Adversarial Learning for Recommendation: Applications for Security and Generative Tasks — Concept to Code

VITO WALTER ANELLI, Polytechnic University of Bari, Italy
YASHAR DELDJOO, Polytechnic University of Bari, Italy
TOMMASO DI NOIA, Polytechnic University of Bari, Italy
FELICE ANTONIO MERRA[*], Polytechnic University of Bari, Italy

Adversarial Machine Learning (AML) has initially emerged as the field of study that investigates security issues of conventional and modern machine learning (ML) models. The objective of this tutorial is to present a comprehensive overview on the application of AML techniques for recommendation in a two-fold categorization: (i) AML for the attack/defense purposes, and (ii) AML to build GAN-based recommender models. A theoretical presentation on the topics is paired with two corresponding hands-on sessions to show the efficacy of AML application and push up novel ideas and advances in recommendation tasks. The tutorial is divided into four parts. We start by introducing a summary on state-of-the-art recommender models, including deep learning ones, and we define the fundamentals of AML. Then, we present the Adversarial Recommendation Framework, to represent attack/defense strategies on RSs, and the GAN-based Recommendation Framework, which is at the basis of novel adversarial-based generative recommenders. The presentation of each framework is followed by a practical session. Finally, we conclude with open challenges and possible future works for both applications.

Additional Key Words and Phrases: Adversarial Machine Learning, Recommender Systems, Security

ACM Reference Format:
Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, and Felice Antonio Merra. 2020. Adversarial Learning for Recommendation: Applications for Security and Generative Tasks — Concept to Code. In *Fourteenth ACM Conference on Recommender Systems (RecSys '20), September 21–26, 2020, Virtual Event, Brazil.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3383313.3411447

1 MOTIVATION AND TOPIC IMPORTANCE

Modern recommender systems utilize a variety of machine learning (ML) models, such as the one based on latent factor models (LFMs), to provide users with relevant suggestions about products in a customized fashion. From a ML perspective, such systems' task can be viewed as a matrix completion task, which aims to predict unknown ratings of an incomplete user-by-item matrix or rank optimization, achieved by placing more relevant items on top of the recommendation ranking list by using a learning-to-rank approach [16,17]. An empirical loss function is used in either of these scenarios to guide the recommendation model learning in the training phase. Notwithstanding the great success of LFMs, which has resulted in a wide variety of models and evaluations metrics, they try to find any available signal in the user interaction data and side-information [11, 22] that can improve recommendation performance accuracy and beyond-accuracy aspects [18]. Unfortunately, these models are not often prepared to handle challenging scenarios,

[*] Authors are listed in alphabetical order. Corresponding author: felice.merra@poliba.it

Manuscript submitted to ACM

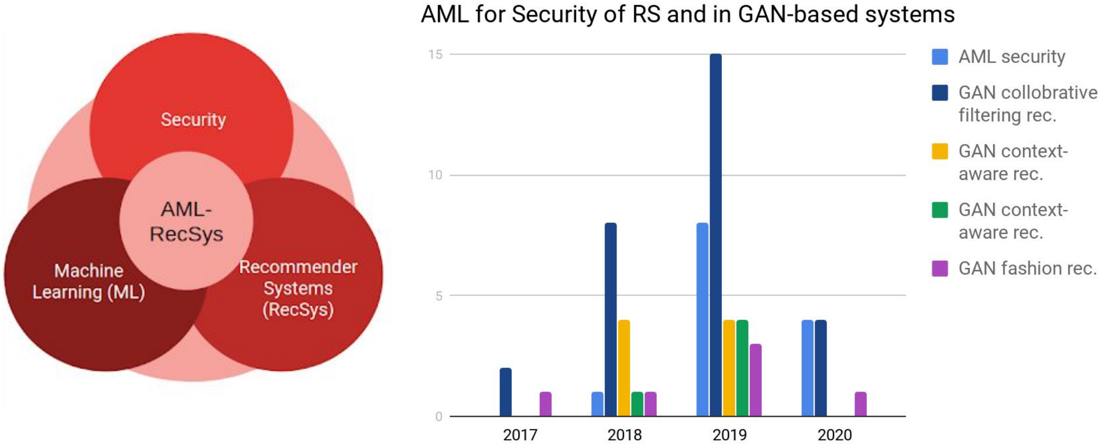

Fig. 1. (Left) AML-RecSys combines best practices in ML and security to (i) improve data security in recommender systems tasks or (ii) better mimic the user profile in GAN-based tasks. (Right) Number of papers published related to AML-RecSys.

