GD-DR Gradient Descent with Dimensionality Reduction

Privacy Enhanced Matrix Factorization for Recommendation with Local Differential Privacy

Hyejin Shin
Samsung Research
hyejin1.shin@samsung.com

Sungwook Kim
Samsung Research
sw14.kim@samsung.com

Junbum Shin
Samsung Research
junbum.shin@samsung.com

Xiaokui Xiao
National University of Singapore
xkxiao@nus.edu.sg

TKDE (2018)

Deficiency

- Protect either uses' ratings or the items rated by user, but not both
  - A user's item set is as sensitive as her ratings since inference from which items a user has rated can discover sensitive information
- Focus on the protection of a single item or rating value among a user's entire items and their ratings
  - Masking the presence or absence of a single item among the correlated items is not secure enough
  - Users often upload a set of ratings at a time

Goal

- To build a recommender system
  - Individual users randomize their data themselves and make a recommendation locally in their devices
  - The recommender computes aggregates of the perturbed data

Privacy requirements

- Privacy of a users' entire ratings and items
  - Protect both user's rating and items
  - Guarantee a stronger degree of privacy: *per-user privacy*
- Privacy over all iterations of gradient descent
  - Inference attack
    - The difference of gradients at the sth and tth iterations ($1 \leq s < t \leq k$) leaks some information about user's ratings or items

Issue [DPMF]

- Which items are rated by a user
  - The recommender requests the users who have rated item *j* to submit their gradients in a private manner
  - The aggregator can learn whether a user has rated item *j*

Solution: Randomized Response Mechanism

- Let y_{ij} be 1, if user i rates item j, and 0, otherwise. Suppose that $r_{ij} > 0$ for item j rated by user i. We let $r_{ij} = 0$ for $(i,j) \notin \mathcal{M}$. Then, observe that

$$\sum_{(i,j)\in\mathcal{M}} \left(r_{ij} - {u_i}^T v_j\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij}\left(r_{ij} - {u_i}^T v_j\right)^2,$$

  So that $\nabla_{v_j}\phi(U,V) = -\frac{2}{M}\sum_{i:\,(i,j)\in\mathcal{M}} u_i\left(r_{ij} - {u_i}^T v_j\right)$ is written as

$$\nabla_{v_j}\phi(U,V) = -2n^{-1}\sum_{i=1} y_{ij}\, u_i\left(r_{ij} - {u_i}^T v_j\right).$$

- Perturb the item vector $Y_i = \left(y_{ij}\right)_{1\le j\le m}$ using the randomized response mechanism. Specifically, we generate y_{ij}^* from y_{ij} as follows:

$$y_{ij}^* = \begin{cases} 0, & with\ probability\ p/2 \\ 1, & with\ probability\ p/2 \\ y_{ij}, & with\ probability\ 1-p \end{cases}$$

  for $j = 1, \ldots, m$.

Solution: Randomized Response Mechanism

- To protect $g_{ij} = (g_{ijl})_{1 \le l \le d} = -2u_i(r_{ij} - {u_i}^T v_j)$, individual users perturb the g_{ij} in their devices such that

$$g^*_{ijl} = g_{ij} + \eta_{ij},$$

where the η_{ij} are independent random variables that follow a Laplace distribution with scale σ. If each user submits $\{(y^*_{ij}, g^*_{ij1}, \dots, g^*_{ijd}) : j = 1, \dots, m\}$ to the server, then it can be shown that the mean of the individual gradients, $\nabla_{v_j} = n^{-1} \sum_{i=1}^{n} y_{ij} g_{ij}$, could be estimated by

$$\nabla^*_{v_j} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y^*_{ij} - p/2}{1-p} \right) g_{ij} .$$

The server updates the item profiles v_j with the mean of the noisy gradients, i.e., $v_j^t = v_j^{t-1} - \gamma_t \cdot \left\{ \nabla^{*,t-1}_{v_j} + 2\lambda_v \mathrm{v}_j^{t-1} \right\}$.

