BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY

Using Ensemble Method to Detect Attacks in the Recommender System

IEEE Access, Volume: 11, October 2023
Link: https://ieeexplore.ieee.org/abstract/document/10268947

Jaid Monwar Chowdhury
1905083, CSE'19

Ahmad Farhan Shahriar Chowdhury
1905081, CSE'19

Problem Statement

Recommender Systems (RSs):

- Widely used in e-commerce, entertainment, and content platforms.
- Collaborative filtering (CF) is a popular approach but vulnerable to shilling attacks.

Shilling Attacks:

- Malicious users (fake profiles) manipulate system output by inflating or deflating item ratings.

Problem:

- Traditional detection methods have limitations in accuracy, especially against sophisticated attacks (obfuscated).

Dataset

Yahoo Movie Dataset

- Link: https://webscope.sandbox.yahoo.com/catalog.php?datatype=r
- Statistics: Dataset consisting of 7642 unique users and 11916 unique movies

MovieLens 1M Dataset

- Link: https://grouplens.org/datasets/movielens/
- Statistics: MovieLens 1M movie ratings. Stable benchmark dataset. 1 million ratings from 6000 users on 4000 movies. Released 2/2003.

Overall Architecture

SYNTHETIC DATA GENERATION

Use Adversarial Network to generate synthetic profiles

Models Used: TGAN, CTGAN

Train TGAN, CTGAN models with real data and generate synthetic data from the trained model

CHOOSE ATTACK MODEL

Attack Models:

1. **Random Attack**
- Rate 5 to a target item; give filler items random ratings conforming to a Gaussian Distribution with mean 3.6 and standard deviation 1.1
2. **Average Attack**
- Rate 5 to a target item; the ratings for filler items are distributed around the mean for each item

DETECT FAKE PROFILES

Detection Method:

1. **Base Classifiers**
- RDAM, WDA, cosine similarity, LenVar, TF-IDF, RDMA similarity, RDMA_LenVar, and degree of similarity
2. **Voting Classifier**
- Use voting to choose the final classification

TEST RECOMMENDATION ACCURACY

Test Recommendation system accuracy and other evaluation metrics after recalling fake profiles

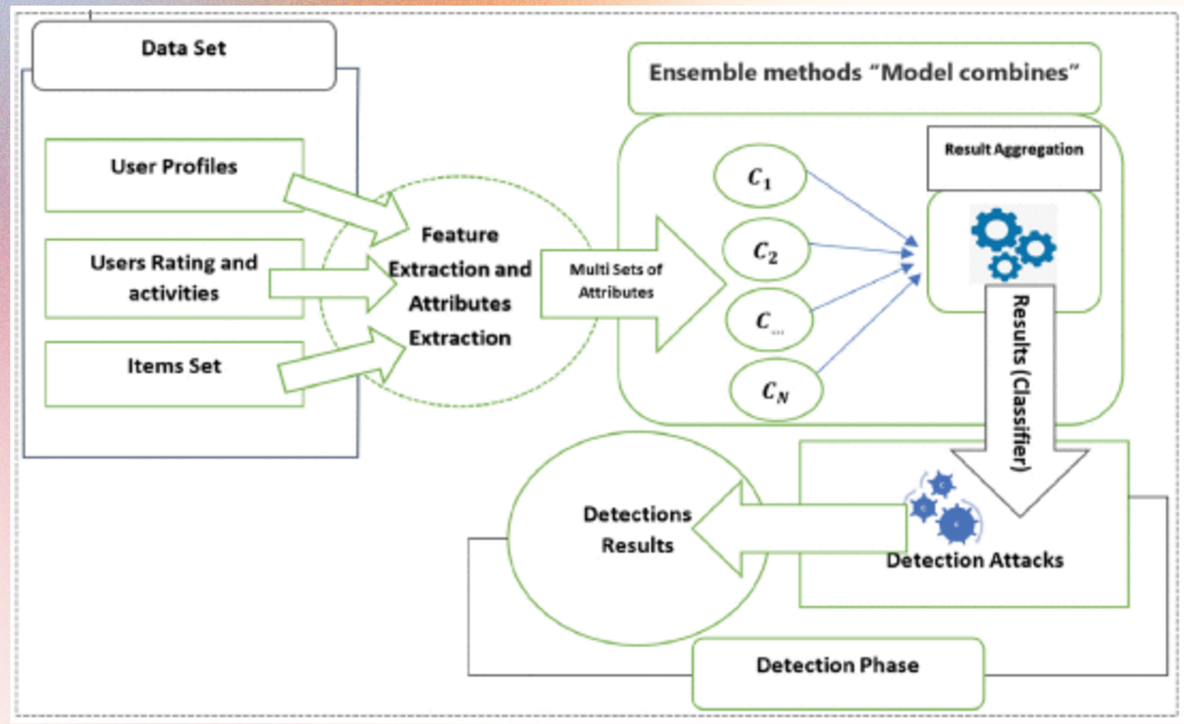
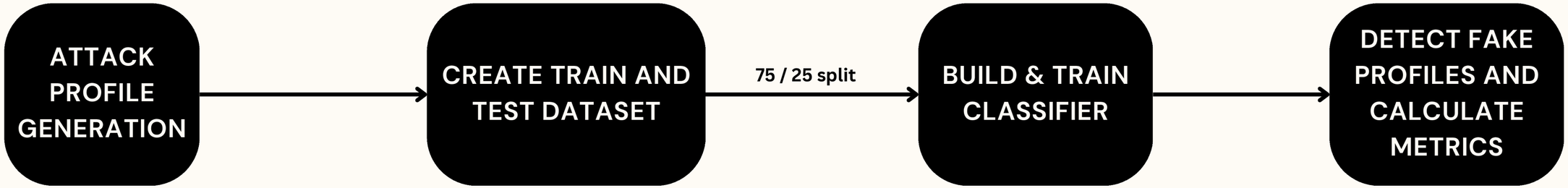