e.g., malicious data injection that can be performed by a malicious agent to guide the recommendation toward an engineered outcome. Security of recommender systems (RS) has been studied in the context of shilling attacks (between 2000-2015) [1, 3, 7, 8] and attacks based on adversarial machine learning (AML) (between 2015-2020) [14, 17, 27]. ML-based approaches for shilling attacks [16, 19], which are, to some extent, similar to AML from the perspective of using ML-learned techniques. Shilling attacks focus on leveraging the similarities and correlations between user rating profiles to design hand-crafted rating patterns identical to those already in the system to impact the training phase and push the desired item into the recommendation list of users (push attack) or to recommend non-relevant items (to create a mistrust on a system). Attacks based on AML, on the other hand, focus on learning ML-driven perturbations to be added to the data to attack recommendation models at *test time*. As shown in Figure 1, AML-RecSys combines best practices in the field of security, ML, and RecSys to learn about adversarial attacks and defensive mechanisms. Adversarial attacks are realized by learning non-random norm-constrained perturbations added to raw data (in the case of images) or model parameters of an ML predictor (e.g., user-item embeddings in a LFM) to reduce the prediction quality.

A drastic reduction of recommendation accuracy due to adversarial attacks has been verified in several recent works on RSs, i.e., APR [17], AMR [23], FG-ACAE [26], ATF [6], and TAaMR [12]. On the other side, defense mechanisms have been tested to improve the robustness of RSs under adversarial attacks (e.g., adversarial training [17], defensive distillation [13]). In the first part of this tutorial, we aim to help researchers and practitioners in understanding how to find possible vulnerabilities of RSs, and, eventually, to identify or propose defense strategies.

The adversarial learning paradigm is also the critical element of a novel deep neural generative model, named Generative Adversarial Network (GAN). In this tutorial, we devote a considerable portion of the presentation to introduce a summary of the main advantages of GAN-based recommender systems including more informative negative sampling step in learning-to-rank models [15, 25]; learn the generator to estimate missing ratings by leveraging both temporal [2, 28] and side-information [5, 24]; or reducing cold-start problems by augmenting the training dataset [4, 13]. Figure 1 summarizes the number of AML-RecSys works published in the last few years, which were reviewed and studied in our recent survey [10] and previously presented in a tutorial at WSDM'20 [9].

Summing up, in this tutorial we give the audience an introduction on the application of adversarial learning for RSs: (i) the one focused on security issues; (ii) the other investigating the adoption of generative models in a recommendation scenario. In both cases, the attendees will be supported by practical sessions (by means of Jupyter notebooks) designed for a deeper understanding of the main aspects of AML-RecSys.

2 OUTLINE OF THE TUTORIAL

The tutorial is scheduled in five main slots.

- **Introduction to Recommender Systems and Deep Learning in RS (15 mins).** We start the tutorial with a brief introduction to recommender systems (RSs) and their evolution over the past decade, starting from recommending service described in Grundy (1972) [21], we will briefly show the moving from the **classical non-neural era** characterized by enhancing the recommendation accuracy to the **post neural era** where we assisted to the transition from classical learning to deep learning and AML techniques.
- **Foundations of Adversarial Machine Learning in Recommender Systems (30 mins).** We present here the basic definitions and notations used across the following sections. Then, we show a brief overview of the more than 60 publications of AML applications in Recommender Systems collected from top-tier conferences (e.g., SIGIR, RecSys, KDD, WSDM, and IJCAI). Then, we comment on the categorization of AML usage in recommendation scenarios:
  - *AML for the security of RS*: This is the "principal application" of AML in RS, which focuses on adversarial attacks and defense models in RS.
  - *Application of AML in GANs*: This is a topic derived from AML, that is focused on "generative" learning models.

