client u hosts a user-specific training set $\mathcal{K}_u : \mathcal{U} \times \mathcal{I} \times \mathcal{I}$ defined by $\mathcal{K}_u = \{(u, i, j) \mid x_{ui} = 1 \wedge x_{uj} = 0\}$, where x_{ui} represents the i-th element of x_u. Please note that we refer to $X^+ = \sum_{u \in \mathcal{U}} |\{x_{ui} \mid x_{ui} = 1\}|$ as the number of positive interactions in the system.

The classic BPR-MF learning procedure [14] for model training can not be applied to the federated learning scheme [13]. Instead, we propose a novel learning paradigm that works by rounds of communication and envisages **Distribution** → **Computation** → **Transmission** → **Aggregation** sequences between the server and the clients, whose details are as follows.

(1) **Distribution.** S randomly selects a subset of users $\mathcal{U}^- \subseteq \mathcal{U}$ and delivers them the model Θ_S.
(2) **Computation.** Each user u generates T triples (u, i, j) from her dataset $\mathcal{K}_u$ and for each of them performs BPR stochastic optimization to compute the updates for the local $\mathbf{p}_u$ vector of Θ_u, and for $\mathbf{q}_i$, b_i, $\mathbf{q}_j$, and b_j of the received Θ_S, following:
$$\Delta\theta = \frac{e^{-\hat{x}_{uij}}}{1 + e^{-\hat{x}_{uij}}} \cdot \frac{\partial}{\partial\theta}\hat{x}_{uij} - \lambda\theta,$$
$$\text{with} \quad \frac{\partial}{\partial\theta}\hat{x}_{uij} = \begin{cases} (\mathbf{q}_i - \mathbf{q}_j) & \text{if } \theta = \mathbf{p}_u, \\ \mathbf{p}_u & \text{if } \theta = \mathbf{q}_i, \qquad -\mathbf{p}_u \quad \text{if } \theta = \mathbf{q}_j, \\ 1 & \text{if } \theta = b_i, \qquad -1 \quad \text{if } \theta = b_j. \end{cases}$$
It is worth noticing that Rendle [14] suggests, in a centralized scenario, to adopt a uniform distribution (over $\mathcal{K}$) to choose the training triples randomly. The purpose is to avoid data is traversed item-wise or user-wise, since this may lead to slow convergence. Conversely, in a federated approach, we required to train the model user-wise since the training of each round of communication is performed separately on each client u knowing only data in $\mathcal{K}_u$. This is the reason why, in FPL, the designer can control of the number of triples T used for training, to tune the degree of local computation — i.e., how much the sampling is user-wise traversing.
(3) **Transmission.** The clients in $\mathcal{U}^-$ send back to S a portion of the updates for the computed item factor vector and item bias. More in detail, since the training output of a triple (u, i, j) in BPR lets the server distinguish the consumed item i from the non-consumed one j (for example just by analyzing the positive and the negative sign of Δb_i and Δb_j), while they show the same absolute value, we argue that sending all the updates computed by u may allow S to reconstruct $\mathcal{K}_u$ thus raising a privacy issue. Since our primary goal is to put users in control of their data, FPL proposes a solution to overcome this vulnerability. By sending the sole update $(\Delta\mathbf{q}_j, \Delta b_j)$ of each training triple (u, i, j), user u would share with S indistinguishably negative or missing values,

Table 1: Characteristics of the datasets used for experiments: $|\mathcal{U}|$, $|\mathcal{I}|$, and X^+ are the number of users, items, and records.

Dataset	$\|\mathcal{U}\|$	$\|\mathcal{I}\|$	X^+	$\frac{X^+}{\|\mathcal{U}\|}$	$\frac{X^+}{\|\mathcal{I}\|}$	$\frac{X^+}{\|\mathcal{L}\|\cdot\|\mathcal{U}\|}\%$
Brazil	17,473	47,270	599,958	34.34	12.69	0.00073%
Canada	1,340	29,518	63,514	47.40	2.15	0.00161%
Italy	1,353	25,522	54,088	39.98	2.20	0.00157%

which are assumed to be *non-sensitive* data. Furthermore, in FPL we introduce the parameter π, which allows users to control of the number of consumed items to share with the central server S. In detail, π works as a probability that clients send a specific positive item update $(\Delta\mathbf{q}_i, \Delta b_i)$ in addition to $(\Delta\mathbf{q}_j, \Delta b_j)$.

(4) **Global aggregation.** S aggregates all the received updates in $\mathbf{Q}$ and $\mathbf{b}$ to build the new model $\Theta_S \leftarrow \Theta_S + \alpha \sum_{u\in\mathcal{U}^-} \Delta\Theta_u$, with α being the learning rate (each row of the matrix $\mathbf{Q}$ and each element of $\mathbf{b}$ is updated by summing up the contribution of all clients in $\mathcal{U}^-$ for the corresponding item).

