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

Antonela Tommasel[a], Daniela Godoy[a] and Arkaitz Zubiaga[b]

[a]*ISISTAN Research Institute (CONICET/UNCPBA), Tandil, Bs. As., Argentina*
[b]*Queen Mary University of London, London, UK*

Abstract

This volume contains the papers presented at the Workshop on Online Misinformation-and Harm-Aware Recommender Systems (OHARS'2020) co-located with the 14th ACM Recommender Systems Conference (RecSys'2020). These proceedings describe the specific workshop goals, format and contain the papers that were presented during the online event held on September 25th, 2020.

Keywords

Recommender systems, online harms, misinformation, hate speech

1. Introduction

Misinformation and online harms are not only widespread on social media but also on other platforms as e-commerce sites, and have shown to have serious damaging effects on individuals and society at large [1]. Online harms include the distribution of false and misleading information (such as hoaxes, conspiracy theories and fake news), harmful content (such as abusive or offensive comments) and the augmentation of societal biases and inequalities online, among others.

Recommender systems play a central role in the process of online information consumption as they leverage massive user-generated content to assist users in finding relevant information as well as social contacts. Thus, they are both affected by the proliferation of low-quality content in social media, which hinders their capacity of achieving accurate predictions and, at the same time, become unintended means for the amplification and spread of online harms.

In addition, in their attempt to deliver relevant and engaging suggestions about items, recommendation algorithms might introduce biases [2], and further foster phenomena as filter bubbles and echo chambers. Biases in data, algorithm design, evaluation, and interaction [3] limit the exposure of users to diverse points of view and make them more vulnerable to manipulation by disinformation [4].

OHARS 2020 was the first edition of the Workshop on Online Misinformation- and Harm-

Aware Recommender Systems[1]. The aim of this workshop was to bring together researchers in the recommender systems community interested in tackling online harms and mitigating their impact on recommendation and, thus, to facilitate discussion about the major challenges and opportunities that will shape future research in the area.

2. Accepted Papers

Five papers were accepted for presentation in the workshop, covering a broad range of technical aspects related to harm-aware recommender systems, and the detection of harmful content and its credibility.

The position paper by Fernandez and Bellogín [5] describes the key challenges behind assessing and measuring the effect of existing recommendation algorithms on the recommendation of misinforming articles and translating successful misinformation management strategies from social science research into computational recommendation models. The authors present their vision on how to address these problems based on four research dimensions: (1) misinformation: problem dimensions, aims to understand what are the different dimensions of the misinformation problem, and within them, the aspects that may affect the behaviour of recommendation algorithms; (2) analysis of recommendation algorithms, refers to the need of in-depth investigations of the internal mechanisms of existing recommendation algorithms that favours the spread of misinformation; (3) human-centred evaluation, refers to the need of modifying existing evaluation methods and metrics to appropriately deal with misinformation; and (4) adaptation, modification and vigilance, refers to the investigation of how recommendation algorithms could be modified and adapted to counter their misinformation recommendation behaviour.

Ghanem et al. [6] address the detection of online users spreading hate, fake, and deceptive online messages, usually called trolls. Text-based features, both affective and lexical ones, coupled with topic modeling were used in this work to enhance the performance of detection models under the assumption that they may characterize the trolls' language changes across topics. The study focus the experimentation on IRA (Internet Research Agency) trolls, originated in Russia for affecting the results of the US 2016 presidential elections, and analyzes the use of NLI (Native Language Identification) features to identify IRA trolls from their writing style.

The identification of malicious bots that degrade the performance of personalization and recommendation algorithms on e-commerce sites with machine learning models is addressed by Sinha et al. [7]. The authors propose two modifications to Positive-Unlabeled learning (PU learning) framework to handle problems where the random sampling assumption is violated. Simulation studies were conducted for validating the approaches and, then, an application of the most general model proposed based on a large real-word dataset with traffic logs of an e-commerce website was described.

A platform to mitigate the dangers of self-disclosure through the use of a personalized harm-aware recommender system is proposed by Salem and Hage [8]. The recommendation algorithm evaluates the risks of disclosing personal information and, if necessary, suggest users ways to reduce such risks. As users are guided towards privacy preservation through nudges tailored

---

[1]https://ohars-recsys2020.isistan.unicen.edu.ar/

according to their perception, the evaluation of the approach shows how users' choices impact the personalization process.

A common concern in recommendation systems is the filter bubble phenomenon, which occurs when the system filters out information, thus narrowing the user's perspective. This problem is tackled by Gharahighehi and Vens [9] in the context of session-based recommenders, which suggest the next item of interest given a sequence of previous items in the active session. The study proposes three scenarios to make the session-based k nearest neighbor method diversity-aware, and evaluates the scenarios across three different news datasets.