  Starting from this categorization, we conduct the rest of the tutorial.
- **Adversarial Machine Learning for Security of RS (60 mins).**
  - *Literature review and main concepts AML for security (30 mins).* In this initial section, we will discuss the main publications, show task-based and model-based graphical representations and present a general framework to classify AML approaches in a recommendation task.
  - *Hands-on session on AML for security (30 mins).* After the identification of the main works and the presentation of a basic adversarial framework of AML in RSs, we will provide to the attendees a Jupyter notebook where they can verify step-by-step how much the recommendation performance is reduced by an adversarial perturbation. Then, the hand-son will continue by showing how to implement the **adversarial training** procedure to make the model robust to the previously experimented **adversarial attack**. The goal of this practice section is to make the attender more confident and aware of state-of-the-art adversarial attack issues and defense mechanisms.
- **Adversarial Learning for GAN-based Recommendation (60 mins).**
  - *Literature review and main concepts of GAN-based recommender model (30 mins).* In this initial section, we will discuss the main publications, show task-based and architectural-based categorizations, and present a GAN-based framework to make evident how the adversarial learning may be used in GAN models to address different recommendation tasks (e.g., complementary recommendation, cross-domain recommendation) and solve cold-start issues (e.g., more informative dynamic sampling, data augmentation).
  - *Hands-on session on GAN-based RS (30 mins).* We provide to the attendees a Jupyter notebook where they can verify step-by-step how the typical adversarial *minimax game* between the generator and the discriminator of a GAN is used in recommendation tasks (e.g., with BPR-MF [20]). The purpose of this practical session is to help researchers in understanding how to use the GAN-based generative paradigm in a recommendation model.

- **Conclusions, Grand Challenges and Discussion (15 mins).** We round off the tutorial by giving a brief summary, communicating the main take away messages, and providing some practical guidelines for researchers new to the area of AML in RSs. Via the identification and discussion of the open challenges, we further guide researchers and practitioners new to the topic, and hopefully, help them shape their ideas for future research directions on this interesting field. Such challenges include, among others:

(1) promote the development of approaches to identify possible security issues of ML-based recommendation models;
(2) propose novel defense approaches to improve the robustness of the recommendation system;
(3) encourage the research towards novel recommendation models that can exploit GAN to build novel recommender models.

3 ADDITIONAL INFORMATION

Intended Audience: We target researchers and practitioners willing to bridge the gap in perspectives and advances between the Deep Learning and Recommender System fields. We foresee a full-day tutorial of **180** minutes (3 hours) duration. Particular prerequisite knowledge or skills are not required from the audience, except for a basic understanding of the main concepts in recommender systems and machine learning. In the tutorial, we will cover both academic and industrial points of view reflected in the background of the presenters. Accompanying the tutorial, we will publish online a comprehensive set of slides, including references to state-of-the-art works and open implementations of several of the presented techniques, and two Jupyter notebooks used for the hand-on sessions.

Previous Offering of the Tutorial: An initial version of this tutorial has been previously offered as a half-day tutorial to *the 13th ACM International WSDM Conference in Houston, Texas from February 3-7, 2020*. The slides are publicly available on the GitHub repository [1].

Type of Support Materials: The tutorial will be supported by:

- GitHub Repository with an overview of the program, all the references and presenters detailed.
- Tutorial slides & Two hands-on sessions: one for the security applications, and one for GAN-based RSs.
- A comprehensive overview [10] within its GitHub repository containing the references with the links to the papers and implementations [2].

ACKNOWLEDGMENT

The authors acknowledge partial support of the following projects: Innonetwork CONTACT, ARS01_00821 FLET4.0, Fincons Smart Digital Solutions for the Creative Industry, PON OK-INSAID.

REFERENCES

[1] Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, Eugenio Di Sciascio, and Felice Antonio Merra. 2020. Sasha: Semantic-aware shilling attacks on recommender systems exploiting knowledge graphs. In *European Semantic Web Conference*. Springer, 307–323.

