Ph.D. Defense

Dissertation Title

Adversarial Machine Learning in Recommender Systems

Felice Antonio Merra

felice.merra@poliba.it

Supervisor: Prof. Tommaso Di Noia

Coordinator: Prof. Mario Carpentieri

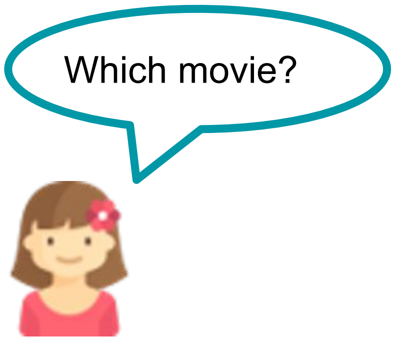
XXXIV Cycle

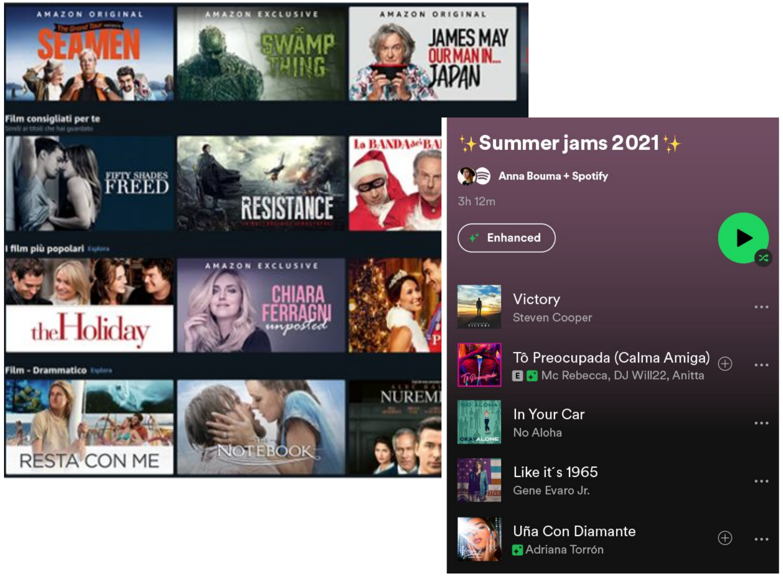
Summary

- Recommender Systems (RS) and Adversarial Machine Learning (AML) Background
- Research Contributions
  - Interpretation of the Impact of Data Characteristics on Robustness
  - Semantics-aware Shilling Attacks
  - Attacks/Defense against Visual RSs
  - Iterative Methods to Perturb Model Parameters.
  - Formal Analysis of Recommendation Quality of Adversarial Recommenders
- Conclusions

Recommender Systems: Goal

Support users' decision-making process in the huge catalogs of e-commerce platforms (e.g., Netflix, Spotify, Google, and Amazon).





Recommender Systems: Techniques

Use of Machine Learning (**ML**) to extrapolate:

- Behavioral Patterns across Users→ **Collaborative Filtering** (CF)
  - Model-based
  - Memory-based
  - Graph-based
- Similarities across Items → **Content-based Filtering** (CBF)
  - Metadata: title, brand name, author
  - Multimedia: product images, sound tracks, videos
  - Semantic Data: knowledge graphs
- CF + CBF → **Hybrid**

Recommender Systems: Assumptions

Collaborative Filtering

↓

Users behave **Honestly**

↓

Good Recommendation
thanks to the
Wisdom of the Crowd

Content-based Filtering

↓

Items Content is **Original**

↓

Good Recommendation
thanks to the
Quality of the Content

“Applications of machine learning are adversarial in nature”

Adversarial Machine Learning
Yevgeniy Vorobeychik and Murat Kantarcioglu
Synthesis Lectures on Artificial Intelligence and Machine Learning, August 2018

Recommender Systems: **Adversarial** Assumptions

Collaborative Filtering

↓

Users behave **Maliciously**

↓

Bad Recommendation because of the **Wickedness of the Crowd**

Content-based Filtering

↓

Items Content is **Adversarial**

↓

Bad Recommendation because of **Manipulated Content**

Recommender Systems: Security Issues

Hand-Engineered

- Injection of Fake Users (Shilling Attacks)
  - Leveraging interaction data

Studied in RSs from the early 2000s…

Robust Collaborative Recommendation
Robin Burke, Michael P. O'Mahony, Neil J. Hurley
Recommender Systems Handbook 2015

of ML Fake Users
of Altered Content Data
of Adversarial Perturbations

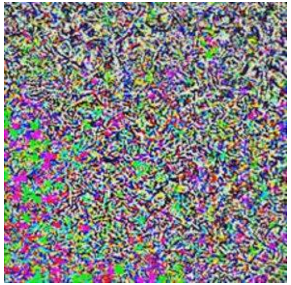
Use of (**Adversarial**) **Machine Learning** Techniques to Attack/Protect Recommender Models

Adversarial Machine Learning

Study of security breaches of ML models in several tasks with a particular (initial) focus on **computer vision (CV)** classification task.

Panda

+

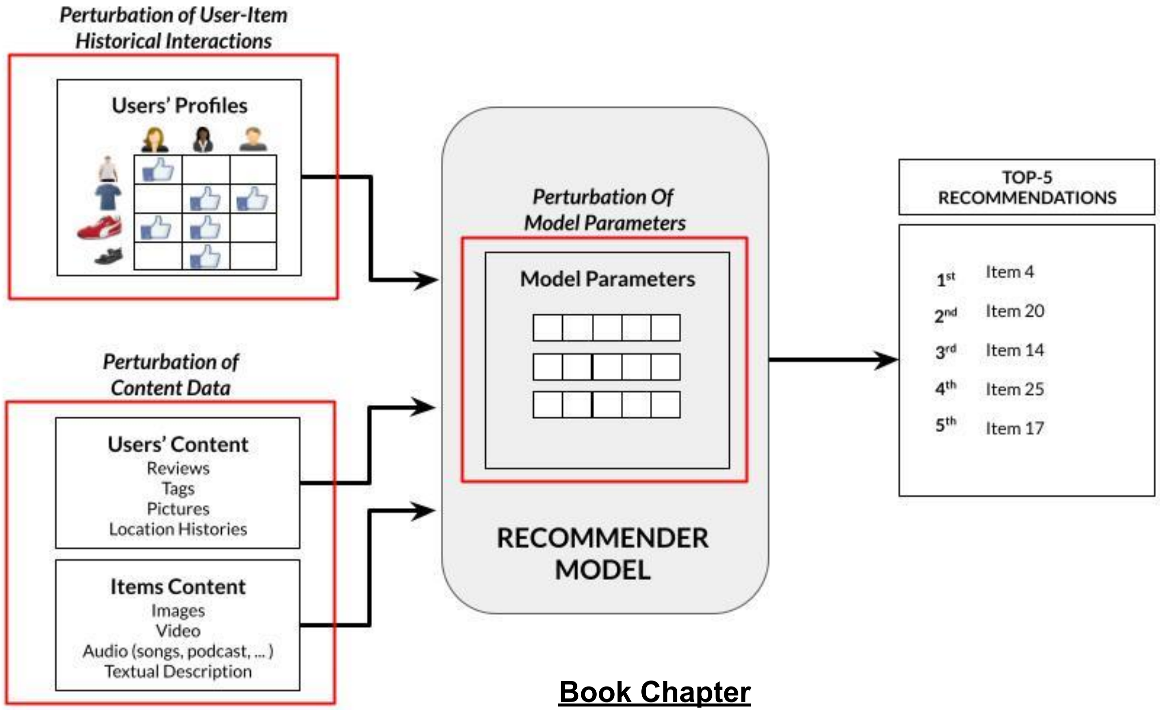


Adversarial Perturbation

=

Gibbon

Adversarial Machine Learning in RSs



<u>**Book Chapter**</u>

Adversarial Recommender Systems: Attack, Defense, and Advances
Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, ***Felice Antonio Merra***
The 3rd Edition of the Recommender Systems Handbook. 2022