Solution: Randomized Response Mechanism

- If we set $p = 2/\left(1 + e^{\epsilon_1/k}\right)$ and $\sigma = \frac{\Delta d}{\epsilon_2/k}$ in each iteration with $\Delta = \max_{i,j} r_{ij} - \min_{i,j} r_{ij}$ and $\epsilon = \epsilon_1 + \epsilon_2$, the final V satisfies ϵ-differential privacy for two datasets D and D' differing in one item or rating of one user.
- If we set $p = 2/\left(1 + e^{(\epsilon_1/k)/m}\right)$ and $\sigma = \frac{m\Delta d}{\epsilon_2/k}$ in each iteration, the final V satisfies ϵ-differential privacy for two datasets D and D' differing in one user's entire items and ratings.
  - *Per-user privacy* requirement results in the estimation error for V linearly growing with m, given as $O\left(\frac{md\sqrt{log(md)}}{(\epsilon_1/k)(\epsilon_2/k)\sqrt{n}}\right)$.

Attempt 1: Gradient Perturbation

- Randomized method
  - Consider the scenario where each user has a multiple-dimensional tuple to be collected by a server
    - When submitting her tuple, the user randomly selects one dimension, and submits a perturbed version of her tuple value on the selected dimension, while ignoring all other dimensions.
  - The perturbed values collected by the server can be used to estimate the mean of all users' values on each dimension, and the estimation error is significantly smaller than the case when each user submits her values on all dimensions

Thông T. Nguyên, Xiaokui Xiao, Yin Yang, Siu Cheung Hui, Hyejin Shin, Junbum Shin: Collecting and Analyzing Data from Smart Device Users with Local Differential Privacy. CoRR 2016

Algorithm 1. Differentially Private Gradient Descent

input: Predefined iteration number k and privacy parameter ϵ
output: Item profile matrix $V \in \mathbb{R}^{m \times d}$

Initialize U, V and a counter $iter = 0$
while $iter \leq k$ **do**
  Initialize $\nabla_V^* \in \{0\}^{m \times d}$
  for $i = 1$ to n **do**
    Initialize $x_i^* \in \{0\}^{m \times d}$
    Sample j uniformly at random from $\{1, 2, \ldots, m\}$
    Sample l uniformly at random from $\{1, 2, \ldots, d\}$
    Compute $(x_i)_{j,l} = -2y_{ij}u_{il}(r_{ij} - u_i^T v_j)$
    If $(x_i)_{j,l} \notin [-1, 1]$, project $(x_i)_{j,l}$ onto $[-1, 1]$
    Draw $T \sim Bernoulli\left(\frac{(x_i)_{j,l}(e^{\epsilon/k}-1)+e^{\epsilon/k}+1}{2(e^{\epsilon/k}+1)}\right)$
    if $T = 1$ **then** $(x_i^*)_{j,l} = md\frac{e^{\epsilon/k}+1}{e^{\epsilon/k}-1}$
    else $(x_i^*)_{j,l} = -md\frac{e^{\epsilon/k}+1}{e^{\epsilon/k}-1}$
    end if
    Compute $\nabla_V^* = \nabla_V^* + x_i^*$
  end for
  Compute $\nabla_V^* = \nabla_V^*/n$
  $iter = iter + 1$
  $V = V - \gamma_t \cdot \{\nabla_V^* + 2\lambda_v V\}$
  for $i = 1$ to n **do**
    Derive ∇_{u_i} in (4) and $u_i = u_i - \gamma_t \cdot \{\nabla_{u_i} + 2\lambda_u u_i\}$
  end for
end while
return V

Private GD

降低了通讯代价

每位用户从 $m \times d$ 的梯度矩阵中随机选择一个元素来更新 V

每位用户向聚合器发送这个元素有噪声的梯度，不管用户的商品，它的值是 $B\left(= md\frac{e^{\epsilon/k}+1}{e^{\epsilon/k}-1}\right)$ 或 $-B$，敌手无法学习她评价的商品或她对商品的偏好

