A Novel Shilling Attack Detection Method

ZEYNEP OZDEMIR

ANADOLU UNıVERSıTY

COMPUTER ENGıNEERıNG DEPARTMENT

Recommender Systems

- An impressive way of overcoming information overload problem
- Choose the most liked items among a huge number of possible items
- Save time
- Help to match users with right items
- Two way to provide recommendations: ***Collaborative Filtering, Content- based approaches***

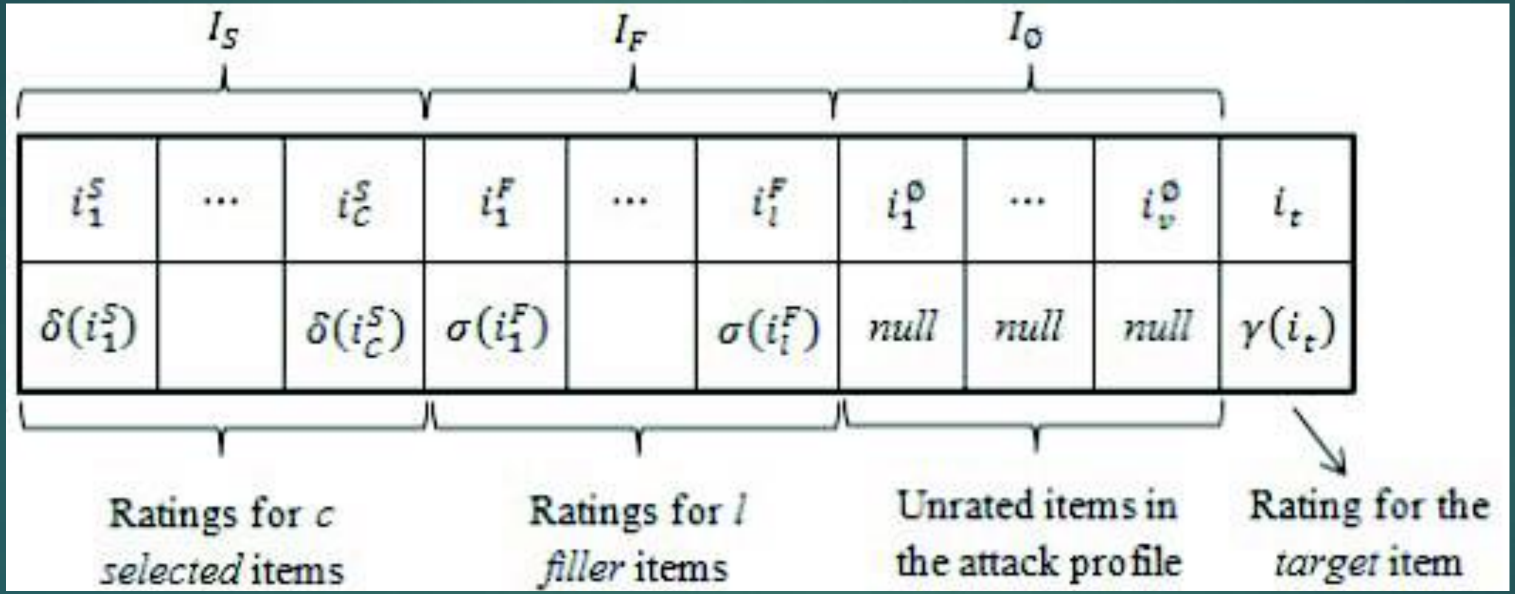
Collaborative Filtering Recommender Systems

- One of the recommendation techniques
- Produce highly accurate predictions
- Based on the assumption
  - Users having similar experiences on past items are tend to agree on new items.
- They are vulnerable to profile injection attacks/shilling attacks.

Shilling Attacks

- Increase/Decrease the popularity of target item.
- Construct fake profiles. Insert them into system's database.
- Effective impact on produced predictions
- ***Filler size*** and ***attack size*** used to design the attacks
- Categorized as push and nuke attacks according to their intends.

General form of an attack profile



Employed shilling attacks

- Shilling attacks we focused on:
  - Segment attack:
    - Designed for a group of users, Low-knowledge, Push attack
  - Bandwagon attack
    - Low-knowledge, Push attack, Popular items are chosen as selected items
  - Average attack
    - Filler items are chosen as randomly,

Importance of detection

- Bogus profiles make data quality worse and affect the accuracy of the predictions.
  - Detection of bogus profiles is extremely important for reliability of the system.
    - A novel shilling attack detection method for specific attacks based on bisecting k-means clustering approach.

A novel shilling attack detection method-Methodology

- Construct a binary decision tree via bisecting k-means clustering algorithm
- Find intra-cluster correlation for each node
- Utilize intra-cluster correlation to detect bogus profiles.

Constructing BDT via bisecting k-means clustering algorithm

- The central server produces a BDT off-line
- K-means clustering is applied to group users into two distinct clusters at each level recursively.
- If any leaf node exceeds the neighbor number(N), the corresponding node is bisected.
- At most N user in each leaf node.

Detection of bogus profiles

- A novel approach: intra-cluster correlation as detection attribute
  - Calculate the intra-cluster correlation coefficient of each sub-cluster for an internal node.
  - Shilling attacks profiles resemble high intra-cluster correlation because of their certain generation strategy.
  - Traverse the BDT to find the shilled cluster.
    - Direct toward higher intra-cluster correlation.
    - Intra cluster correlation of two children nodes ,that consist of totaly or most of fake profiles, can not be diversely different intra-cluster correlation of parent node.

A novel shilling attack detection method-Experiments

- MovieLens Data
- ***Precision*** and ***Recall*** as evaluation metric
- Experiments according to varying ρ parameter, attack size and filler size values.

Table 1. Effects of varying ρ values on overall performance

	Precision					Recall				
ρ	1	2	4	7	10	1	2	4	7	10
Segment	0.955	0.933	0.875	0.850	0.863	0.950	0.955	0.965	0.952	0.967
Bandwagon	0.574	0.577	0.521	0.469	0.396	0.371	0.572	0.815	0.942	0.988
Average	0.746	0.743	0.749	0.751	0.701	0.622	0.619	0.623	0.628	0.638

A novel shilling attack detection method-Experiments

Table 2. Effects of varying *fillersize* values on overall performance

	Precision					Recall				
Fillersize	3	5	10	15	25	3	5	10	15	25
Segment	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.984
Bandwagon	0.904	0.897	0.929	0.984	0.985	0.922	0.914	0.945	1.000	0.999
Average	0.521	0.916	0.992	0.990	0.988	0.498	0.800	0.826	0.877	0.949

Table 3. Effects of varying *attacksize* values on overall performance

	Precision					Recall				
Fillersize	3	5	10	15	25	3	5	10	15	25
Segment	0.622	0.980	0.854	1.000	1.000	0.620	0.984	0.853	1.000	0.982
Bandwagon	0.053	0.085	0.070	0.947	0.985	0.970	0.973	0.352	0.987	0.999
Average	0.161	0.898	0.964	0.982	0.988	0.127	0.765	0.873	0.916	0.951

Summary & Future Work

- Our work is the first one that uses bisecting k-means clustering as detection scheme.
- Very successful at detecting bogus profiles generated from specific attack models like segment, bandwagon and average attacks.
- We want to extend our work to detect shilling attacks in private environments