3. Program

OHARS'2020 was a half-day workshop in the context of RecSys'2020. The workshop program included short presentations and talks with the aim of discussing on the different aspects of harm-aware recommender systems practice and experience. The workshop started with an opening keynote by Bárbara Poblete (University of Chile), entitled *"Mining social networks to learn about rumors, hate speech, bias and polarization"* [10], on mining social networks to learn about rumors, hate speech, bias and polarization. The second part of the workshop started with an invited talk by Martha Larson (Radboud University and Delft University of Technology, Netherlands) entitled *"Moderation meets recommendation: Perspectives on the role of policies in harm-aware recommender ecosystems"* [11], on ways in which recommender systems can use algorithms to more closely connect with moderation policies, allowing for better oversight of system outputs and behavior. Authors of accepted submissions were invited to give a presentations followed by some time for QA 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.

- Esma Aïmeur, Université de Montréal
- Ludovico Boratto, Eurecat
- Ivan Cantador, Universidad Autónoma de Madrid
- Giovanni Luca Ciampaglia, University of South Florida
- Leon Derczynski, University of Copenhagen
- Dagmar Gromann, University of Vienna
- Elena Kochkina, University of Warwick
- Michal Kompan, Slovak University of Technology
- Ana Maguitman, Universidad Nacional del Sur, Argentina
- Lara Quijano Sanchez, Universidad Autónoma de Madrid
- Ravi Shekhar, Queen Mary University of London
- Damiano Spina, RMIT University
- Christoph Trattner, University of Bergen, Norway

- Adam Tsakalidis, Queen Mary University of London
- Marco Viviani, University of Milano-Bicocca
- Marcos Zampieri, Rochester Institute of Technology

5. Acknowledgments

References

[1] C. Shao, G. Ciampaglia, O. Varol, A. Flammini, F. Menczer, The spread of fake news by social bots, CoRR abs/1707.07592 (2017). `arXiv:1707.07592`.

[2] D. Nikolov, M. Lalmas, A. Flammini, F. Menczer, Quantifying biases in online information exposure, Journal of the Association for Information Science and Technology 70 (2019) 218–229.

[3] R. Baeza-Yates, Bias in search and recommender systems, in: Proceedings of the 14th ACM Conference on Recommender Systems (RecSys '20), ACM, Virtual Event, Brazil, 2020, p. 2. doi:`10.1145/3383313.3418435`.

[4] F. Menczer, 4 reasons why social media make us vulnerable to manipulation, in: Proceedings of the 14th ACM Conference on Recommender Systems (RecSys '20), ACM, Virtual Event, Brazil, 2020, p. 1. doi:`10.1145/3383313.3418434`.

[5] M. Fernandez, A. Bellogín, Recommender systems and misinformation: The problem or the solution?, in: Proceedings of the Workshop on Online Misinformation- and Harm-Aware Recommender Systems (OHARS 2020), Virtual Event, Brazil, 2020.

[6] B. Ghanem, D. Buscaldi, P. Rosso, TexTrolls: Identifying trolls on twitter with textual and affective features, in: Proceedings of the Workshop on Online Misinformation- and Harm-Aware Recommender Systems (OHARS 2020), Virtual Event, Brazil, 2020.

[7] S. D. R. Sinha, L. Kumari, M. Savova, Botcha: Detecting malicious non-human traffic in the wild, in: Proceedings of the Workshop on Online Misinformation- and Harm-Aware Recommender Systems (OHARS 2020), Virtual Event, Brazil, 2020.

[8] R. B. Salem, E. A. H. Hage, A nudge-based recommender system towards responsible online socializing, in: Proceedings of the Workshop on Online Misinformation- and Harm-Aware Recommender Systems (OHARS 2020), Virtual Event, Brazil, 2020.

[9] A. Gharahighehi, C. Vens, Making session-based news recommenders diversity-aware, in: Proceedings of the Workshop on Online Misinformation- and Harm-Aware Recommender Systems (OHARS 2020), Virtual Event, Brazil, 2020.

[10] B. Poblete, Mining social networks to learn about rumors, hate speech, bias and polarization - Abstract, in: Proceedings of the Workshop on Online Misinformation- and Harm-Aware Recommender Systems (OHARS 2020), Virtual Event, Brazil, 2020.

[11] M. Larson, Moderation meets recommendation: Perspectives on the role of policies in harm-aware recommender ecosystems - Abstract, in: Proceedings of the Workshop on Online Misinformation- and Harm-Aware Recommender Systems (OHARS 2020), Virtual Event, Brazil, 2020.