Attempt 1: Gradient Perturbation

Lemma 1. *Let x_i be the non-private gradient matrix of user i and x_i^* a noisy gradient matrix of user i in Algorithm 1. Assume that $x_i \in [-1, 1]^{m \times d}$. Then, each submission of user's gradients satisfies ϵ/k-LDP, so noisy gradients submitted by each user over k iterations satisfy ϵ-LDP.*

Proof. Similarly to Nguyen et al. [21], we have

$$\frac{Pr[x_i^* = v|x_i]}{Pr[x_i^* = v|x_i']} \leq \frac{\max_{x_i}\{(x_i)_{j,l}(e^{\epsilon/k} - 1) + e^{\epsilon/k} + 1\}}{\min_{x_i'}\{(x_i')_{j,l}(e^{\epsilon/k} - 1) + e^{\epsilon/k} + 1\}}$$
$$= e^{\epsilon/k}.$$

□

Thông T. Nguyên, Xiaokui Xiao, Yin Yang, Siu Cheung Hui, Hyejin Shin, Junbum Shin: Collecting and Analyzing Data from Smart Device Users with Local Differential Privacy. CoRR 2016

Theorem 1. *Algorithm 1 satisfies ϵ-differential privacy.*

Proof. Suppose that two datasets D and D' differ in the ratings of user p. For $w \in \mathbb{R}^{m \times d}$, observe that

$$\begin{aligned} Pr\left[\nabla_V^* = w|D\right] &= Pr\left[\sum_{i=1}^{n} x_i^* = nw \middle| D\right] \\ &= \sum_{\substack{\tilde{w}_1,\ldots,\tilde{w}_n \\ \tilde{w}_1+\cdots+\tilde{w}_n=nw}} \prod_{i=1}^{n} Pr\left[x_i^* = \tilde{w}_i|D\right] \\ &= \sum_{\substack{\tilde{w}_1,\ldots,\tilde{w}_n \\ \tilde{w}_1+\cdots+\tilde{w}_n=nw}} \prod_{i=1}^{n} Pr\left[x_i^* = \tilde{w}_i|x_i\right], \end{aligned}$$

where the probability space is over the coin flips of the binary randomization in Algorithm 1. Thus, we have

$$\begin{aligned} \frac{Pr[\nabla_V^* = w|D]}{Pr[\nabla_V^* = w|D']} &\leq \max_{\substack{\tilde{w}_1,\ldots,\tilde{w}_n \\ \tilde{w}_1+\cdots+\tilde{w}_n=nw}} \prod_{i=1}^{n} \frac{Pr\left[x_i^* = \tilde{w}_i|x_i\right]}{Pr\left[x_i^* = \tilde{w}_i|x_i'\right]} \\ &= \max_{\tilde{w}_p} \frac{Pr\left[x_p^* = \tilde{w}_p|x_p\right]}{Pr\left[x_p^* = \tilde{w}_p|x_p'\right]} \leq e^{\epsilon/k}, \end{aligned}$$

which means that the derived V at each iteration satisfies ϵ/k-differential privacy. Therefore, for the final item profile V, we have

$$\frac{Pr[V = \bar{V}|D]}{Pr[V = \bar{V}|D']} \leq \max_{a_1,\ldots,a_k} \prod_{t=1}^{k} \frac{Pr[\nabla_V^{*,t} = a_t|D]}{Pr[\nabla_V^{*,t} = a_t|D']} \leq e^{\epsilon}.$$

□

在每轮迭代时聚合用户提交的有噪声的梯度，聚合器无法区分最终的商品特征 V 是否包含某名用户

保证 per-user privacy，抵抗重识别攻击

Theorem 2. *Suppose that the learning rate γ_t decreases as iteration goes and satisfies $\sum_{t=1}^{\infty} \gamma_t^2 < \infty$. Let V^* be the item profile matrix derived by Algorithm 1 and V the item profile matrix in non-private manner after k iterations. With at least $1-\beta$ probability,*

$$\|V^* - V\|_{\max} = O\left(\frac{\sqrt{md\log(md/\beta)}}{(\epsilon/k)\sqrt{n}}\right).$$