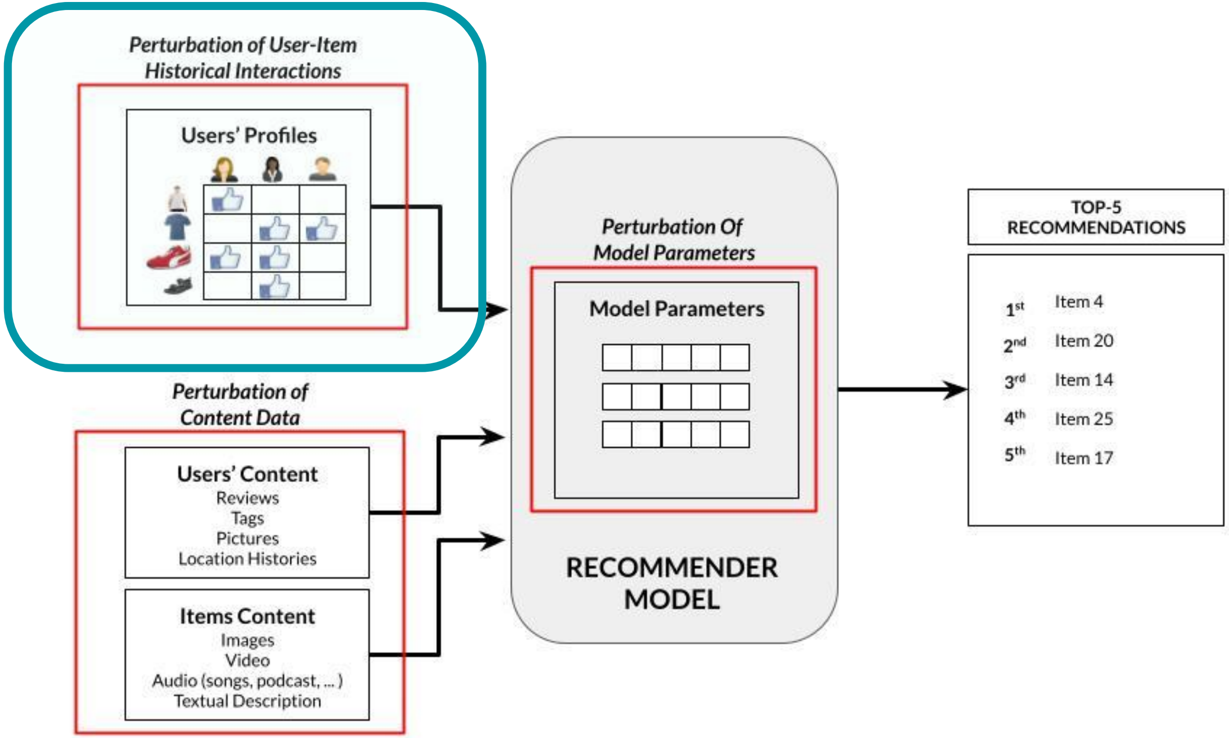
What are the novel adversarial risks of ML-RS?

How can we robustify and defend them to preserve high quality recommendations?

Overall Research Questions

Research Contributions

Focus of the following contributions



Impact of Data Characteristics on the Recommendation Robustness

Motivations

- Existing works on Shilling Attacks have focused on "**win-lose**" scenarios, proposing stronger attacks, as well as stronger defense.
- No attention on understanding possible source of robustness.

Research Question

Is there an underlying relationship between the dataset characteristics and the effectiveness of shilling attack against CF-RSs?

<u>**Reference Publication as Main Author**</u>

How Dataset Characteristics Affect the Robustness of Collaborative Recommendation Models.
Yashar Deldjoo, Tommaso Di Noia, Eugenio Di Sciascio, ***Felice Antonio Merra***
SIGIR 2020

Method: Regression-based Explanatory Model

Dependent Variable (DV)
Quantify Attack's Effects

Regression Coefficients
Statistical Significance
Magnitude
Directionality

Independent Variables (IV)
Data Characteristics

- Size
- Shape
- Density
- Gini-index on Users
- Gini-index on Items
- Ratings Std. Dev.

$$y_i = \epsilon_i + \theta_0 + \sum_{d=1}^{D-1} \theta_d x_{d,i} + \sum_{c=1}^{C} \theta_c x_{c,i}$$

Error

Expected Value of the DV

Regression Coefficients and IVs
For the between dataset analysis

Results

- **IVs** Account for the variations in attack performance ($R^2 > 60\%$)

- **Density** & **Space** have **Negative** Impact
  Increasing the density (or decreasing sparsity) of the dataset REDUCES the attacks' effectiveness.

- **Shape** has **Positive** Impact
  Few items are easy to be manipulated.

$\Delta_{HR@10}$		User-kNN		
		ML-20M	**Yelp**	**LFM-1b**
Random	$R^2(adj.R^2)$	0.761(0.758)	0.838(0.835)	0.673(0.668)
	$Constant$	**.179*** **	**.609*** **	**.717*** **
	$SpaceSize_{log}$	-0.063***	.041	-0.629***
	$Shape_{log}$	.184***	.248***	.288*
	$Density_{log}$	-0.189***	-0.316*	-1.546***
	$Gini_{users}$	.277	-0.012	1.901***
	$Gini_{item}$	-0.102	-0.485	1.753***
	Std_{rating}	-0.072	.287	-0.152
Love-Hate	$R^2(adj.R^2)$	0.806(0.803)	0.839(0.837)	0.673(0.668)
	$Constant$	.267***	.657***	.717***
	$SpaceSize_{log}$	-0.027*	.042	-0.628***
	$Shape_{log}$	.209***	.131*	.287*
	$Density_{log}$	-0.198***	-0.290*	-1.544***
	$Gini_{users}$	.347	.114	1.896***
	$Gini_{item}$	-0.430	-0.150	1.754***
	Std_{rating}	-0.179	.239	-0.151

TAKE HOME MESSAGE

A recommender system designer can quantify the effectiveness of attacks by using the dataset characteristics and manipulating them to guarantee higher robustness.

Semantic-Aware Shilling Attacks on RSs Exploiting Knowledge Graphs.

Motivations

- Publicly available KGs, e.g., DBpedia, have been used as a source of information to enhance recommendation accuracy and diversity.
- A lack of investigation is on verifying if these data can be used for malevolent objectives.

