Attack Detection in Collaborative Filtering Recommender System

By:
Deepti Goel
Anurag Tripathi
Anvaya Rai
Four Pillars...
Recommender System...

“Automate the Circle of Advisors”

<table>
<thead>
<tr>
<th></th>
<th>Item 1</th>
<th>Item 2</th>
<th>Item 3</th>
<th>Item 4</th>
<th>Item 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>User 1</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>User 2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>User 3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>User 4</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>User 5</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>User 6</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

User 1 had already bought Item 1, 2 and 3.
Which Item should User 1 buy next, Item 4 or 5?

Average Rating of Item 4 is \( \frac{10}{4} = 2.5 \) and of Item 5 is \( \frac{14}{4} = 3.5 \)

So, he should go for Item 5 !!!
Recommender System ...

How correct was the previous recommendation?

There could be many more parameters affecting the choice of the USER 1.

Above recommendation would be same for USER 1 and USER 6.

But there choices and needs may be different based on various other things like geographical location, culture, habits, liking etc.

Taking these factors into account is the job of Collaborative Filtering during a recommendation!!!
Recommender System ...

The basic idea...

• Get the set of most similar users from the database

• Calculate the rating of the item by taking a weighted average, by giving more weights to the ratings of the similar users

• Make the recommendation
Collaborative Filtering...

Given User Parameters ($\Theta$), Estimate Item Property Vectors ($x$) such that:

\[
\min_{x^{(i)}} \frac{1}{2} \sum_{j:r(i,j)=1} \left( (\theta(j)^T x^{(i)}) - y^{(i,j)} \right)^2 + \frac{\lambda}{2} \sum_{k=1}^{n} (x_k^{(i)})^2
\]

Where,
\[r(i,j) = 1\text{ if user } j \text{ has rated movie } i\]
\[y^{(i,j)} = \text{ rating given by user } j \text{ to movie } i\]
The Threat... The Fake Profiles

True Database

Fake Users

Distorted Database
True User and Attacker Profile...

- True User’s Profile
  - Unrated Items
  - Rated Items

- Attacker’s Profile
  - Selected Items
  - Filler Items
  - Unrated Items
  - Target Items
The Attack Models...

Random Attack
Average Attack
Bandwagon Attack
Segment Attack
Love-Hate Attack
Reverse Bandwagon Attack

Attack Profiles created by these models effective in promoting an item, but they are highly correlated and hence can be detected by the Recommender System easily
The Correlation Matrix...

True Profiles have huge Variance but low Covariance and in case of Fake Profiles it is vice versa
The Fake Profile Detection...

True Profiles in Green and Fake Profiles in Red
Detection done using PCA
The Hybrid Attack Model...

Attacker’s Profile

Selected Items

Filler1

Filler2

Unrated Items

Target Items

Filler Items

Filler Items split into two sets: Filler1 and Filler2

Filler1 filled using Gaussian with mean and variance of first ‘F1’ items in the list

Filler2 filled using Gaussian with mean and variance of first ‘F1’ items in the list

Items selected randomly and ratings assigned through Gaussian
The New Correlation Matrix...

A reduced amount of correlation is visible among the Fake Profiles
The PCA Analysis...

Poor classification with previous threshold
The PCA Analysis...

Poor classification with Updated threshold as well
The Results...

Reduced correlation between the Fake Profiles

Bandwagon Attack Model

Hybrid Attack Model

Reduced correlation between the Fake Profiles
The Results...

<table>
<thead>
<tr>
<th>Case</th>
<th>PCA</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bandwagon</td>
<td>Hybrid</td>
</tr>
<tr>
<td>True Profiles classified as Fake Profiles</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>Fake Profiles classified as True Profiles</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>True Profiles classified as Fake Profiles (Threshold Updated)</td>
<td>NA</td>
<td>5</td>
</tr>
<tr>
<td>Fake Profiles classified as True Profiles (Threshold Updated)</td>
<td>NA</td>
<td>9</td>
</tr>
</tbody>
</table>

Increased Misclassification i.e. more True Profiles classified as Fake Profiles
The Resources...

Dataset:
MovieLens 100K data set (www.cs.umn.edu/research/GroupLens/data).

OS:
Windows and Linux

Platforms:
MATLAB and Python

Packages:
libSVM, scikit-learn and python-matplotlib

2. Recommender System Handbook, 2010


4. Unsupervised shilling detection for collaborative filtering http://dl.acm.org/citation.cfm?id=1619870

5. Attacks and Remedies in Collaborative Recommendation
Thank You!

Time for You to Attack...
The Questions?