Proof. First observe that $V^* - V = -\sum_{t=1}^{k} \gamma_t c_t (\nabla_V^{*,t} - \nabla_V^t)$ with $c_t = \prod_{j=t+1}^{k}(1 - 2\lambda_v \gamma_j)$. The random variables $(\nabla_V^{*,t})_{j,l}$, $1 \le t \le k$, are independent and bounded by $O(\frac{md}{\epsilon/k})$. Also, $(\nabla_V^{*,t})_{j,l}$ is an unbiased estimator of $(\nabla_V^t)_{j,l}$, and $\mathrm{Var}((\nabla_V^{*,t})_{j,l}) = O(\frac{md}{(\epsilon/k)^2 n})$. Then, by Bernstein's inequality, we have

$$\begin{aligned}
&Pr\left[\left\|\sum_{t=1}^{k} \gamma_t c_t (\nabla_V^{*,t} - \nabla_V^t)\right\|_{\max} > \lambda \,\middle|\, D\right] \\
&\le \sum_{j=1}^{m}\sum_{l=1}^{d} Pr\left[\left|\sum_{t=1}^{k} \gamma_t c_t ((\nabla_V^{*,t})_{j,l} - (\nabla_V^t)_{j,l})\right| > \lambda \,\middle|\, D\right] \\
&\le 2 \cdot md \cdot \exp\left(-\frac{\lambda^2}{2\sum_{t=1}^{k} \gamma_t^2 c_t^2 \mathrm{Var}((\nabla_V^{*,t})_{j,l}) + \frac{2}{3}\lambda md \frac{e^{\epsilon/k}+1}{e^{\epsilon/k}-1}}\right) \\
&= O\Big(md \cdot \exp\big(-n\lambda^2(\epsilon/k)^2/md\big)\Big)
\end{aligned}$$

since $\sum_{t=1}^{k} \gamma_t^2 < \infty$ and $|c_t| \le 1$ for small $\lambda_v \downarrow 0$. Taking $\lambda = O(\frac{\sqrt{md\log(md/\beta)}}{(\epsilon/k)\sqrt{n}})$, the desired rate for $\|V^* - V\|_{\max}$ is achieved. □

在每轮迭代中，梯度 ∇_{v_j} 的估计误差从 $O(m)$ 减少到 $O(\sqrt{m})$

k 轮迭代后，减少了最终商品特征 V 的估计误差

Attempt 2: Dimension Reduction

- The error may still be excessively large
  - m is an enormous number, typically ranging from 10^4 to 10^6
- Random projection
  - Reduce the dimensionality of users' data without prior knowledge of the data
  - Restricted isometry property so that the distances between the points in a randomly selected subspace of suitably high dimension are approximately preserved

Attempt 2: Dimension Reduction

- For $q \ll m$, let Φ be a $q \times m$ random matrix whose elements ϕ_{kj} are independent random variables from Gaussian distribution with mean 0 and variance $1/q$. From the Johnson-Lindenstrauss lemma, we have

$$(1-\delta)\left\|r_i - V_{u_i}\right\|^2 \le \left\|\Phi\left(r_i - V_{u_i}\right)\right\|^2 \le (1+\delta)\left\|r_i - V_{u_i}\right\|^2,$$

for $q = O(\delta^{-2}\log n)$. Since $(\Phi a)^T(\Phi b) = (\|\Phi(a+b)\|^2 - \|\Phi a\|^2 - \|\Phi b\|^2)/2$, we have

$$|(\Phi a)^T(\Phi b) - a^T b| \le \frac{3}{2}\delta(\|a\|^2 + \|b\|^2),$$

for any $a, b \in \mathbb{R}^m$. It follows that

$$\sum_{j=1}^{m} y_{ij}\left(r_{ij} - {u_i}^T v_j\right)^2 = {r_i}^T \Phi^T \Phi_{r_i} - 2{r_i}^T \Phi^T \Phi V_{u_i} + {u_i}^T V^T \Phi^T \Phi D_i \Phi^T \Phi V_{u_i} + O(\delta)$$