Research Question

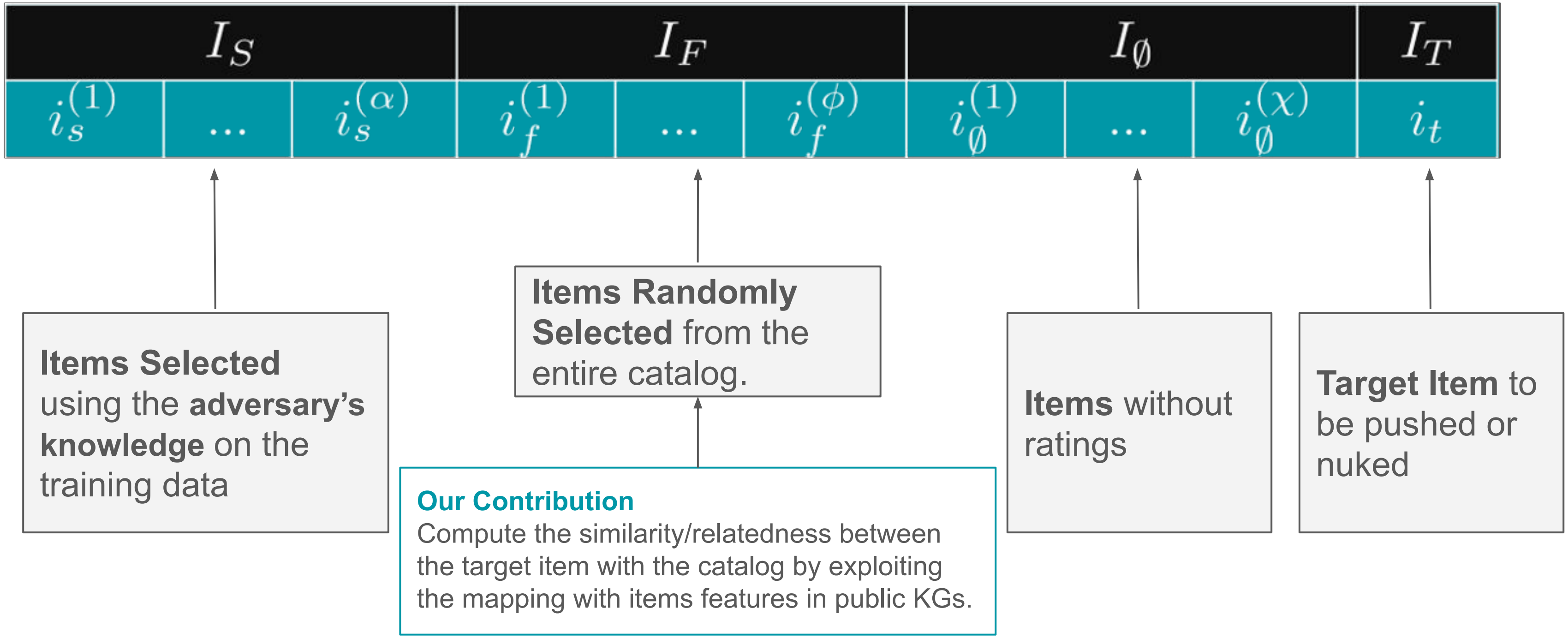
Can public available semantic information be exploited to develop more effective shilling attacks against CF models?

Reference Publication as Main Author

SAShA: Semantic-Aware Shilling Attacks on Recommender Systems Exploiting Knowledge Graphs.
*Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, Eugenio Di Sciascio, **Felice Antonio Merra***
ESWC 2020

Method: Semantic-Aware SHilling Attacks (SAShA)

The Structure of a Shilling Profile (the Fake Users)



Results

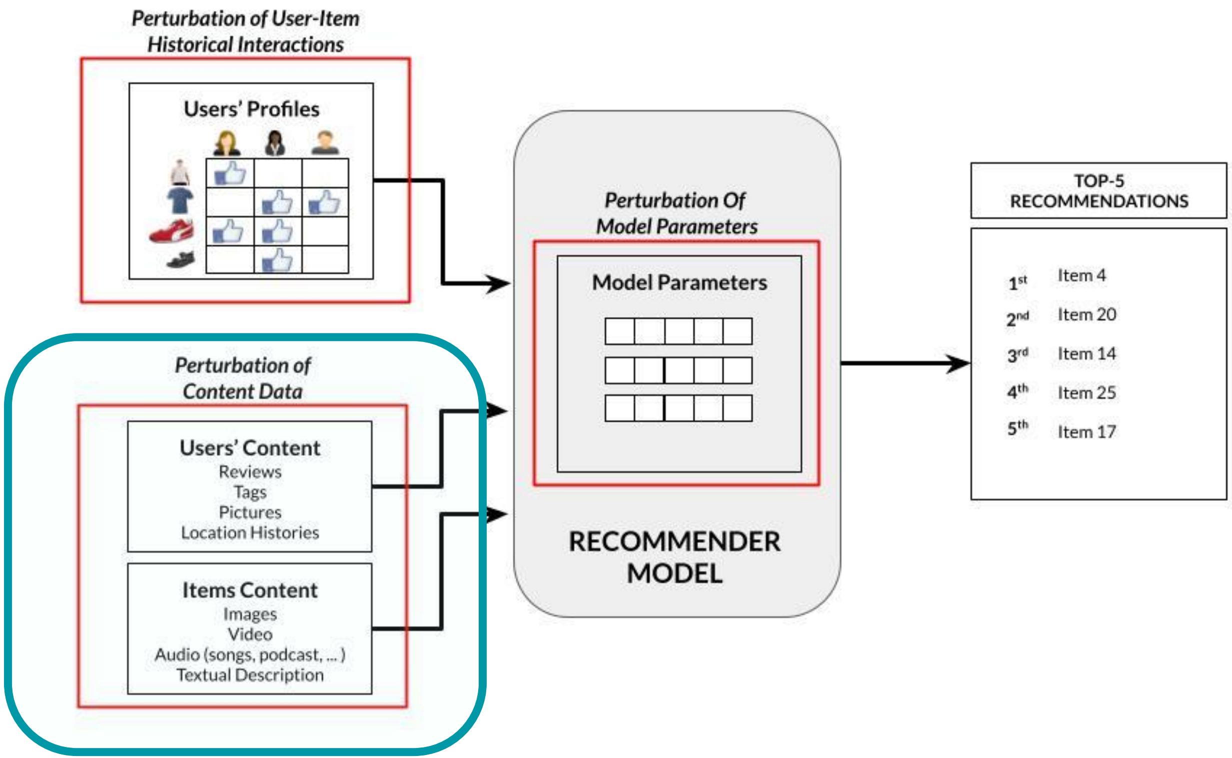
			User-kNN			Item-kNN			MF			NeuMF		
Attack	**Feature**	**Sim.**	1	2.5	5	1	2.5	5	1	2.5	5	1	2.5	5
R	Baseline		*.0736*	*.1570*	*.2301*	*.2885*	*.4588*	*.5590*	*.7660*	*.8987*	*.9419*	*.0612*	*.1130*	*.2216*
	Cat.	Cosine	.0745	.1576	.2311	.2804	.4575	.5687	.7837	.9014	**.9439**	.0802	.1324	.1653
		Katz	.0808	.1698	.2441	.2862	.4610	.5691	.7885	.9021	.9418	.0808	.1105	.1812
		Excl.	**.0816**	**.1703**	**.2456**	**.2915**	.4635	.5707	.7897	.8993	.9427	.0886	.1479	**.2417**
	Ont.	Cosine	.0709	.1503	.2252	.2748	.4483	.5634	.7720	.8979	.9423	.0561	**.1493**	.1926
		Katz	.0774	.1622	.2355	.2837	.4592	.5670	.7845	.9021	.9416	.0751	.1392	.1857
		Excl.	.0766	.1619	.2349	.2848	.4602	.5686	.7846	.9010	.9433	**.1091**	.0999	.2240
	Fact.	Cosine	.0740	.1558	.2280	.2786	.4528	.5642	.7835	.9023	.9419	.0676	.1009	.1285
		Katz	.0760	.1591	.2319	.2823	.4570	.5662	.7839	.9015	.9417	.0685	.1366	.1823
		Excl.	.0793	.1672	.2425	.2890	**.4646**	**.5722**	**.7888**	**.9029**	.9434	.0921	.1034	.2143

