

Second Workshop on Online Misinformation- and Harm-Aware Recommender Systems: Preface

Antonela Tommasel¹, Daniela Godoy¹ and Arkaitz Zubiaga²

¹*ISISTAN Research Institute (CONICET/UNCPBA), Tandil, Bs. As., Argentina*

²*Queen Mary University of London, London, UK*

Abstract

This volume contains the proceedings with the research contributions presented at the Second Workshop on Online Misinformation-and Harm-Aware Recommender Systems (OHARS'2021) co-located with the 15th ACM Recommender Systems Conference (RecSys'2021). These proceedings describe the specific workshop goals and format, and contain the papers presented during the online event held on October 2nd, 2021.

Keywords

Recommender systems, online harms, misinformation, hate speech

1. Introduction

In recent years, there has been an increase in the dissemination of false news, rumors, deception, and other forms of misinformation, as well as abusive language, incitements of violence, harassment, and other forms of hate speech throughout online platforms. While these phenomena are widely observed in social media, they affect users' experience on multiple online platforms. For example, collaborative filtering approaches in e-commerce sites are vulnerable to low-quality reviews, manipulation, and attacks.

Recommender systems play a central role in online information consumption and user decision-making by leveraging user-generated information at scale. As a result, they are affected by different forms of online harms, which may hinder the accuracy of predictions while, at the same time, become unintended means for their spread and amplification.

OHARS 2021 was the second edition of the Workshop on Online Misinformation- and Harm-Aware Recommender Systems¹, following the first edition, also co-located with ACM Recommender Systems Conference in 2020 [1]. This workshop aimed to bring together researchers in the recommender systems community interested in tackling online harms and mitigating their impact on recommendation, with a special interest in research tackling the negative effects of recommending fake or harmful content linked to the COVID-19 crisis. The end goal was to

OHARS'21: Second Workshop on Online Misinformation- and Harm-Aware Recommender Systems, October 2, 2021, Amsterdam, Netherlands

✉ antonela.tommasel@isistan.unicen.edu.ar (A. Tommasel); daniela.godoy@isistan.unicen.edu.ar (D. Godoy); a.zubiaga@qmul.ac.uk (A. Zubiaga)

🆔 0000-0002-5185-4570 (D. Godoy); 0000-0003-4583-3623 (A. Zubiaga)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

¹<https://ohars-recsys.isistan.unicen.edu.ar/>

facilitate the discussion about the major challenges and opportunities that will shape future research.

2. Accepted Papers

Six papers were accepted for presentation in the workshop, covering a broad range of technical aspects related to harm-aware recommender systems. Based on collaborative filtering, the most popular approach used in e-commerce platforms to improve user experience, different contributions analyze the impact of shilling and adversarial attacks as well as social polarization. Other contributions focus on fake news detection and recommendation as a means for helping users to take privacy-preserving decisions while using social media.

Shrestha et al. [2] analyze the effect of shilling attacks, i.e., malicious users creating fake profiles to provide fraudulent reviews, on recommender systems. The work explores the robustness of collaborative recommender systems to shilling attacks. Instead of simulating attacks, the impact of fraudulent reviews was quantified and evaluated on multiple real-world datasets, in which spam reviews were used as ground truth. In addition, the analysis studies whether non-mainstream users are more affected than mainstream users by spammers.

In Anelli et al. [3], authors study the collateral and negative impact of adversarial attacks against deep/convolutional neural networks (DNNs/CNNs) used in visually-aware recommender systems (VRSs). VRSs integrate products' image features with historical users' feedback to enhance recommendation performance. However, their integrity can be harmed by uploading item images with human-imperceptible adversarial perturbations capable of pushing a target item into higher recommendation positions. The paper presents an extensive evaluation of three state-of-the-art adversarial attacks against visual-based recommendations in multiple settings, varying the adversary knowledge (i.e., black- and white-box), the adversarial capability; and evaluating their performance on groups of target items. Given the importance of items' popularity on the recommendation performance, the work also analyzes whether items' popularity influences the effectiveness of the attacks.

Sun and Nasraoui [4] addressed the problem of polarization in collaborative filtering. A user polarization score is calculated based on specific rating patterns and then used for depolarizing a recommendation system. A user-polarization-aware matrix factorization (UpaMF) algorithm and a weighted alternative (WUpaMF) are presented to make recommendations that are less biased by extreme polarization. Algorithms were evaluated in terms of rank-based and value-based metrics and their capacity to improve the recommendation lists' diversity and reduce the blind spots induced by the recommendations. In the same line of research, Badami and Nasraoui [5] propose a novel polarization-aware recommender interactive system (PaRIS) to recommend relevant items, while at the same time including opposite views in case the user is interested in a different perspective. PaRIS uses a modified objective function that considers both relevance and polarization.