where $D_i = diag(y_{i1}, \dots, y_{im}), r_i = \left(r_{ij}\right)_{1 \le j \le m}$ and $V = \left(v_{jl}\right) = [v_1, \dots, v_m]^T$ with the d-dimensional vector v_j. Letting $z_i = \Phi_{r_i} = (z_{ik})_{1 \le k \le q}, B = \Phi V$, and $W_i = \Phi D_i \Phi^T = \left(w_{i,kk'}\right)_{1 \le k,k' \le q}$, the problem of finding the item profiles V becomes the problem of finding $B = \left[b_1, \dots, b_q\right]^T$, where $b_k \in \mathbb{R}^d$, that minimizes

$$n^{-1}\sum_{i=1}^{n}\{{z_i}^T z_i - 2{z_i}^T B u_i + {u_i}^T B^T W_i B u_i\} + \lambda_v \sum_{k=1}^{q}\|b_k\|^2,$$

for given $\{u_i \in \mathbb{R}^d, 1 \le i \le n\}$. Then, the update formula is

$$B^t = B^{t-1} - \gamma_t \cdot \{\nabla_B \psi(U^{t-1}, B^{t-1}) + 2\lambda_v B^{t-1}\},$$

where $\nabla_B \psi(U, B) = -2n^{-1}\sum_{i=1}^{n}(z_i - W_i B u_i)\,{u_i}^T$.

Algorithm 2. Differentially Private Gradient Descent with Dimension Reduction

input: Positive integer q, predefined iteration number k, and privacy parameter ϵ
output: Item profile matrix $V \in \mathbb{R}^{m \times d}$

Generate a $q \times m$ random matrix Φ whose entries are drawn from Gaussian distribution with mean 0 and standard deviation $1/\sqrt{q}$ and send Φ to users
Initialize U, V and a counter $iter = 0$
while $iter \leq k$ **do**
  Initialize $\nabla_B^* \in \{0\}^{q \times d}$
  for $i = 1$ to n **do**
    Initialize $x_i^* \in \{0\}^{q \times d}$
    Derive $\nabla_V^i = \{-2y_{ij}u_i(r_{ij} - u_i^T v_j)\}_{1 \leq j \leq m}$
    Compute $x_i = \Phi \nabla_V^i$
    Sample s uniformly at random from $\{1, 2, \ldots, q\}$
    Sample l uniformly at random from $\{1, 2, \ldots, d\}$
    If $(x_i)_{s,l} \notin [-1, 1]$, project $(x_i)_{s,l}$ onto $[-1, 1]$
    Draw $T \sim Bernoulli\left(\frac{(x_i)_{s,l}(e^{\epsilon/k}-1)+e^{\epsilon/k}+1}{2(e^{\epsilon/k}+1)}\right)$
    if $T = 1$ **then** $(x_i^*)_{s,l} = qd\frac{e^{\epsilon/k}+1}{e^{\epsilon/k}-1}$
    else $(x_i^*)_{s,l} = -qd\frac{e^{\epsilon/k}+1}{e^{\epsilon/k}-1}$
    end if
    Compute $\nabla_B^* = \nabla_B^* + x_i^*$
  end for
  Compute $\nabla_B^* = \nabla_B^*/n$ and send ∇_B^* to users
  $iter = iter + 1$
  for $i = 1$ to n **do**
    Derive ∇_{u_i} in (4) and $u_i = u_i - \gamma_t \cdot \{\nabla_{u_i} + 2\lambda_u u_i\}$
    $V = V - \gamma_t \cdot \{\Phi^\dagger \nabla_B^* + 2\lambda_v V\}$
  end for
end while
return V

