Differentially Private Recommendation Systems

Anupam Datta

Fall 2015
Netflix $1,000,000 Prize Competition

<table>
<thead>
<tr>
<th>User/Movie</th>
<th>300</th>
<th>The Notebook</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>John</td>
<td>4</td>
<td>Unrated</td>
</tr>
<tr>
<td>Mary</td>
<td>Unrated</td>
<td>Unrated</td>
</tr>
<tr>
<td>Sue</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Joe</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Queries: On a scale of 1 to 5 how would John rate “The Notebook” if he watched it?
**Netflix Prize Competition**

<table>
<thead>
<tr>
<th>User/Movie</th>
<th>….</th>
<th>13,537</th>
<th>13,538</th>
<th>….</th>
</tr>
</thead>
<tbody>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>258,964</td>
<td>(4, 10/11/2005)</td>
<td>Unrated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>258,965</td>
<td>Unrated</td>
<td>Unrated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>258,966</td>
<td>(2, 6/16/2005)</td>
<td>(5, 6/18/2005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

**Note:** N x M table is very sparse (M = 17,770 movies, N = 500,000 users)

**To Protect Privacy:**
- Each user was randomly assigned to a globally unique ID
- Only 1/10 of the ratings were published
- The ratings that were published were perturbed a little bit
Root Mean Square Error

\[
RMSE(P) = \sqrt{\frac{1}{k} \sum_{i=1}^{k} (p_i - a_i)^2}
\]

- \( p_i \in [1,5] \) - predicted ratings
- \( a_i \in [1,5] \) - actual ratings
Netflix Prize Competition

**Goal:** Make accurate predictions as measured by Root Mean Squared Error (RMSE)

\[
RMSE(P) = \sqrt{\frac{1}{k} \sum_{i=1}^{k} (p_i - a_i)^2}
\]

- \( p_i \in [1, 5] \) - predicted ratings
- \( a_i \in [1, 5] \) - actual ratings

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>BellKor's Pragmatic Chaos</td>
<td>0.8567 &lt; 0.8572</td>
</tr>
<tr>
<td>Challenge: 10% Improvement</td>
<td>0.8572</td>
</tr>
<tr>
<td>Netflix’s Cinematch (Baseline)</td>
<td>0.9525</td>
</tr>
<tr>
<td>Rank</td>
<td>Team Name</td>
</tr>
<tr>
<td>------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>1</td>
<td>BellKor's Pragmatic Chaos</td>
</tr>
<tr>
<td>2</td>
<td>The Ensemble</td>
</tr>
<tr>
<td>3</td>
<td>Grand Prize Team</td>
</tr>
<tr>
<td>4</td>
<td>Opera Solutions and Vandelay United</td>
</tr>
<tr>
<td>5</td>
<td>Vandelay Industries !</td>
</tr>
<tr>
<td>6</td>
<td>PragmaticTheory</td>
</tr>
<tr>
<td>7</td>
<td>BellKor in BigChaos</td>
</tr>
<tr>
<td>8</td>
<td>Dace</td>
</tr>
<tr>
<td>9</td>
<td>Feeds2</td>
</tr>
<tr>
<td>10</td>
<td>BigChaos</td>
</tr>
<tr>
<td>11</td>
<td>Opera Solutions</td>
</tr>
<tr>
<td>12</td>
<td>BellKor</td>
</tr>
</tbody>
</table>

**Progress Prize 2008 - RMSE = 0.8627 - Winning Team: BellKor in BigChaos**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Team Name</th>
<th>Best Test Score</th>
<th>% Improvement</th>
<th>Best Submit Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>xiangliang</td>
<td>0.8642</td>
<td>9.27</td>
<td>2009-07-15 14:53:22</td>
</tr>
<tr>
<td>14</td>
<td>Gravity</td>
<td>0.8643</td>
<td>9.26</td>
<td>2009-04-22 18:31:32</td>
</tr>
<tr>
<td>15</td>
<td>Ces</td>
<td>0.8651</td>
<td>9.18</td>
<td>2009-06-21 19:24:53</td>
</tr>
<tr>
<td>16</td>
<td>Invisible Ideas</td>
<td>0.8653</td>
<td>9.15</td>
<td>2009-07-15 15:53:04</td>
</tr>
<tr>
<td>17</td>
<td>Just a guy in a garage</td>
<td>0.8662</td>
<td>9.06</td>
<td>2009-05-24 10:02:54</td>
</tr>
<tr>
<td>18</td>
<td>J Dennis Su</td>
<td>0.8666</td>
<td>9.02</td>
<td>2009-03-07 17:16:17</td>
</tr>
<tr>
<td>19</td>
<td>Craig Carmichael</td>
<td>0.8666</td>
<td>9.02</td>
<td>2009-07-25 16:00:54</td>
</tr>
<tr>
<td>20</td>
<td>acmehill</td>
<td>0.8668</td>
<td>9.00</td>
<td>2009-03-21 16:20:50</td>
</tr>
</tbody>
</table>