The paper by Mifsud et al. [6] addresses the problem of fake news detection. It investigates whether the application of transformer models, such as BERT, RoBERTa and ALBERT, can be leveraged to classify short claims according to six levels of veracity. The potential to enhance the overall classification results by adding neural network layers that use both the transformer's

output and the source’s reputation score is evaluated. Finally, the authors posit that considering language-based fake news classification on such short statements is potentially an ill-posed problem.

In the context of privacy-preserving recommender systems, Salem et al. [7] present an approach to mitigate the risks of self-disclosure of sensitive data. The work introduces the notion of disclosure appetite, a user-specific term encompassing user perception and drive to reveal their private information. Leveraging this information and the sensitivity of private data along with the context, disclosure mitigating recommendations are generated. A survey was carried out to evaluate the system’s effectiveness.

3. Program

OHARS’2021 was a half-day workshop in the context of RecSys’2021. The workshop program included short presentations and talks to discuss the different aspects of harm-aware recommender systems practice and experience. The workshop started with an opening keynote by Alexandra Olteanu (Microsoft Research) entitled “*What do we need to effectively measure computational harms?*”, on challenges in measuring harms. The second part of the workshop started with an invited talk by Paolo Rosso (Universitat Politècnica de València) entitled “*Detecting online harmful information: fake news, conspiracy theories and misogyny*”, on the identification of harmful items of information and the characterization of its spreaders. Authors of accepted submissions were invited to give presentations followed by Q&A and discussion.

4. Program Committee

We would like to thank the members of the Program Committee for their valuable contribution in providing timely and high-quality reviews.

- Esmá Aïmeur, Université de Montréal, Canada
- Giannis Bekoulis, ETRO-VUB, Belgium
- Ludovico Boratto, Eurecat
- Ivan Cantador, Universidad Autónoma de Madrid, Spain
- Tommaso Caselli, University of Groningen, Netherlands
- Lara Quijano Sanchez, Universidad Autónoma de Madrid, Spain
- Ana Maguitman, Universidad Nacional del Sur, Argentina
- Barbara Poblete, University of Chile, Chile
- Ravi Shekhar, Queen Mary University of London, UK
- Damiano Spina, RMIT University, Australia
- Marco Viviani, University of Milano-Bicocca, Italy

5. Acknowledgments

We would like to thank the RecSys 2021 workshop chairs, Jen Golbeck, Marijn Koolen and Denis Parra, for giving us the opportunity to host this workshop and all their assistance in the workshop organization.

The organizers are in part supported by the CONICET–Royal Society International Exchange (IEC\R2\192019).

References

- [1] A. Tommasel, D. Godoy, A. Zubiaga, Workshop on online misinformation- and harm-aware recommender systems, in: Proceedings of the 14th ACM Conference on Recommender Systems (RecSys '20), Virtual Event, Brazil, 2020, pp. 638–639.
- [2] A. Shrestha, F. Spezzano, M. S. Pera, An empirical analysis of collaborative recommender systems robustness to shilling attacks, in: Proceedings of the 2nd Workshop on Online Misinformation- and Harm-Aware Recommender Systems (OHARS 2021), Amsterdam, Netherlands, 2021.
- [3] V. W. Anelli, T. D. Noia, E. D. Sciascio, D. Malitesta, F. A. Merra, Adversarial attacks against visual recommendation: an investigation on the influence of items' popularity, in: Proceedings of the 2nd Workshop on Online Misinformation- and Harm-Aware Recommender Systems (OHARS 2021), Amsterdam, Netherlands, 2021.
- [4] W. Sun, O. Nasraoui, User polarization aware matrix factorization for recommendation systems, in: Proceedings of the 2nd Workshop on Online Misinformation- and Harm-Aware Recommender Systems (OHARS 2021), Amsterdam, Netherlands, 2021.
- [5] M. Badami, O. Nasraoui, PaRIS: Polarization-aware recommender interactive system, in: Proceedings of the 2nd Workshop on Online Misinformation- and Harm-Aware Recommender Systems (OHARS 2021), Amsterdam, Netherlands, 2021.
- [6] M. Mifsud, C. Layfield, J. Azzopardi, J. Abela, “To trust a LIAR”: Does machine learning really classify fine-grained, fake news statements?, in: Proceedings of the 2nd Workshop on Online Misinformation- and Harm-Aware Recommender Systems (OHARS 2021), Amsterdam, Netherlands, 2021.
- [7] R. B. Salem, E. Aïmeur, H. Hage, The privacy versus disclosure appetite dilemma: Mitigation by recommendation, in: Proceedings of the 2nd Workshop on Online Misinformation- and Harm-Aware Recommender Systems (OHARS 2021), Amsterdam, Netherlands, 2021.