Fig: Architecture Diagram

Overall Architecture



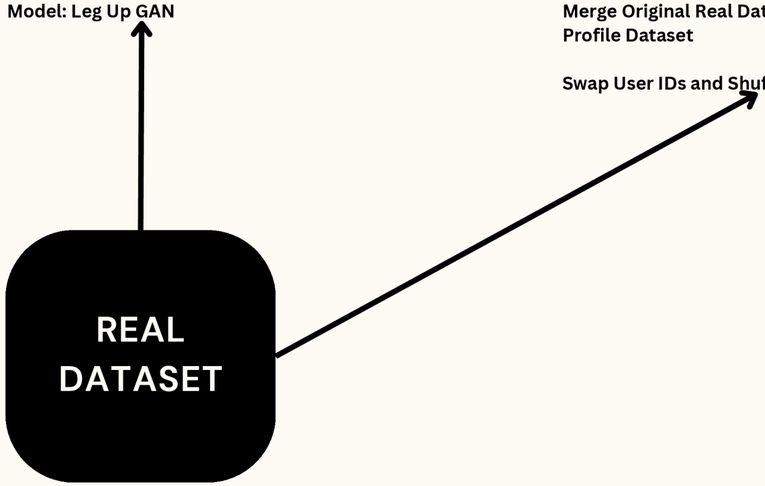
Model: Leg Up GAN

Merge Original Real Dataset with Attack Profile Dataset

Swap User IDs and Shuffle

Detection Method:

1. **Base Classifiers**
- RDAM, WDA, cosine similarity, LenVar, TF-IDF, RDMA similarity, RDMA_LenVar, and degree of similarity
2. **Stacking**
- Use stacking to choose the final classification



Evaluation Metrics

Detection Rate

$$= \frac{\text{No. of Detection}}{\text{No. of Attacks}}$$

False Positive Rate

$$= \frac{\text{No.of FalsePositives}}{\text{No. of GenuineProfiles}}$$

Precision

$$= \frac{\text{True Positive}}{\text{TruePositive + FalsePositive}}$$

Recall

$$= \frac{\text{True Positive}}{\text{TruePositive + FalseNegative}}$$

F1-Measure

$$= \frac{\text{2*Precision*Recall}}{\text{Precision + Recall}}$$

Feedback

The presentation was accepted by the teachers.

Everything was okay.

No additional feedback received from the teachers.

Results: Movie Lens 100K

	Precision	Recall	F1-Score	Support
0	0.93	1.00	0.96	25
1	1.00	0.99	1.00	236
Accuracy			0.99	261
Macro Average	0.96	1.00	0.98	261
Weighted Average	0.99	0.99	0.99	261

Results: Tool Home

	Precision	Recall	F1-Score	Support
0	1.00	0.97	0.99	38
1	1.00	1.00	1.00	223
Accuracy			1.00	261
Macro Average	1.00	0.99	0.99	261
Weighted Average	1.00	1.00	1.00	261

Results: Grocery Food

	Precision	Recall	F1-Score	Support
0	1.00	0.98	0.99	63
1	1.00	1.00	1.00	444
Accuracy			1.00	261
Macro Average	1.00	0.99	0.99	261
Weighted Average	1.00	1.00	1.00	261