Private GD-DR

用户发送从 $q \times d$ 的矩阵中随机选择的元素有噪声的梯度

迭代 k 次

聚合器收集所有用户降维的梯度

Corollary 1. *Algorithm 2 satisfies ϵ-differential privacy.*

Proof. Let $x_i, x_i' \in [-1,1]^{q\times d}$ be any two gradient matrices of user i. Also, let x_i^* be a perturbed gradient matrix of user i by Algorithm 2. As shown in Theorem 1, we have

$$\frac{Pr[x_i^* = v|x_i]}{Pr[x_i^* = v|x_i']} \le e^{\epsilon/k}.$$

Now let D and D' be two datasets differing in the ratings of user p. For $w \in \mathbb{R}^{m\times d}$, observe that

$$Pr\left[\Phi^\dagger \nabla_B^{*,t} = w|D\right] = Pr\left[\frac{1}{n}\sum_{i=1}^n x_i^{*,t} = \Phi w \middle| D\right] = \sum_{\substack{\tilde{w}_1,\ldots,\tilde{w}_n \\ \tilde{w}_1+\cdots+\tilde{w}_n = n\Phi w}} \prod_{i=1}^n Pr\left[x_i^{*,t} = \tilde{w}_i|x_i\right],$$

where the probability space is over the coin flips of the binary randomization in Algorithm 2. Note that a random matrix Φ is generated by a server and then shared with users, so we treat Φ as a given matrix. As in Theorem 1, for the final item profile V, we have

$$\begin{aligned}\frac{Pr[V=\bar{V}|D]}{Pr[V=\bar{V}|D']} &\le \max_{a_1,\ldots,a_k} \prod_{t=1}^k \frac{Pr[\Phi^\dagger\nabla_B^{*,t} = a_t|D]}{Pr[\Phi^\dagger\nabla_B^{*,t} = a_t|D']} \\ &\le \max_{a_1,\ldots,a_k} \prod_{t=1}^k \max_{\tilde{w}_p^t} \frac{Pr\left[x_p^{*,t} = \tilde{w}_p^t|x_p\right]}{Pr\left[x_p^{*,t} = \tilde{w}_p^t|x_p'\right]} \\ &\le e^\epsilon. \end{aligned}$$

□

Corollary 2. *Suppose that the learning rate γ_t decreases as iteration goes and satisfies $\sum_{t=1}^\infty \gamma_t^2 < \infty$. Let V^* be the item profile matrix derived by Algorithm 2 and V the item profile matrix in non-private manner after k iterations. With at least $1-\beta$ probability,*

$$\|V^* - V\|_{\max} = O\left(\frac{qd\sqrt{\log(qd/\beta)}}{(\epsilon/k)\sqrt{n}}\right).$$

Proof. Observe that $V^* - V = -\sum_{t=1}^k \gamma_t c_t(\Phi^\dagger\nabla_B^{*,t} - \nabla_V^t)$ with $c_t = \prod_{j=t+1}^k (1-2\lambda_v\gamma_j)$ for $1 \le t < k$ and $c_k = 1$. From the Johnson-Lindenstrauss lemma [14], we have

$$\begin{aligned}&\left\|\sum_{t=1}^k \gamma_t c_t(\Phi^\dagger\nabla_B^{*,t} - \nabla_V^t)\right\|_{\max} \\ &\le \left\|\sum_{t=1}^k \gamma_t c_t(\Phi^\dagger\nabla_B^{*,t} - \nabla_V^t)\right\|_F \\ &\le (1-\delta)^{-1/2}\left\|\sum_{t=1}^k \gamma_t c_t(\nabla_B^{*,t} - \Phi\nabla_V^t)\right\|_F \\ &\le (1-\delta)^{-1/2}\sqrt{qd}\left\|\sum_{t=1}^k \gamma_t c_t(\nabla_B^{*,t} - \Phi\nabla_V^t)\right\|_{\max}.\end{aligned}$$