4 EXPERIMENTS

Experimental Setting. FPL needs to be evaluated in a domain that guarantees the availability of transaction data the user may prefer to protect. In our view, the optimal domain would be that of the Point-of-Interest (PoI), which concerns data that users usually perceive as sensitive. Among the many available datasets, a good candidate is the *Foursquare* dataset [16], which is often considered as a reference for evaluating PoI recommendation models. To mimic a federation of devices in a single country, we have extracted check-ins for three countries, namely Brazil, Canada, and Italy, in order to obtain datasets with different size/sparsity characteristics. To fairly evaluate FPL against the baselines, we have kept users with more than 20 interactions[3]. Moreover, we have split the datasets by adopting a realistic temporal hold-out 80-20 splitting on a per-user basis [6]. Table 1 shows the details of the resulting training sets.

To evaluate the efficacy of FPL, we have conducted the experiments by considering non-personalized methods (random and most popular recommendation), and different recommendation approaches, including the centralized **BPR-MF** implementation [14], **VAE** [12], and **FCF** [2], which is, to date, the only federated recommendation approach based on MF (since no source code is available, we reimplemented and considered it in the reader's interest). To evaluate the impact of feedback deprivation on recommendation accuracy, we have evaluated different values of π in the range $[0.0, 1.0]$, with $\pi = 0.0$ meaning that u is not sharing any positive feedback with S, and $\pi = 1.0$ meaning that u is sharing the updates on all positive items. Hence, we have considered two different configurations regarding computation and communication:

- ***s*FPL**: it reproduces the centralized stochastic learning, where the central model is updated sequentially; thus, we set $|\mathcal{U}^-| = 1$ to involve just one random client per round, and it extracts solely one triple (u, i, j) from its dataset ($T = 1$) for the training phase;
- ***p*FPL**: we enable parallelism by involving all clients in each round of communication ($\mathcal{U}^- = \mathcal{U}$); we keep $T = 1$.

In Rendle *et al.* [14], authors suggest to set the number of triples in one epoch of BPR to X^+, which corresponds to the number of optimizations steps. A particular choice is to randomly sampling $T = \frac{X^+}{|\mathcal{U}|}$ triples per user. To compare federated training with BPR and among configurations, we consider rpe rounds of communication of FPL to be equivalent to one epoch of centralized BPR, if rpe is set such that we perform the same overall number of optimization steps. This results in $rpe = X^+$ for *s*FPL, and $rpe = \frac{X^+}{|\mathcal{U}^-|}$ for *p*FPL.

Reproducibility. For the splitting strategy, we have adopted a **temporal hold-out** 80/20 to separate our datasets in training and test set. Moreover, to find the most promising learning rate α, we have further split the training set, adopting a temporal hold-out 80/20 strategy on a user basis to extract her validation set. **VAE** has been trained by considering three autoencoder topologies, with the following number of neurons per layer: 200-100-200, 300-100-300, 600-200-600. We have chosen candidate models by considering the best models after training for 50, 100, and 200 epochs, respectively. For the **factorization models**, we have performed a grid search in BPR-MF for $\alpha \in \{0.005, 0.05, 0.5\}$ varying the number of latent factors in $\{10, 20, 50\}$. Then, to ensure a fair comparison, we have exploited the same learning rate and number of latent factors to train FPL and **FCF**, and we explored the models in the range of $\{10, \ldots, 50\}$ iterations. We have set *user*- and *positive item*-regularization parameter to $\frac{1}{20}$ of the learning rate. The *negative item*-regularization parameter is $\frac{1}{200}$ of the learning rate, as suggested in *mymedialite*[4] implementation as well as by Anelli *et al.* [4].

Evaluation Metrics. We have evaluated the performance of FPL under the accuracy and diversity perspective. The accuracy of the models is measured by exploiting Precision ($P@N$) and Recall ($R@N$). They respectively represent, for each user, the proportion of relevant recommended items in the recommendation list, and the fraction of relevant items that have been altogether suggested. We have assessed the statistical significance of results by adopting Student's paired T-test considering p-values < 0.05[5]. The results are in general statistically significant but the differences among BPR-MF, *s*FPL, and *p*FPL, which is a very important result. To measure the diversity of recommendations, we have measured the Item Coverage ($IC@N$), and the Gini Index ($G@N$). IC provides the number of diverse items recommended to users. It also conveys the sense of the degree of personalization [1]. Gini measures distributional inequality, i.e., how unequally different items a RS provides users with [5]. A higher value of G corresponds to higher personalization [6].