- **KGs** data **improve** by a large margin the attacker’s performance.
- **Graph-based** measures make **attacks stronger** and stronger capturing imperceptible similarities.
- **Single-hop** exploration is **sufficient** to outperform the SOTA techniques.
- **Similarity**-based and classical Factorization RSs heavily suffer from semantic attacks.

TAKE HOME MESSAGE

Since public KG can be also maliciously used by an adversary, novel defense solutions have to be made considering this always available source of adversary knowledge.

Focus of the following contributions



Training Time Adversarial Attacks and Defenses against Visual-based RSs

Motivations

- Visual recommenders rely on visual features extracted from product images to enhance the recommendation performance since users' taste is influenced by the aesthetic appearance of products.
- Despite AML emerged in the computer vision domain, no works have been focused on **poisoning** Visual RSs.

Research Question

*Can an adversary **poison** the data of multimedia recommender systems with adversarial samples?*

*Do adversarial perturbations of product images **confuse** multimedia recommenders?*

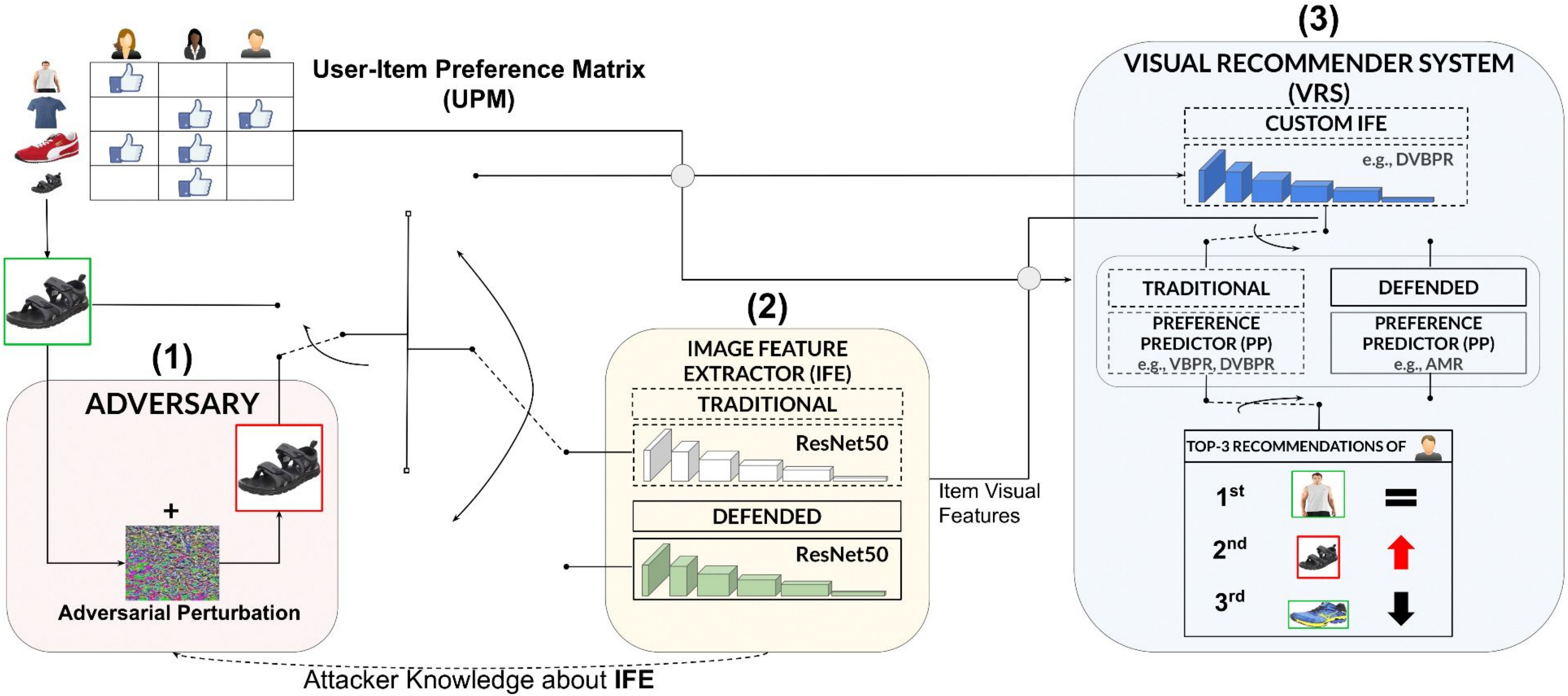
*Can we **protect** the model integrity?*

Reference Publications as Main Author

A Study of Defensive Methods to Protect Visual Recommendation Against Adversarial Manipulation of Images.
*Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, Daniele Malitesta, **Felice Antonio Merra***
SIGIR 2021

TAaMR: Targeted Adversarial Attack against Multimedia Recommender Systems.
*Tommaso Di Noia, Daniele Malitesta, **Felice Antonio Merra***
DSML 2020

Framework: Visual Adversarial Recommendation (VAR)



Results

SOTA VRSs are not robust against Training Time Tttacks.

The proposed adversarial robustification strategies have partially reduced the impact of adversarial attacks.

TAKE HOME MESSAGE
The adaptation of sota adversarial robustification procedures are not the final solution and novel studies need to be performed in the next years.

