Preface to the special issue on fair, accountable, and transparent recommender systems

Robin Burke¹ · Michael D. Ekstrand² · Nava Tintarev³ · Julita Vassileva⁴

Published online: 24 July 2021

© The Author(s), under exclusive licence to Springer Nature B.V. 2021

1 Introduction

Personalization has become a ubiquitous and essential part of systems that help users find relevant information in today's highly complex information-rich online environments. Recommender systems are a key enabling technology that allows intelligent systems to learn from users and adapt their output to users' needs and preferences. However, there has been a growing recognition that these underlying technologies raise novel ethical, legal, and policy challenges. It has become apparent that a single-minded focus on the user preferences has obscured other important and beneficial outcomes such systems must be able to deliver. System properties such as fairness, transparency, balance, openness to diversity, and other social welfare considerations are not captured by typical metrics based on which data-driven personalized models are optimized. Indeed, widely used personalization systems in such popular sites such as Facebook, Google News, and YouTube have been heavily criticized for personalizing information delivery too heavily at the cost of these other objectives.

This special issue therefore addresses research on responsible design, maintenance, evaluation, and study of recommender systems. It is a venue for work that has evolved

✉ Robin Burke
robin.burke@colorado.edu

Michael D. Ekstrand
michaelekstrand@boisestate.edu

Nava Tintarev
n.tintarev@maastrichtuniversity.nl

Julita Vassileva
jiv@cs.usask.ca

¹ Department of Information Science, University of Colorado, Boulder, CO 80309, USA

² Department of Computer Science, People and Information Research Team, Boise State University, Boise, ID 83725, USA

³ Department of Data Science and Knowledge Engineering, Maastricht University, Maastricht 6229 EN, The Netherlands

⁴ Department of Computer Science, University of Saskatchewan, Saskatoon, SK S7N 5C9, Canada

out of recent workshops and conferences (e.g., FairUMAP, FATRec, FATML, FAccT) on fair, accountable, and transparent (FAccT) recommender systems. In particular, it addresses what it means for a recommender system to be responsible, and how to assess the social and human impact of recommender systems. We solicited manuscripts addressing questions in each of these areas:

- *Fairness* what might “fairness” mean in the context of recommendation? How could a recommender be unfair, and how could we measure such unfairness?
- *Accountability* to whom, and under what standard, should a recommender system be accountable? How can or should it and its operators be held accountable? What harms should such accountability be designed to prevent?
- *Transparency* what is the value of transparency in recommendation, and how might it be achieved? How might it trade off with other important concerns?

2 Accepted articles

Interestingly, the manuscripts that authors submitted concentrated on the areas of fairness and transparency, but accountability was notably missing. More about this issue below. In the end, the special issue contains seven accepted articles, with four articles on transparency and explanations for recommendations and three on fairness.

2.1 Transparency and explanations

The authors in this issue have approached explanation from a number of perspectives. In “The Effects of Controllability and Explainability in a Social Recommender System” Tsai and Brusilovsky link explanation closely to a user interface that allows users to control aspects of the recommendation process and, using structural equation modeling, are able to study how different factors moderate the effect of explanations and control on user experience. The controllability as transparency aspect of recommendation is also explored by Zheng and Toribio in “The Role of Transparency in Multi-Stakeholder Educational Recommendations.” In this paper, the authors extend the challenge of explanation to multi-stakeholder recommendation environments and show that transparency about the goals of other stakeholders can lead to an acceptance of the reciprocal aspects of the interaction. As the authors note, there has been very little study of transparency and explanation in multi-stakeholder contexts, and a great deal more work is needed to extend these findings in other application settings.

A very different approach to explanation is found in the paper by Musto et al. “Generating Post-Hoc Review-based Natural Language Justifications for Recommender Systems.” The authors here take up the challenge of developing *justifications* for recommendations, treating the recommender system as a black box whose inner workings are not available for inspection either by the user or by the explanation module. This research relies on user reviews as a source for textual elements to be repurposed into these explanations. As earlier research in recommender systems explanation found, users often find these justifications acceptable despite their disconnection from the underlying recommendation processes. The paper raises interesting questions regard-

ing the accountability of recommender systems which use explanations generated specifically to persuade the user in the virtues of the recommendation without revealing how the recommendation was generated.

The problem of explanations for reciprocal recommendation is addressed in “Supporting Users in Finding Successful Matches in Reciprocal Recommender Systems” by Kleinerman et al. Looking specifically at online dating, the authors develop a reciprocal recommendation algorithm and an associated explanation scheme that incorporates the viewpoint of both the user receiving recommendations and the individual being recommended as a potential partner. The system is evaluated in various user experiments including a test involving users of a dating app, and the authors find some gendered differences in the usage of the explanations.