Discussion. The goal of the experiments is assessing whether it is possible to obtain a recommendation performance comparable to a centralized pair-wise learning approach while allowing the users to control their data. In this respect, Table 2 shows the accuracy and diversity results of the comparison between the state-of-the-art baselines and the experimental configurations of FPL presented in Section 4. By focusing on accuracy metrics, we may notice that VAE outperforms the other approaches in the three datasets. However, who is familiar with VAE knows that, since it restricts training data by applying k-core, it does not always produce recommendations for all the users. On the other hand, it is important to investigate the differences of FPL with respect to BPR-MF, which is a pair-wise centralized approach, being FPL the first federated pair-wise

[3]The limitations of the CF in a cold-start user setting are well-known in the literature.

[4]http://www.mymedialite.net/

[5]The complete results are available at https://bit.ly/3kGiXFz.

Table 2: Results of accuracy and beyond-accuracy metrics for baselines and FPL on the three datasets. For each configuration of FPL and for each dataset, the experiment with the best π is shown (see the bottom part for details). For all metrics, the greater the better.

	Brazil				Canada				Italy			
	P@10	R@10	IC@10	G@10	P@10	R@10	IC@10	G@10	P@10	R@10	IC@10	G@10
Random	0.00013	0.00015	46120	0.70946	0.00030	0.00035	10815	0.26809	0.00030	0.00029	10478	0.28914
Top-Pop	0.01909	0.02375	19	0.00020	0.04239	0.04679	18	0.00030	0.04634	0.05506	19	0.00035
VAE *	0.10320	0.13153	5503	0.02117	0.06060	0.06317	1044	0.00652	0.10421	0.21324	165	0.02336
BPR-MF	0.07702	0.09494	2552	0.00756	0.03694	0.03650	1216	0.00998	0.04560	0.05458	19	0.00036
FCF	0.03089	0.03749	911	0.00095	0.03724	0.03836	504	0.00174	0.03126	0.03708	403	0.00158
***s*FPL**	0.07757	0.09581	1581	0.00561	0.04515	0.04550	451	0.00243	0.04701	0.05600	18	0.00036
***p*FPL**	0.07771	0.09582	2114	0.00638	0.04582	0.04637	425	0.00213	0.04642	0.05465	96	0.00056

Best π obtained for each the proposed FPL variations across three countries (Brazil, Canada, and Italy) are: *s*FPL = (0.5, 0.1, 0.4), *s*FPL+ = (0.9, 0.4, 0.2), *p*FPL = (0.8, 0.1, 1), *p*FPL+ = (0.8, 0.3, 0.1)
* VAE does not always produce recommendations for all the users. For Italy, the reported results cover the 14% of the users

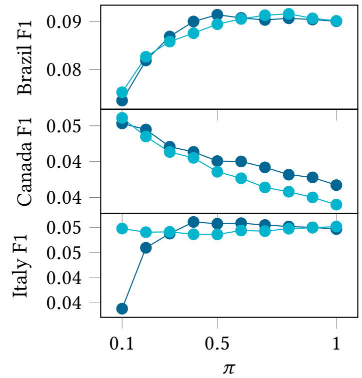

Figure 1: F1 performance at different values of π in the range $[0.1, 1]$. Dark blue is *s*FPL, light blue *p*FPL.

recommender based on a factorization model. The performance of BPR-MF against FPL, in the configuration *s*FPL, shows how Precision and Recall in *s*FPL are slightly outperforming BPR-MF while achieving very similar diversity values. The consideration that the performance is comparable is surprising since the two methods share the sequential training, but *s*FPL exploits a π reduced to 0.5, 0.1, and 0.4, respectively, for Brazil, Canada, and Italy. This behavior is more evident in Figure 1, where the harmonic mean between Precision and Recall (F1) is plotted for different values of π. If we look at the dark blue line, we may observe how the best result does not correspond to $\pi = 1$. When comparing *p*FPL with *s*FPL, we observe that the increased parallelism does not affect the performance significantly. As a concluding remark, we may affirm that *the proposed system can generate recommendations with a quality that is comparable with the centralized pair-wise learning approach*, and that *the training parallelism does not significantly affect results.* Afterwards, we varied π in the range $[0.1, \ldots, 1.0]$ to investigate how removal of the updates for consumed items affects the final recommendation accuracy, and we plotted the accuracy performance by considering F1 in Figure 1. The best performance rarely corresponds to $\pi = 1$. On the contrary, the training reaches a peak for a certain value of π, and then the system performance decays in accuracy when increasing the amount of shared positive updates. In rare cases, e.g., *s*FPL, and *p*FPL for Brazil dataset, the decay is absent, but results that are very close for different values of π. The general behavior suggests that the system learning exploits the updates of positive items to absorb information about popularity. This consideration is coherent with the mathematical formulation of the learning procedure, and it is also supported by the observation that for Canada and Italy FPL reaches the peak before with respect to Brazil. Indeed, Canada and Italy datasets are less sparse than Brazil, and the increase of information about positive items may lead to push up too much the popular items (this is a characteristic of pair-wise learning), while the same behavior in Brazil can be observed for $\pi \simeq 1$. Ultimately, *users can receive high-quality recommendations, also when disclosing a small amount of sensitive data.*