**Progress Prize 2007 - RMSE = 0.8723 - Winning Team: KorBell**

**Cinematch score - RMSE = 0.9525**
Netflix Privacy Woes

3/12/2010 @ 12:35PM | 2,590 views

Netflix Settles Privacy Lawsuit, Cancels Prize Sequel

Taylor Buley, Contributor

+ Comment Now + Follow Comments

On Friday, Netflix announced on its corporate blog that it has settled a lawsuit related to its Netflix Prize, a $1 million contest that challenged machine learning experts to use Netflix’s data to produce better recommendations than the movie giant could serve up themselves.

The lawsuit called attention to academic research that suggests that Netflix indirectly exposed the movie preferences of its users by publishing anonymized customer data. In the suit, plaintiff Paul Navarro and others sought an injunction preventing Netflix from going through the so-called “Netflix Prize II,” a follow-up challenge that Netflix promised would offer up even more personal data such as genders and zipcodes.
Outline

- Recap: *Differential Privacy* and define *Approximate Differential Privacy*
- Prediction Algorithms
- Privacy Preserving Prediction Algorithms
- Remaining Issues
Privacy in Recommender Systems

- Netflix might base its recommendation to me on both:
  - My own rating history
  - The rating history of other users

- Goal: not leak other users’ ratings to me

- Basic recommendation systems leak other users’ information
Recall Differential Privacy [Dwork et al 2006]

Randomized sanitization function $\kappa$ has $\varepsilon$-differential privacy if for all data sets $D_1$ and $D_2$ differing by at most one element and all subsets $S$ of the range of $\kappa$,

$$\Pr[\kappa(D_1) \in S] \leq e^\varepsilon \Pr[\kappa(D_2) \in S]$$
Review: Laplacian Mechanism

\[ K(D) = f(D) + \text{Lan}(GSf/\epsilon) \]

Thm: \( K \) is \( \delta \)-differentially private

Picture Proof: Probability Density Function

Question: The Gaussian (Normal) distribution is nicer because it is more tightly concentrated around its mean. Can we use that distribution instead?
Gaussian Mechanism

\[ \kappa(D) = f(D) + N(\text{GSf}/\epsilon) \]

Thm? \( \kappa \) is \( \epsilon \)-differentially private?

Probability Density Function

\[ N(x,0,\sigma) \propto 1/2\sigma\sqrt{2\pi} \exp(-x^2/2\sigma^2) \]

Problem: The ratio can be huge at the tails!

But these events are very unlikely…
Approximate Differential Privacy

Randomized sanitization function $\kappa$ has $(\varepsilon, \delta)$-differential privacy if for all data sets $D_1$ and $D_2$ differing by at most one element and all subsets $S$ of the range of $\kappa$,

$$\Pr[\kappa(D_1) \in S] \leq e^{\varepsilon} \Pr[\kappa(D_2) \in S] + \delta$$
Gaussian Mechanism

\[ K(D) = f(D) + N(\sigma^2) \]

**Thm** \( K \) is \( (\varepsilon, \delta) \)-differentially private as long as \( \sigma \geq \sqrt{2 \ln(2/\delta)} / \varepsilon \times GSf \)

**Idea** Use \( \delta \) to exclude the tails of the gaussian distribution
Multivariate Gaussian Mechanism

Suppose that $f$ outputs a length $d$ vector instead of a number

$$K(D) = f(D) + N(\sigma^2)d$$

**Thm** $K$ is $(\varepsilon, \delta)$-differentially private as long as

$$\sigma \geq \sqrt{2\ln(2/\delta)} / \varepsilon \times \max_{D1 \approx D2} \|f(D1) - f(D2)\|_2$$

Remark: Similar results would hold with the Laplacian Mechanism, but we would need to add more noise (proportional to the larger $L_1$ norm)
Approximate Differential Privacy

- Key Difference
  - Approximate Differential Privacy does NOT require that:

\[
\text{Range}(\kappa(D1)) = \text{Range}(\kappa(D2))
\]

- The privacy guarantees made by \((\varepsilon, \delta)\)-differential privacy are not as strong as \(\varepsilon\)-differential privacy, but less noise is required to achieve \((\varepsilon, \delta)\)-differential privacy.
Achieving Approximate Differential Privacy

Key Differences:

• Use of the L2 norm instead of L1 norm to define the sensitivity of $\Delta f$

$$\max_{A \approx B} \| f(A) - f(B) \|_2$$

• Use of Gaussian Noise instead of Laplace Noise
Differential Privacy for Netflix Queries

- What level of granularity to consider? What does it mean for databases D1 and D2 to differ on at most one element?
  - One user (column) is present in D1 but not in D2
  - One rating (cell) is present in D1 but not in D2

- Issue 1: Given a query “how would user i rate movie j?” Consider: $K(D-u[i])$ - how can it possibly be accurate?