Data	VRS	Att.	Image Feature Extractor					
			Traditional		Adv. Train.		Free Adv. Train.	
			CHR	*CnDCG*	*CHR*	*CnDCG*	*CHR*	*CnDCG*
Amazon Men	FM	*Base*	*0.4960*	*0.0246*	*0.4082*	*0.0204*	*0.4048*	*0.0202*
		FGSM	**0.5309** *	**0.0266***	0.3886	0.0198*	0.3821*	0.0194*
		PGD	0.5293*	**0.0266***	0.3795*	0.0193*	0.3811*	0.0193*
		C&W	0.5258*	0.0263*	0.3837*	0.0194*	0.3871*	0.0194*
	VBPR	*Base*	*0.6531*	*0.0293*	*0.3074*	*0.0141*	*0.3775*	*0.0159*
		FGSM	0.5824*	0.0299	0.6164*	0.0323*	0.5860*	0.0283*
		PGD	**1.1480**	**0.0538***	0.6410*	0.0324*	0.5918*	0.0286*
		C&W	0.6132*	0.0290	**0.6880***	**0.0336***	**0.6642***	**0.0348***
	AMR	*Base*	*0.3944*	*0.0196*	*0.5037*	*0.0232*	*0.1076*	*0.0038*
		FGSM	0.3347*	0.0150*	0.4426*	0.0235	0.4178*	0.0187*
		PGD	**0.8365**	**0.0418***	0.4519*	**0.0242**	0.4263*	0.0193*
		C&W	0.3678	0.0170*	0.4371*	0.0230	**0.4451***	**0.0202***
	ACF	*Base*	*0.5574*	*0.0278*	*0.3560*	*0.0176*	*0.3565*	*0.0176*
		FGSM	**0.5692***	**0.0282***	**0.3773***	**0.0185***	0.3517	0.0172*
		PGD	0.5610	0.0280	0.3731*	0.0183*	0.3521	0.0172*
		C&W	0.5628	0.0279	0.3690*	0.0181*	0.3471*	0.0169*
	DVBPR	*Base*	*0.6945*	*0.0359*	–	–	–	–
		FGSM	0.6579*	0.0329*	–	–	–	–
		PGD	0.5549*	0.0281*	–	–	–	–
		C&W	0.6414*	0.0306*	–	–	–	–

Test Time Adversarial Attacks and Defenses against Visual-based RSs

Motivations

- No works have been focused on protecting Visual RSs from **Test Time Attacks**.
- No defenses have been proposed to protect from test time adversarial attacks.

Research Questions

Can test time adversarial attacks misuse the behavior of trained recommenders?

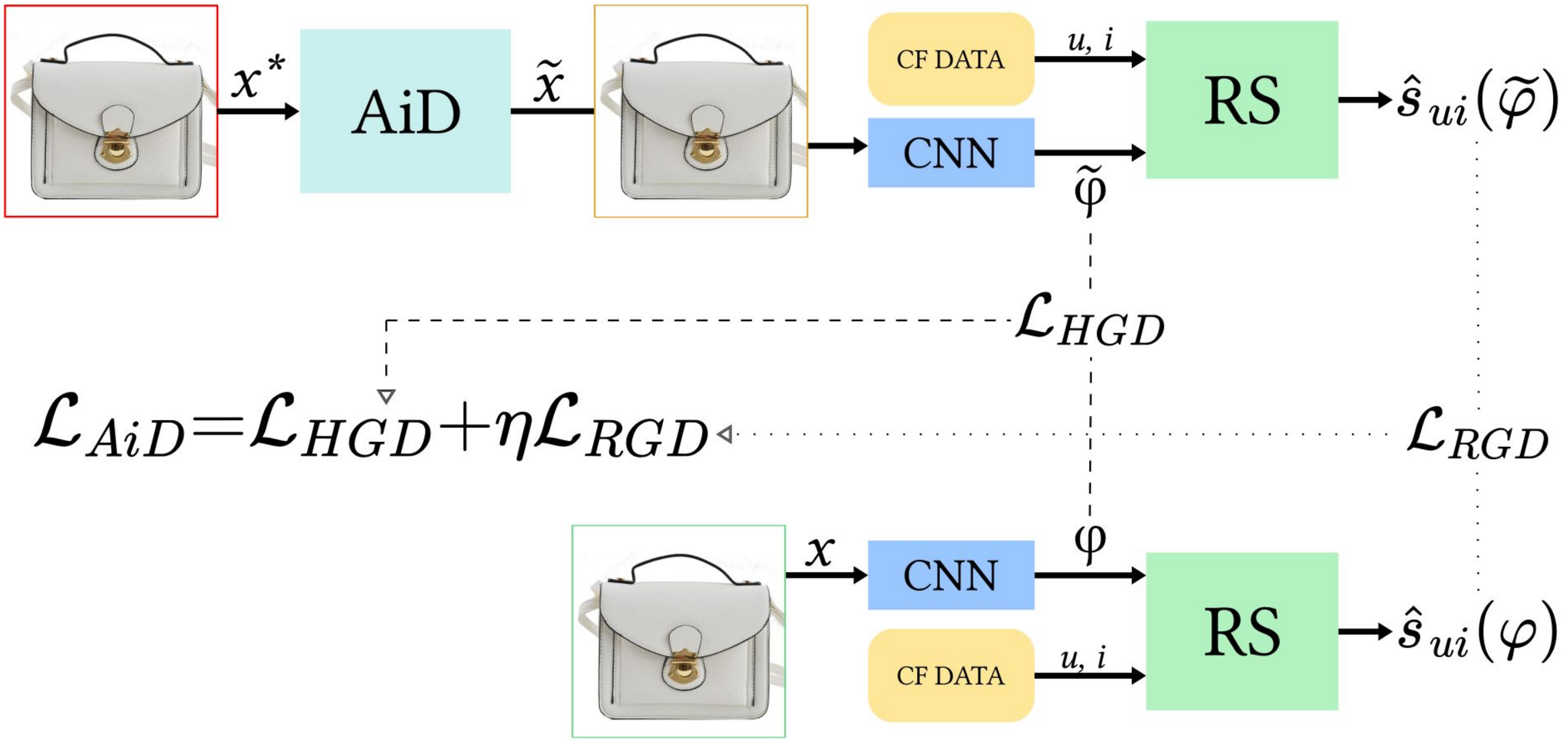
Can an adversarial image denoiser (our proposal) reduce the effectiveness of adversaries?

Reference Publications as Main Author

AiD: Adversarial Image Denoiser to Protect Visual-based Recommender Systems
***Felice Antonio Merra**, Vito Walter Anelli, Tommaso Di Noia, Daniele Malitesta, Alberto Carlo Maria Mancino*
Under Review

Adversarial Attacks against Visual Recommendation: an Investigation on the Influence of Items' Popularity
*Vito Walter Anelli, Tommaso Di Noia, Eugenio Di Sciascio, Daniele Malitesta, **Felice Antonio Merra***
OHARS@RecSys 2021

Method: Adversarial Image Denoiser (AiD)



Results

AiD is an **Effective Defensive Solution**:

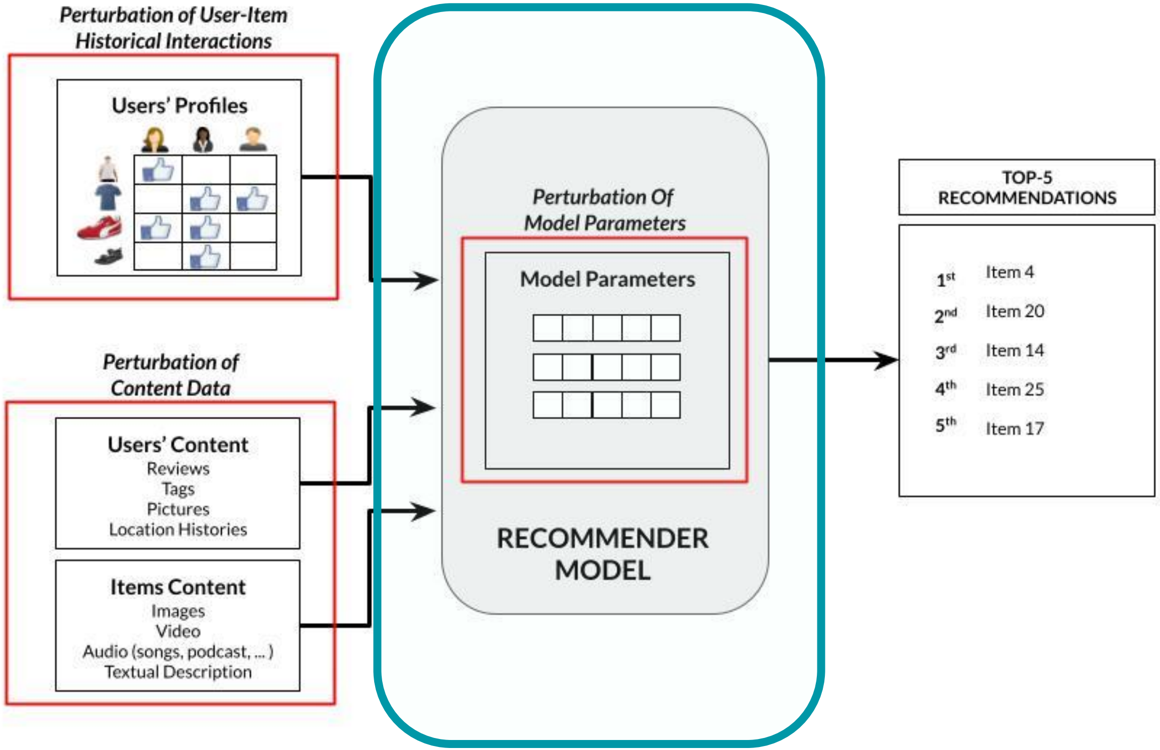
- **Attack effects** measured with the Prediction Shift **are lower** than the not defended case.
- **Predicted scores get low variations** also with stronger and stronger attacks.
- **Accuracy** and beyond-accuracy values are mostly **preserved** when compared with the not-defended recommender.

TAKE HOME MESSAGE

Protecting a VRS by removing the noise from images has been shown to be an effective solution.

Dataset	Model	Attack	PS^{wo}	PS^{w}
Amazon Boys& Girls	VBPR	BB-TAaMR	-0.1437	**0.0507**
		WB-INSA	0.8250	**0.1410**
		WB-SIGN	1.8466	**1.2668**
	AMR	BB-TAaMR	0.4643	0.6648
		WB-INSA	1.0432	**0.2193**
		WB-SIGN	1.3349	**1.1183**
Amazon Men	VBPR	BB-TAaMR	-0.1072	0.1105
		WB-INSA	2.2217	**0.5560**
		WB-SIGN	2.2413	**1.0005**
	AMR	BB-TAaMR	-0.0803	**-0.0423**
		WB-INSA	2.2418	**0.6057**
		WB-SIGN	2.5066	**1.0969**
Pinterest	VBPR	BB-TAaMR	0.4784	**0.1729**
		WB-INSA	1.9113	**0.4931**
		WB-SIGN	1.8929	**0.6434**
	AMR	BB-TAaMR	0.7163	**0.1470**
		WB-INSA	1.3108	**0.2205**
		WB-SIGN	1.2817	**0.3345**

Focus of the following contributions



Iterative Adversarial Perturbations on the Parameters of Model-based RSs

Motivations

- Model based RSs are not robust to **adversarial perturbations added on the learned parameters**.
- An adversarial training procedure has been propose to robustify a RS against this attack, however, no studies have been conducted on iterative versions that have be demonstrated to be much more dangerous.

Research Questions

How vulnerable are the parameters to iterative gradient-based adversarial methods?

Is Adversarial Personalized Ranking effective in robustifying the model against iterative methods?

Reference Publication as Main Author

MSAP: Multi-Step Adversarial Perturbations on Recommender Systems Embeddings.

Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, ***Felice Antonio Merra***

AdvML@KDD 2021

Method: Multi-Step Adversarial Perturbation (MSAP)

FGSM-based *multi-step* strategy and create more effective ϵ-clipped perturbations.

The initial model parameters are defined as $\mathbf{\Theta}_0^{adv} = \mathbf{\Theta} + \mathbf{\Delta}_0$

Let $Clip_{\mathbf{\Theta},\epsilon}$ be an element-wise clipping function to limit the perturbation in $[-\epsilon, +\epsilon]$

Let α be the step size which is the maximum perturbation budget of each iteration

Let L be the number of iterations

MSAP

$$\boxed{\mathbf{\Theta}_l^{adv} = Clip_{\mathbf{\Theta},\epsilon}\left\{\mathbf{\Theta}_{l-1}^{adv} + \alpha\frac{\Pi}{\|\Pi\|}\right\} \text{ where } \Pi = \frac{\partial\mathcal{L}(\mathbf{\Theta}+\mathbf{\Delta}_{l-1}^{adv})}{\partial\mathbf{\Delta}_{l-1}^{adv}}}$$

where $l \in [1, 2, ..., L]$

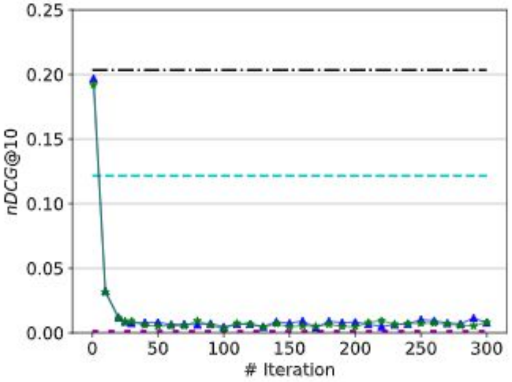
$\mathbf{\Delta}_l^{adv}$ is the adversarial perturbation at the l-th iteration

$\mathbf{\Theta}_l^{adv}$ is the sum of the original model parameters $\mathbf{\Theta}$ with the perturbation at the l-th iteration

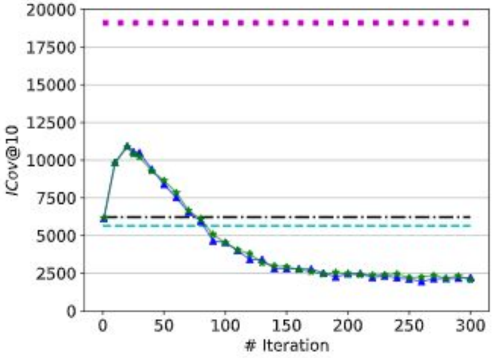
Results

MSAP:

- Impaired an ML-RS making worse thar a **random** recommender,
- Impacted the performance of the adversarial protected version by **-50%**,
- Produced the same performance drop as of **FGSM** with **5-time smaller budget**



(a) $nDCG$ on BPR-MF



(c) $ICov$ on BPR-MF

TAKE HOME MESSAGE

How to robustify the recommender against MSAP to avoid minimal variations that can make a ML-RS working in a random way?