Now note that the random variables $(\nabla_B^{*,t})_{j,l}, 1 \le t \le k$, are independent and bounded by $O(\frac{qd}{\epsilon/k})$, $(\nabla_B^{*,t})_{j,l}$ is an unbiased estimator of $(\Phi\nabla_V^t)_{j,l}$, and $\mathrm{Var}((\nabla_B^{*,t})_{j,l}) = O(\frac{qd}{(\epsilon/k)^2 n})$. Similarly to the proof of Theorem 2, by Bernstein's inequality, we have

$$\begin{aligned}&Pr\left[\left\|\sum_{t=1}^k \gamma_t c_t(\nabla_B^{*,t} - \Phi\nabla_V^t)\right\|_{\max} > \lambda \,\middle|\, D\right] \\ &\le 2\cdot qd\cdot\exp\left(-\frac{\lambda^2}{2\sum_{t=1}^k \gamma_t^2 c_t^2 \mathrm{Var}((\nabla_B^{*,t})_{j,l}) + \frac{2}{3}\lambda qd\frac{e^{\epsilon/k}+1}{e^{\epsilon/k}-1}}\right) \\ &= O\Big(qd\cdot\exp\big(-n\lambda^2(\epsilon/k)^2/qd\big)\Big).\end{aligned}$$

Taking $\lambda = O\left(\frac{\sqrt{qd\log(qd/\beta)}}{(\epsilon/k)\sqrt{n}}\right)$, we have

$$\left\|\sum_{t=1}^k \gamma_t c_t(\nabla_B^{*,t} - \Phi\nabla_V^t)\right\|_{\max} = O\left(\frac{\sqrt{qd\log(qd/\beta)}}{(\epsilon/k)\sqrt{n}}\right),$$

so the desired rate for $\|V^* - V\|_{\max}$ is achieved. □

Experiments

- Baseline
  - Non-private GD
- Comparison
  - Hua et al's [DPMF]
- Metric
  - RMSE

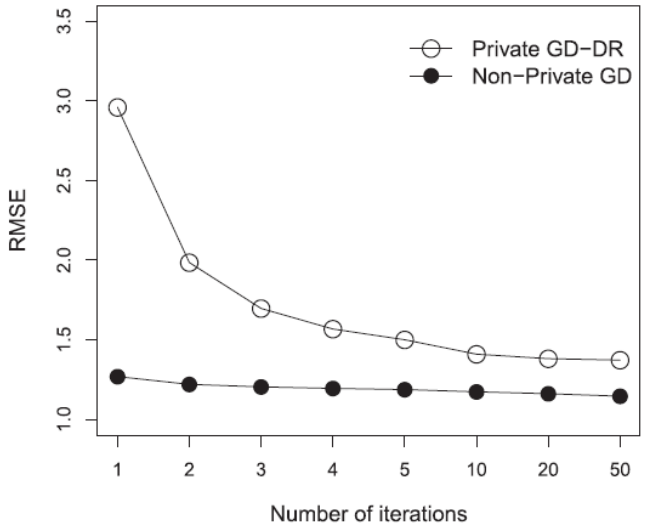
Experimental Setup

- Datasets
  - MovieLens
    - 20M ratings (0.5 to 5)
    - 26,744 movies
    - 138,493 users
  - LibimSeTi dating recommendation dataset
    - 13,689,250 / 17,359,346 ratings (1 to 10)
    - 26,509 / 168,791 profiles
    - 135,359 users
    - 0.38 / 0.07 percent density
      - Profiles having more than 140 ratings
- Choice of parameters
  - $\lambda_u = \lambda_v = 10^{-8}$
  - $q = 2700$ for dimension reduction
  - $d = 15, 20$ low rank approximation of MovieLens and LibimSeTi, respectively
  - γ_t to $O(1/t)$ at iteration t
    - "The learning rate γ_t is a constant, typically set to $O(1/t)$."
  - ϵ from 0.05 to 1.6

Herbert Robbins, Sutton Monro: A Stochastic Approximation Method. Ann. Math. Statist. 1951

Evaluation Results

- Prediction accuracy over various numbers of iterations
  - RMSE decreases as iteration goes when $\epsilon = 0.1$



	k					
	1	3	5	10	20	50
Non-Private GD	1.269	1.204	1.187	1.173	1.161	1.145
Private GD-DR	2.960	1.697	1.500	1.409	1.381	1.372

Fig. 9. MovieLens - Prediction RMSEs of private GD-DR over k iterations when $\epsilon = 0.1$. The values in the table are RMSEs for private GD-DR and non-private GD.