- Issue 2: If the definition of differing in at most one element is taken over cells, then what privacy guarantees are made for a user with many data points?
Netflix Predictions – High Level

- **Q(i,j)** – “How would user i rate movie j?”

- Predicted rating may typically depend on
  - Global average rating over all movies and all users
  - Average movie rating of user i
  - Average rating of movie j
  - Ratings user i gave to *similar* movies
  - Ratings *similar users* gave to movie j

- Sensitivity may be small for many of these queries
Personal Rating Scale

- For Alice a rating of 3 might mean the movie was really terrible.
- For Bob the same rating might mean that the movie was excellent.
- How do we tell the difference?

\[ r_{im} - \bar{r}_i > 0 \]
How do we tell if two users are similar?

Pearson’s Correlation is one metric for similarity of users i and j
- Consider all movies rated by both users
- Negative value whenever i likes a movie that j dislikes
- Positive value whenever i and j agree

\[
S(i, j) = \sum_{m \in L_i \cap L_j} (r_{im} - \bar{r}_i)(r_{jm} - \bar{r}_j)
\]

We can use similar metrics to measure the similarity between two movies.
Netflix Predictions Example

- **Collaborative Filtering**
  - Find the k-nearest neighbors of user i who have rated movie j by Pearson’s Correlation:

    \[
    S(i, j) \quad \text{similarity of users i and j}
    \]

    \[
    N_i(k, j) = \{u_1, \ldots, u_k\} \quad \text{k most similar users}
    \]

- **Predicted Rating**

    \[
    p_{ij} = \bar{r}_i + \frac{1}{k} \sum_{u \in N_i(k, j)} (r_{uj} - \bar{r}_u)
    \]
Netflix Prediction Sensitivity Example

\[ p_{ij} = \bar{r}_i + \frac{1}{k} \sum_{u \in N_i(k, j)} (\bar{r}_{uj} - \bar{r}_u) \]

- Pretend the query Q(i,j) included user i’s rating history
- At most one of the neighbors ratings changes, and the range of ratings is 4 (since ratings are between 1 & 5). The L1 sensitivity of the prediction is:

\[ \Delta p = 4/k \]
Similarity of Two Movies

- Let $U$ be the set of all users who have rated both movies $i$ and $j$ then

\[ S(i, j) = \sum_{u \in U} (r_{uj} - \bar{r}_u) \times (r_{ui} - \bar{r}_u) \]
K-Nearest Users or K-Nearest Movies?

Find $k$ most similar users to $i$ that have also rated movie $j$?

Find $k$ most similar movies to $j$ that user $i$ has rated?

Either way, after some pre-computation, we need to be able to find the $k$-nearest users/movies quickly!
Covariance Matrix

Movie-Movie Covariance Matrix
- \((M \times M)\) matrix
- \(\text{Cov}[i][j]\) measures similarity between movies \(i\) and \(j\)
- \(M \approx 17,000\)
- More accurate

User-User Covariance Matrix?
- \((N \times N)\) Matrix to measure similarity between users
- \(N \approx 500,000\)
- Less accurate
What do we need to make predictions?

For a large class of prediction algorithms it suffices to have:

- $G_{avg}$ – average rating for all movies by all users
- $M_{avg}$ – average rating for each movie by all users
- Average Movie Rating for each user
- Movie-Movie Covariance Matrix (COV)
Differentially Private Recommender Systems (High Level)

To respect approximate differential privacy publish

- $G_{avg} + \text{NOISE}$
- $M_{avg} + \text{NOISE}$
- $\text{COV} + \text{NOISE}$

- $\Delta G_{avg}, \Delta M_{avg}$ are very small so they can be published with little noise
- $\Delta \text{COV}$ requires more care (our focus)

Don’t publish average ratings for users (used in per-user prediction phase using k-NN or other algorithms)

Source: Differentially Private Recommender Systems (McSherry and Mironov)
Movie-Movie Covariance Matrix

\[ \text{Cov} = \sum_{u} \sum (r_{iu} - \bar{r}_{u})(r_{iu} - \bar{r}) \]

\[ r_{iu} = r_{iu} - \bar{r} \]