Theoretical Modeling of Adversarial Training on Recommendations

Motivations

- ML-RSs are not robust to **adversarial perturbations added on the learned parameters**.
- **Adversarial Regularization** is the widely adopted solutions in more than 15 novel RSs.
- No studies have been conducted to understand the **reasons of robustness**.

Research Questions

Since adversarial training has been demonstrated to disturb the model accuracy in the image classification task, how does it influence the recommendation performance on accuracy and beyond-accuracy perspectives?

Reference Publications as Main Author

Understanding the Effects of Adversarial Personalized Ranking Optimization Method on Recommendation Quality.
*Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, **Felice Antonio Merra***
AdvML@KDD 2021

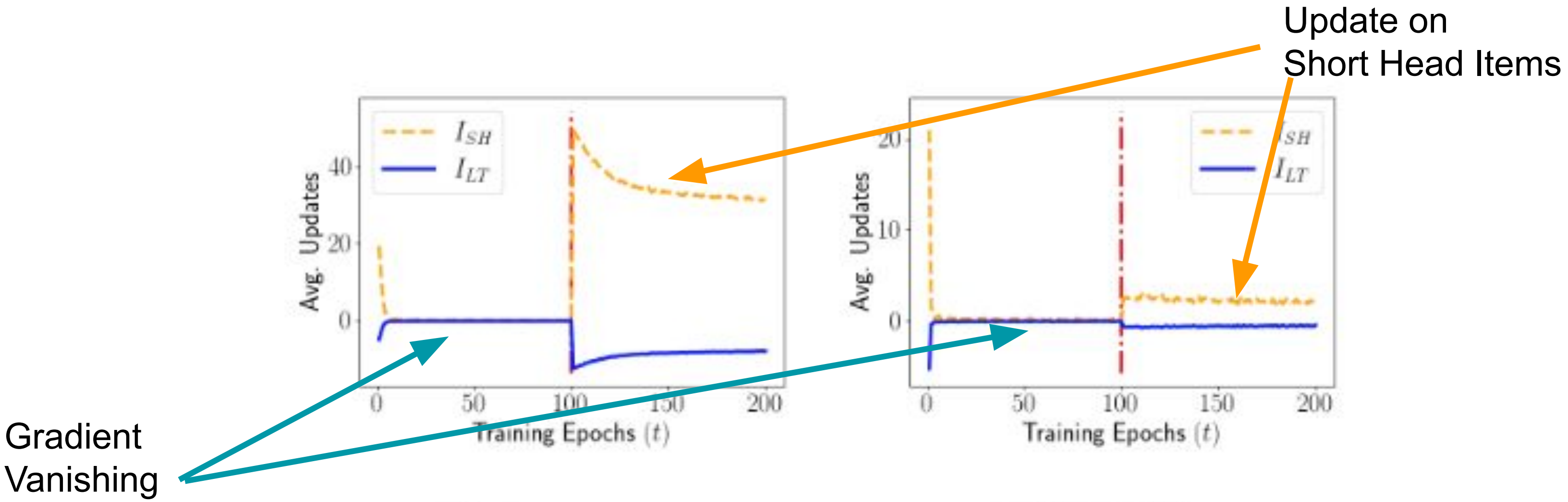
A Formal Analysis of Recommendation Quality of Adversarially-trained Recommenders.
*Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, **Felice Antonio Merra***
CIKM 2021

Gradient Magnitudes

$$\Theta \leftarrow \Theta + \eta \underbrace{\left(1 - \sigma\left(\hat{s}_{uij}(\Theta)\right)\right)}_{\omega_{uij}\text{: Bayesian Grad. Magnitude}} \frac{\partial \hat{s}_{uij}(\Theta)}{\partial \Theta}$$

$$\Theta \leftarrow \Theta + \eta \Big[\left(1 - \sigma\left(\hat{s}_{uij}(\Theta)\right)\right) \frac{\partial \hat{s}_{uij}(\Theta)}{\partial \Theta} + \alpha \underbrace{\left(1 - \sigma\left(\hat{s}_{uij}(\Theta + \Delta_{adv})\right)\right)}_{\omega_{uij}^{adv}\text{: Adversarial Grad. Magnitude}} \frac{\partial \hat{s}_{uij}(\Theta + \Delta_{adv})}{\partial \Theta} \Big]$$

Amplification of Popularity Bias



(a) Amazon

(b) ML100K.

The Global number of positive updates on short-head items is higher than the one on long-tail ones

Results

Model	Accuracy			Beyond		Popularity Bias		
	Rec	*Prec*	*nDCG*	*Nov*	$Cov_{\%}$	$ARP \uparrow$	$APLT \downarrow$	$ACLT \downarrow$
ML100K								
BPR-MF	0.3871	0.0077	0.1222	2.7653	71.22	176.64	0.2890	14.4486
APR-MF	0.3966	0.0079	0.1260*	2.7577*	71.22*	177.33*	0.2841*	14.2068*
R.V.	**+2.47%**	**+2.47%**	**+3.15%**	**-0.27%**	**0.00%**	**+0.39%**	**-1.67%**	**-1.67%**
Amazon								
BPR-MF	0.2077	0.0042	0.0656	6.0431	99.37	106.59	0.3541	17.7055
APR-MF	0.2130	0.0043	0.0687*	5.6805*	90.58*	131.30*	0.2829*	14.1471*
R.V.	**+2.58%**	**+2.58%**	**+4.63%**	**-6.00%**	**-8.85%**	**+23.18%**	**-20.10%**	**-20.10%**

- APR can negatively influence the beyond-accuracy recommendation performance
- APR can amplify the popularity bias more than BPR

TAKE HOME MESSAGE

It is fundamental to understand the effects of defenses also on beyond-accuracy metrics.

Contributions from the review on AML in RSs

Book Chapter

Adversarial Recommender Systems: Attack, Defense, and Advances
*Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, **Felice Antonio Merra***
The 3rd Edition of the Recommender Systems Handbook. 2022

Survey

A Survey on Adversarial Recommender Systems: From Attack/Defense Strategies to Generative Adversarial Networks
*Yashar Deldjoo, Tommaso Di Noia, **Felice Antonio Merra***
ACM Computing Survey 2021

Tutorials

Adversarial Learning for Recommendation
*Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, **Felice Antonio Merra***
ECIR 2021

Adversarial Learning for Recommendation: Applications for Security and Generative Tasks - Concept to Code
*Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, **Felice Antonio Merra***
RecSys 2020

Adversarial Machine Learning in Recommender Systems (AML-RecSys)
*Yashar Deldjoo, Tommaso Di Noia, **Felice Antonio Merra***
WSDM 2020

Resource

Elliot: A Comprehensive and Rigorous Framework for Reproducible Recommender Systems Evaluation.
Vito Walter Anelli, Alejandro Bellogín, Antonio Ferrara, Daniele Malitesta, Felice Antonio Merra, Claudio Pomo, Francesco Maria Donini, Tommaso Di Noia:
SIGIR 2021

Conclusions

The research contributions presented in my dissertation pave the way towards more robust recommender systems.

We have shown the limits of the existing defenses against novel adversarial attacks we have proposed possible solutions.

