Adversarial Learning for Recommendation

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

Polytechnic University of Bari, Italy
{name.surname}@poliba.it

Abstract. Modern recommender systems (RSs) utilize a variety of machine learning (ML) models to provide users with relevant, personalized suggestions about products in a vast catalog. Notwithstanding the great success of ML models to make recommendations, they are often no-robust to adversarial actors, e.g., competitors, that might act to alter recommendations toward a malicious outcome. While the injection of hand-engineered fake profile, i.e., shilling attacks, [2,1] has been the core of investigation between years 2000 and 2015, the last years have been characterized by the rise of Adversarial Machine Learning (AML) techniques, i.e., ML-based approaches for attacking and defending RSs. In this tutorial, we present an overview of more than 75 publications on AML applications in RSs reviewed in our recent survey [3]. In particular, we introduce a twofold categorization of AML uses in RSs: the one based on the study of adversarial attacks, and defenses, against either the model parameters [5], content data [4], or user-item interactions [6]; the other one related to the use of Generative Adversarial Networks (GAN) to propose novel recommender models [7]. All the material is publicly available at github.com/sisinflab/amlrecsys-tutorial.

Keywords: Adversarial Machine Learning · Recommender Systems

References

1. Anelli, V.W., Deldjoo, Y., Di Noia, T., Di Sciascio, E., Merra, F.A.: Sasha: Semantic-aware shilling attacks on recommender systems exploiting knowledge graphs. In: ESWC. Lecture Notes in Computer Science, vol. 12123, pp. 307–323. Springer (2020)
2. Deldjoo, Y., Di Noia, T., Di Sciascio, E., Merra, F.A.: How dataset characteristics affect the robustness of collaborative recommendation models. In: ACM SIGIR 2020
3. Deldjoo, Y., Di Noia, T., Merra, F.A.: A survey on adversarial recommender systems: from attack/defense strategies to generative adversarial networks. ACM Computing Surveys (2021), https://doi.org/10.1145/3439729
4. Di Noia, T., Malitesta, D., Merra, F.A.: Taamr: Targeted adversarial attack against multimedia recommender systems. In: DSN Workshops. pp. 1–8. IEEE (2020)
5. He, X., He, Z., Du, X., Chua, T.: Adversarial personalized ranking for recommendation. In: SIGIR. pp. 355–364. ACM (2018)
6. Li, B., Wang, Y., Singh, A., Vorobeychik, Y.: Data poisoning attacks on factorization-based collaborative filtering. In: NIPS. pp. 1885–1893 (2016)
7. Wang, J., Yu, L., Zhang, W., Gong, Y., Xu, Y., Wang, B., Zhang, P., Zhang, D.: IRGAN: A minimax game for unifying generative and discriminative information retrieval models. In: SIGIR. pp. 515–524. ACM (2017)