- User u’s rating for each movie
- Average rating for each movie
Movie-Movie Covariance Matrix

\[ \text{Cov} = \sum_u (r_u)(r_u)^T \]

\[ r = \langle 3.2 @ 2 @ 3 \rangle \]

\[ r_{u1} = \langle 4.2 @ 2 @ 3 \rangle \quad r_{u2} = \langle 1.5 @ 4.5 @ 2 \rangle \]
Movie-Movie Covariance Matrix

\[ \text{Cov} = \sum_{u} \left( r_{\downarrow u} \right) \left( r_{\downarrow u} \right)^T \]

\[ r = \langle 3.2, 2, 3 \rangle \]

\[ r_{\downarrow u} = \langle 1, 0, 0.5, -1 \rangle \]
Example

\[ r\downarrow u_1 \ (r\downarrow u_1 )^T = \begin{pmatrix} -1.7 & 2.5 & -1 \\ -1.7 & 2.5 & -1 \end{pmatrix} \]

\[ -4.25 = -1.7 \times 2.5 \]

\[ 2.89 \times -4.25 \ ]
Example

\[ \text{Cov} = r_{u1} (r_{u1})^T + r_{u2} (r_{u2})^T \]

\[ = \begin{bmatrix}
3.89 & -4.25 & 1.7 \\
-4.25 & 6.25 & -2.5 \\
1.7 & -2.5 & 1
\end{bmatrix} \]
Covariance Matrix Sensitivity

\[
\text{Cov} = \sum_u r_u r_u^T
\]

\[
\|\text{Cov}^a - \text{Cov}^b\| = \|r_u^a r_u^a T - r_u^b r_u^b T\|
\leq \|r_u^a - r_u^b\| \times (\|r_u^a\| + \|r_u^b\|)
\]

- Could be large if a user’s rating has large spread or if a user has rated many movies
Covariance Matrix Trick I

- Center and clamp all ratings around averages. If we use clamped ratings then we reduce the sensitivity of our function.

\[
\tilde{r}_{ui} = \begin{cases} 
-B, & \text{if } r_{ui} - \bar{r}_u < -B, \\
 r_{ui} - \bar{r}_u, & \text{if } -B \leq r_{ui} - \bar{r}_u < B, \\
 B, & \text{if } B \leq r_{ui} - \bar{r}_u.
\end{cases}
\]
Example (B = 1)

User 1:

\[ r_{lu1} = \langle 4.2, 2, 3 \rangle \]
\[ r_{lu1} = 4.2 + 2 + 3/3 \approx 3.07 \]

\[ r_{lu1} = \langle 1, -1, -0.07 \rangle \]

\[ \min r \{B, 4.2 - 3.07\} \]
\[ \max r \{-B, 2 - 3.07\} \]
Covariance Matrix Trick II

- Carefully weight the contribution of each user to reduce the sensitivity of the function. Users who have rated more movies are assigned lower weight.

\[
\text{Cov} = \sum_u w_u \hat{r}_u \hat{r}_u^T + \text{Noise}^{d \times d}
\]

- Where \( e_{ui} \) is 1 if user \( u \) rated movie \( i \) and

\[
w_u = 1/\|e_u\|_2
\]
Publishing the Covariance Matrix

- Theorem (roughly):

\[ \| w_u^a r_u^a r_u^a T - w_u^b r_u^b r_u^b T \|_2 \leq (1 + 2\sqrt{2}) B^2 \]

- Add independent Gaussian noise proportional to this sensitivity bound to each entry in covariance matrix
Experimental Results

Privacy decreases

Source: Differentially Private Recommender Systems (McSherry and Mironov)
Note About Results

- **Granularity:** One *rating* present in D1 but not in D2
  - Accuracy is much lower when one user is present in D1 but not in D2
  - Intuition: Given query Q(i,j) the database D-u[i] gives us no history about user i.

- **Approximate Differential Privacy**
  - Gaussian Noise added according to L2 Sensitivity
  - Clamped Ratings (B =1) to further reduce noise
Acknowledgment

- A number of slides are from Jeremiah Blocki
Global Averages

\[
\begin{align*}
\text{GSum} &= \sum_{u,i} r_{ui} + \text{Noise}, \\
\text{GCnt} &= \sum_{u,i} e_{ui} + \text{Noise}, \\
\text{MSum} &= \sum_u r_u + \text{Noise}^d, \\
\text{MCnt} &= \sum_u e_u + \text{Noise}^d. \\
\end{align*}
\]

\[
G = \frac{\text{GSum}}{\text{GCnt}},
\]

\[
\text{MAvg}_i = \frac{\text{MSum}_i + \beta_m G}{\text{MCnt}_i + \beta_m}.
\]
**Theorem 4.** Let $r^a$ and $r^b$ differ on one rating, present in $r^b$. Let $\alpha$ be the maximum possible difference in ratings$^2$. For centered and clamped ratings $\tilde{r}^a$ and $\tilde{r}^b$, we have

\[
\|\tilde{r}^a - \tilde{r}^b\|_1 \leq \alpha + B,
\]

\[
\|\tilde{r}^a - \tilde{r}^b\|_2 \leq \frac{\alpha^2}{4\beta_p} + B^2.
\]

$^2$For the Netflix Prize data set $\alpha = 4.$