[2] Homanga Bharadhwaj, Homin Park, and Brian Y. Lim. 2018. RecGAN: recurrent generative adversarial networks for recommendation systems. In *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2-7, 2018*. 372–376.

[3] Robin Burke, Michael P. O'Mahony, and Neil J. Hurley. 2015. Robust Collaborative Recommendation. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer, 961–995. https://doi.org/10.1007/978-1-4899-7637-6_28

[4] Dong-Kyu Chae, Jin-Soo Kang, Sang-Wook Kim, and Jaeho Choi. 2019. Rating Augmentation with Generative Adversarial Networks towards Accurate Collaborative Filtering. In *WWW*. ACM, 2616–2622.

[1] https://github.com/sisinflab/amlrecsys-tutorial
[2] https://github.com/sisinflab/adversarial-recommender-systems-survey

[5] Dong-Kyu Chae, Jin-Soo Kang, Sang-Wook Kim, and Jung-Tae Lee. 2018. CFGAN: A Generic Collaborative Filtering Framework based on Generative Adversarial Networks. In *CIKM*. ACM, 137–146.

[6] Huiyuan Chen and Jing Li. 2019. Adversarial tensor factorization for context-aware recommendation. In *RecSys*. ACM, 363–367.

[7] Yashar Deldjoo, Tommaso Di Noia, Felice Antonio Merra, and Eugenio Di Sciascio. 2020. How Dataset Characteristics Affect the Robustness of Collaborative Recommendation Models. In *Proc. of ACM SIGIR 2020 - 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM Press. http://sisinflab.poliba.it/publications/2020/DDMD20 to appear.

[8] Yashar Deldjoo, Tommaso Di Noia, and Felice Antonio Merra. 2019. Assessing the Impact of a User-Item Collaborative Attack on Class of Users. In *Proceedings of the 1st Workshop on the Impact of Recommender Systems co-located with 13th ACM Conference on Recommender Systems, ImpactRS@RecSys 2019), Copenhagen, Denmark, September 19, 2019.* http://ceur-ws.org/Vol-2462/paper2.pdf

[9] Yashar Deldjoo, Tommaso Di Noia, and Felice Antonio Merra. 2020. Adversarial Machine Learning in Recommender Systems (AML-RecSys). In *WSDM '20: The Thirteenth ACM International Conference on Web Search and Data Mining, Houston, TX, USA, February 3-7, 2020*, James Caverlee, Xia (Ben) Hu, Mounia Lalmas, and Wei Wang (Eds.). ACM, 869–872. https://doi.org/10.1145/3336191.3371877

[10] Yashar Deldjoo, Tommaso Di Noia, and Felice Antonio Merra. 2020. Adversarial Machine Learning in Recommender Systems: State of the art and Challenges. *CoRR* abs/2005.10322 (2020). arXiv:2005.10322 https://arxiv.org/abs/2005.10322

[11] Yashar Deldjoo, Markus Schedl, Paolo Cremonesi, and Gabriella Pasi. 2020. Recommender Systems Leveraging Multimedia Content. *Comput. Surveys* (2020). https://doi.org/10.1145/3407190

[12] Tommaso Di Noia, Daniele Malitesta, and Felice Antonio Merra. 2020. TAaMR: Targeted Adversarial Attack against Multimedia Recommender Systems. In *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) (2020)*. IEEE, IEEE Digital Library. http://sisinflab.poliba.it/publications/2020/DMM20

[13] Yali Du, Meng Fang, Jinfeng Yi, Chang Xu, Jun Cheng, and Dacheng Tao. 2019. Enhancing the Robustness of Neural Collaborative Filtering Systems Under Malicious Attacks. *IEEE Trans. Multimedia* 21, 3 (2019), 555–565. https://doi.org/10.1109/TMM.2018.2887018

[14] Negin Entezari, Saba A. Al-Sayouri, Amirali Darvishzadeh, and Evangelos E. Papalexakis. 2020. All You Need Is Low (Rank): Defending Against Adversarial Attacks on Graphs. In *WSDM 2020*.