5 CONCLUSION AND FUTURE WORK

We proposed FPL, a novel federated learning framework that exploits pair-wise learning for factorization models. We have designed a model that leaves the user-specific information of the original factorization model in the clients' devices so that a user may be completely in control of her sensitive data and could share no positive feedback with the server. The framework can be envisioned as a general factorization model in which clients can tune the amount of information shared among devices. We have conducted an exploratory, but extensive, experimental evaluation to analyze the degree of accuracy, the diversity of the recommendation results, the trade-off between accuracy, and amount of shared transactions. We have assessed that the proposed model shows performance comparable with several state-of-the-art baselines and the classic centralized factorization model with pair-wise learning. The evaluation shows that clients may share a small portion of their data with the server and still receive high-performance recommendations. We believe that the proposed privacy-oriented paradigm may open the doors to a new class of ubiquitous recommendation engines.

REFERENCES

[1] G. Adomavicius and Y. Kwon. Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Trans. Knowl. Data Eng.*, 24(5):896–911, 2012.
[2] M. Ammad-ud-din, E. Ivannikova, S. A. Khan, W. Oyomno, Q. Fu, K. E. Tan, and A. Flanagan. Federated collaborative filtering for privacy-preserving personalized recommendation system. *CoRR*, abs/1901.09888, 2019.
[3] V. W. Anelli, Y. Deldjoo, T. Di Noia, and A. Ferrara. Towards effective device-aware federated learning. In *International Conference of the Italian Association for Artificial Intelligence*, pages 477–491. Springer, 2019.
[4] V. W. Anelli, T. D. Noia, E. D. Sciascio, C. Pomo, and A. Ragone. On the discriminative power of hyper-parameters in cross-validation and how to choose them. In *Proc. of the 13th ACM Conf. on Recommender Systems*, pages 447–451. ACM, 2019.
[5] P. Castells, N. J. Hurley, and S. Vargas. Novelty and diversity in recommender systems. In F. Ricci, L. Rokach, and B. Shapira, editors, *Recommender Systems Handbook*, pages 881–918. Springer, 2015.
[6] A. Gunawardana and G. Shani. Evaluating recommender systems. In F. Ricci, L. Rokach, and B. Shapira, editors, *Recommender Systems Handbook*, pages 265–308. Springer, 2015.
[7] A. J. P. Jeckmans, M. Beye, Z. Erkin, P. H. Hartel, R. L. Lagendijk, and Q. Tang. Privacy in recommender systems. In N. Ramzan, R. van Zwol, J. Lee, K. Clüver, and X. Hua, editors, *Social Media Retrieval*, Computer Communications and Networks, pages 263–281. Springer, 2013.
[8] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, et al. Advances and open problems in federated learning. *arXiv preprint arXiv:1912.04977*, 2019.
[9] J. Konecný, H. B. McMahan, D. Ramage, and P. Richtárik. Federated optimization: Distributed machine learning for on-device intelligence. *CoRR*, abs/1610.02527, 2016.
[10] Y. Koren, R. M. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8):30–37, 2009.

[11] D. kumar Bokde, S. Girase, and D. Mukhopadhyay. Role of matrix factorization model in collaborative filtering algorithm: A survey. *CoRR*, abs/1503.07475, 2015.
[12] D. Liang, R. G. Krishnan, M. D. Hoffman, and T. Jebara. Variational autoencoders for collaborative filtering. In *Proceedings of the 2018 World Wide Web Conference*, pages 689–698, 2018.
[13] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, et al. Communication-efficient learning of deep networks from decentralized data. *arXiv preprint arXiv:1602.05629*, 2016.
[14] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. BPR: bayesian personalized ranking from implicit feedback. In J. A. Bilmes and A. Y. Ng, editors, *UAI 2009, Proc. of the Twenty-Fifth Conf. on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, June 18-21, 2009*, pages 452–461. AUAI Press, 2009.
[15] S. Weiss. The need for a paradigm shift in addressing privacy risks in social networking applications. In *IFIP International Summer School on the Future of Identity in the Information Society*, pages 161–171. Springer, 2007.
[16] D. Yang, D. Zhang, and B. Qu. Participatory cultural mapping based on collective behavior data in location-based social networks. *ACM TIST*, 7(3):30:1–30:23, 2016.