The attention towards a complete analysis of the recommendation quality of defended models should motivate defense proposals that also consider beyond-accuracy aspects.

Accepted Publications (Bold when Main Author)

2022

1. Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, **Felice Antonio Merra**, Adversarial Recommender Systems: Attack, Defense, and Advances, The 3rd Edition of the Recommender Systems Handbook. 2022
2. Yashar Deldjoo; Tommaso Di Noia; Daniele Malitesta; Felice Antonio Merra, Leveraging Content-Style Item Representation for Visual Recommendation, ECIR 2022
3. Vito Walter Anelli, Alejandro Bellogin, Tommaso Di Noia, Francesco Donini, Antonio Ferrara, Daniele Malitesta, Felice Antonio Merra, Claudio Pomo, V-Elliot: Speeding up Visual Recommendation via a GPU-powered Data Input Pipeline, NVIDIA GTC 2022

2021

4. Vito Walter Anelli; Yashar Deldjoo; Tommaso Di Noia; **Felice Antonio Merra**, A Formal Analysis of Recommendation Quality of Adversarially-trained Recommenders, CIKM 2021
5. Vito Walter Anelli; Tommaso Di Noia; **Felice Antonio Merra**, The Idiosyncratic Effects of Adversarial Training on Bias in Personalized Recommendation Learning, RecSys 2021
6. Vito Walter Anelli, Alejandro Bellogin, Tommaso Di Noia, Francesco Donini, Antonio Ferrara, Daniele Malitesta, Felice Antonio Merra, Claudio Pomo, V-Elliot: Build, Evaluate and Tune Visual Recommender Systems, RecSys 2021
7. Vito Walter Anelli , Tommaso Di Noia, Eugenio Di Sciascio, Daniele Malitesta and **Felice Antonio Merra**, Adversarial Attacks against Visual Recommendation: an Investigation on the Influence of Items' Popularity, OHARS@RecSys2021
8. Vito Walter Anelli; Yashar Deldjoo; Tommaso Di Noia; **Felice Antonio Merra**, Understanding the Effects of Adversarial Personalized Ranking Optimization Method on Recommendation Quality, 3rd Workshop on Adversarial Learning Methods for Machine Learning and Data Mining @ KDD 2021 (virtual workshop)
9. Yashar Deldjoo, Tommaso Di Noia, Eugenio Di Sciascio, **Felice Antonio Merra**, A Regression Framework to Interpret the Robustness of Recommender Systems Against Shilling Attacks, IIR 2021
10. Adversarial Learning for Recommendation, Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, **Felice Antonio Merra**, ECIR 2021
11. Vito Walter Anelli, Alejandro Bellogin, Tommaso Di Noia, Francesco Donini, Antonio Ferrara, Daniele Malitesta, Felice Antonio Merra, Claudio Pomo, How to perform reproducible experiments in the ELLIOT recommendation framework: data processing, model selection, and performance evaluation, IIR 2021
12. Giuseppe De Candia, Tommaso Di Noia, Eugenio Di Sciascio, **Felice Antonio Merra**, AMFLP: Adversarial Matrix Factorization-based Link Predictor in Social Graphs, SEBD 2021.
13. Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, Daniele Malitesta, **Felice Antonio Merra**, A Study of Defensive Methods to Protect Visual Recommendation Against Adversarial Manipulation of Images,The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval
14. Vito Walter Anelli, Alejandro Bellogin, Tommaso Di Noia, Francesco Donini, Antonio Ferrara, Daniele Malitesta, Felice Antonio Merra, Claudio Pomo, Elliot: a Comprehensive and Rigorous Framework for Reproducible Recommender Systems Evaluation, SIGIR 2021
15. Yashar Deljoo,, Tommaso Di Noia, Daniele Malitesta, Felice Antonio Merra, A Study on the Relative Importance of Convolutional Neural Networks in Visually-Aware Recommender Systems,The 4th CVPR Workshop on Computer Vision for Fashion, Art, and Design

Accepted Publications (Bold when Main Author)

16. Vito Walter Anelli, Alejandro Bellogin, Yashar Deldjoo, Tommaso Di Noia, **Felice Antonio Merra**. MSAP: Multi-Step Adversarial Perturbations on Recommender Systems Embeddings, FLAIRS 2021
17. Yashar Deldjoo, Tommaso Di Noia, **Felice Antonio Merra**, A survey on Adversarial Recommender Systems: from Attack/Defense strategies to Generative Adversarial Networks, ACM Computing Surveys, March 2021

2020

18. Vito Walter Anelli, Tommaso Di Noia, Daniele Malitesta, **Felice Antonio Merra**, Assessing Perceptual and Recommendation Mutation of Adversarially-Poisoned Visual Recommenders, WSCD@NeurIPS2020, Vancouver, Canada (Virtual Event) Code.
19. Vito Walter Anelli, Tommaso Di Noia, Daniele Malitesta, **Felice Antonio Merra**, An Empirical Study of DNNs Robustification Inefficacy in Protecting Visual Recommenders, arXiv.
20. Vito Walter Anelli, Alejandro Bellogín, Yashar Deldjoo, Tommaso Di Noia, **Felice Antonio Merra**, Multi-Step Adversarial Perturbations on Recommender Systems Embeddings , arXiv.
21. Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, **Felice Antonio Merra**, Adversarial Learning for Recommendation: Applications for Security and Generative Tasks - Concept to Code, To appear in Proceedings of the 14th ACM Conference on Recommender Systems, RecSys 2020, Virtual Conference (Brazil).
22. Yashar Deldjoo, Tommaso Di Noia, Eugenio Di Sciascio, **Felice Antonio Merra**, How Dataset Characteristics Affect the Robustness of Collaborative Recommendation Models, SIGIR 2020. X'ian, China, July, 2020. Video
23. Tommaso Di Noia, Daniele Malitesta, **Felice Antonio Merra**, TAaMR: Targeted Adversarial Attack against Multimedia Recommender Systems, The 3rd International Workshop on Dependable and Secure Machine Learning – DSML 2020 Co-located with the 50th IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2020). Code Video
24. Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, **Felice Antonio Merra**, Eugenio Di Sciascio, Giuseppe Acciani, Knowledge-enhanced Shilling Attacks for recommendation, To appear in Proceedings of the 28th Italian Symposium on Advanced Database Systems, June 21-24, 2020 - Villasimius, Sardinia, Italy
25. Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, **Felice Antonio Merra**, Eugenio Di Sciascio, SAShA: Semantic-Aware Shilling Attacks on Recommender Systems exploiting Knowledge Graphs, The 17th Extended Semantic Web Conference. Springer, Cham., Heraklion, Greece, May 31- June 4, 2020. Video
26. Yashar Deldjoo, Tommaso Di Noia, **Felice Antonio Merra**, Adversarial Machine Learning in Recommender Systems, Slide, The 13th ACM International WSDM Conference, Texas, February 3-7, 2020.

2019

27. Yashar Deldjoo, Tommaso Di Noia, **Felice Antonio Merra**, Assessing Knowledge-enhanced Shilling Attacks for recommendation the Impact of a User-Item Collaborative Attack on Class of Users, ImpactRS@RecSys 2019, Copenhagen, Denmark, September 19, 2019.