[15] Wenqi Fan, Tyler Derr, Yao Ma, Jianping Wang, Jiliang Tang, and Qing Li. 2019. Deep Adversarial Social Recommendation. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019.* 1351–1357. https://doi.org/10.24963/ijcai.2019/187

[16] Minghong Fang, Guolei Yang, Neil Zhenqiang Gong, and Jia Liu. 2018. Poisoning Attacks to Graph-Based Recommender Systems. In *Proceedings of the 34th Annual Computer Security Applications Conference, ACSAC 2018, San Juan, PR, USA, December 03-07, 2018.* ACM, 381–392. https://doi.org/10.1145/3274694.3274706

[17] Xiangnan He, Zhankui He, Xiaoyu Du, and Tat-Seng Chua. 2018. Adversarial Personalized Ranking for Recommendation. In *SIGIR*. ACM, 355–364.

[18] Yehuda Koren and Robert Bell. 2015. Advances in collaborative filtering. In *Recommender systems handbook*. Springer, 77–118.

[19] Bo Li, Yining Wang, Aarti Singh, and Yevgeniy Vorobeychik. 2016. Data Poisoning Attacks on Factorization-Based Collaborative Filtering. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, Daniel D. Lee, Masashi Sugiyama, Ulrike von Luxburg, Isabelle Guyon, and Roman Garnett (Eds.). 1885–1893. http://papers.nips.cc/paper/6142-data-poisoning-attacks-on-factorization-based-collaborative-filtering

[20] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, June 18-21, 2009.* 452–461. https://dslpitt.org/uai/displayArticleDetails.jsp?mmnu=1&smnu=2&article_id=1630&proceeding_id=25

[21] Elaine Rich. 1979. User Modeling via Stereotypes. *Cognitive Science* 3, 4 (1979), 329–354. https://doi.org/10.1207/s15516709cog0304_3

[22] Yue Shi, Martha A. Larson, and Alan Hanjalic. 2014. Collaborative Filtering beyond the User-Item Matrix: A Survey of the State of the Art and Future Challenges. *ACM Comput. Surv.* 47, 1 (2014), 3:1–3:45. https://doi.org/10.1145/2556270

[23] J. Tang, X. Du, X. He, F. Yuan, Q. Tian, and T. Chua. 2019. Adversarial Training Towards Robust Multimedia Recommender System. *IEEE Transactions on Knowledge and Data Engineering* (2019), 1–1. https://doi.org/10.1109/TKDE.2019.2893638

[24] Jun Wang, Lantao Yu, Weinan Zhang, Yu Gong, Yinghui Xu, Benyou Wang, Peng Zhang, and Dell Zhang. 2017. IRGAN: A Minimax Game for Unifying Generative and Discriminative Information Retrieval Models. In *SIGIR*. ACM, 515–524.

[25] Qinyong Wang, Hongzhi Yin, Zhiting Hu, Defu Lian, Hao Wang, and Zi Huang. 2018. Neural Memory Streaming Recommender Networks with Adversarial Training. In *KDD*. ACM, 2467–2475.

[26] Feng Yuan, Lina Yao, and Boualem Benatallah. 2019. Adversarial Collaborative Auto-encoder for Top-N Recommendation. In *International Joint Conference on Neural Networks, IJCNN 2019 Budapest, Hungary, July 14-19, 2019.* 1–8. https://doi.org/10.1109/IJCNN.2019.8851902

[27] Hengtong Zhang, Yaliang Li, Bolin Ding, and Jing Gao. 2020. Practical Data Poisoning Attack against Next-Item Recommendation. In *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*, Yennun Huang, Irwin King, Tie-Yan Liu, and Maarten van Steen (Eds.). ACM / IW3C2, 2458–2464. https://doi.org/10.1145/3366423.3379992

[28] Wei Zhao, Benyou Wang, Jianbo Ye, Yongqiang Gao, Min Yang, and Xiaojun Chen. 2018. PLASTIC: Prioritize Long and Short-term Information in Top-n Recommendation using Adversarial Training. In *IJCAI*. ijcai.org, 3676–3682.