2.2 Fairness

The papers on fairness are diverse in their subject matter, covering both consumer-side fairness (fairness concerns relative to the receivers of recommendations) and provider-side fairness (concerns having to do with the creators or others providing items to be recommended). The paper from Ekstrand et al. “Exploring Author Gender in Book Rating and Recommendation” expands on the authors’ original 2018 case study of provider unfairness in recommendation delivery. They show that some recommendation algorithms fail to deliver recommendations the match users’ level of interest in female authors, thereby denying these authors opportunities to reach their most preferred audience. The authors are careful to make their methodology accessible and reusable for the study of other fairness concerns.

An examination of the fairness-aware recommendation and ranking literature shows an extensive variety of different ways of conceptualizing and quantifying recommender system fairness. Deljoo et al.’s “A Flexible Framework for Evaluating User and Item Fairness in Recommender Systems” proposes a general class of fairness measures based on generalized cross-entropy, which unifies a wide variety of metrics that might be applied in consumer-side and provider-side group fairness problems.

While it is often noted that unfairness in recommendation outcomes can be addressed at various points in the implementation pipeline, it is not common to see solutions of different types compared and synthesized. Boratto, Fenu, and Marras examine a type of pre-processing intervention (upsampling of the protected group) and a type of model-based intervention (regularization) in their paper “Interplay between Upsampling and Regularization for Provider Fairness in Recommender Systems.” They show that the combination of these techniques can be more beneficial for provider fairness than each independently.

3 Steps ahead

Social responsibility—including but not limited to fairness, accountability, and transparency—in recommender systems, information retrieval, and user-adapted computing is a young and rapidly expanding topic. The papers in this issue provide a useful

sampling of the kinds of work happening so far in fair, accountable, and transparent recommender systems. Our hope is that they are a valuable starting point for building a significant, impactful body of knowledge over the next years. There is a great deal of important work left to be done, including:

Purposes and target audiences of transparency. Recommender systems have a long history of research on explanations and similar transparency mechanisms and therefore have much to contribute to the broader discussions of explainable and transparent AI. The work collected in this special issue makes useful contributions in studying the roles and possibilities of transparency in particular application settings. There is still much to learn, however, about how to design and evaluate transparency mechanisms for different stakeholders to advance particular purposes.

Accountable recommendation and personalization. Of the three “pillars” of FAccT, accountability is relatively under-studied, particularly in recommender systems and other personalization applications, and the submissions collected here reflect that. What does it mean to hold a recommender system accountable? Who should be held accountable (designers, operators, content providers)? To whom should they be held accountable, and how? What properties should a system be accountable for, and how are those enforced? What technical and policy mechanisms can allow for external verification and accountability for socially beneficial properties?

Working out fairness in many contexts. The literature so far has identified a broad taxonomy of fairness problems that can arise in recommendation and related applications. Recent work has also studied fairness in particular contexts and developed metrics that may be useful in a variety of contexts. What remains to be done is working out what fairness means, specifically, in many of these domains and applications, to provide an inductive basis for guidance that practitioners can use when determining appropriate fairness objectives and methods for their specific application.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Robin Burke is a professor of information science at the University of Colorado Bolder. He conducts research in personalized recommender systems, with his most recent projects exploring fairness, accountability and transparency in recommendation through the integration of objectives from diverse stakeholders.

Michael D. Ekstrand is an assistant professor of computer science at Boise State University. He studies how people and intelligent information systems interact, working with the People and Information Research Team (PIReT) to answer a wide array of questions in human-computer interaction, recommender systems, information retrieval, and user-facing artificial intelligence applications. His research seeks to build better recommendation and information access experiences for users by applying psychological insights and algorithmic innovation, understand the ways in which the system and its users affect and change each other, and promote the reproducibility of recommender systems research.

Nava Tintarev is a professor of explainable AI at Maastricht University, and a visiting Professor at TU Delft. She leads and contributes to several projects in the field of human-computer interaction in artificial advice giving systems, such as recommender systems; specifically developing the state-of-the-art for automatically generated explanations (transparency) and explanation interfaces (recourse and control).

Julita Vassileva is a professor of computer science at the University of Saskatchewan, Canada. Her research areas involve human issues in decentralized software environments: user modeling and personalization, designing incentive mechanisms for encouraging participation and facilitating trust in decentralized software applications, such as online communities, social networks, multi-agent systems and peer-to-peer systems.