Responsible Recommendation and Search Systems

Alex Beutel
alexbeutel@google.com
Google
New York, NY, USA

Ed H. Chi
edchi@google.com
Google
Mountain View, CA, USA

Fernando Diaz
diazf@acm.org
Microsoft Research
Montreal, Quebec, CA

Robin Burke
robin.burke@colorado.edu
University of Colorado, Boulder
Boulder, CO, USA

ABSTRACT

Recommendation and search systems play pivotal roles in how users access information, using them to discover news, entertainment, tutorials, housing, and employment, to name a few. As such, these systems influence social processes related to politics, culture, education, health, and economic well-being. The impact and risks across these systems are widely varied, from shaping the information consumed by users to uncertainty about what users want to challenges in simultaneously supporting a breadth of stakeholders. The potential for adverse impacts has resulted in increased attention from multiple stakeholders, including the academic community, policymakers, industry, and civil society.

This tutorial will cover four main topics: (1) content and experience quality, (2) bias and fairness, (3) diversity and filter bubbles, and (4) ecosystem effects. Each of these topics is complex and may be unfamiliar to many researchers and engineers designing these systems. However, there is a growing body of work in computer science and existing work in related disciplines that can inform the design of information access systems. For each, we will frame the set of concerns within each topic and then survey recent work for both measurement and modeling of these concerns. While each of these topics has been studied independently in the literature, we hope that by presenting them together we can give a more complete picture of how they interact and come together in recommendation and search system design and evaluation.

ACM Reference Format:
Alex Beutel, Ed H. Chi, Fernando Diaz, and Robin Burke. 2020. Responsible Recommendation and Search Systems. In . ACM, New York, NY, USA, 3 pages.

1 TOPIC

In this tutorial we will cover four main topics covering a wide variety of effects and issues that arise in the design and use of recommendation and search systems. We give a high-level characterization of each of these topics below.

Quality & Objectives. One central goal of recommendation and search systems is to provide users with a good user experience. However, measuring and understanding what is a "good" user experience turns out to be quite difficult, and getting this wrong can even be harmful. A large amount of work has focused on predicting what will a user click on [31], but this is significantly incomplete information. For example, "clickbait" may grab users' attention and receive lots of clicks while not offering a positive user experience [44]. A myriad of work has explored how to detect low quality items [44], discover more informative signals of user engagement [7, 36, 53, 57], as well as more general forms of user satisfaction [26, 27]. More recently, new issues have arisen as misinformation or disinformation may in the short-term be appealing to individual users and thus require new frameworks for evaluating quality beyond engagement [21, 34].

Bias & Fairness. Even while optimizing for a positive user experience, it is important to consider the distribution of items and item producers that the system surfaces. For example, it would be concerning if a book recommender system had a bias for male authors [23]. A significant amount of work has studied how to measure and correct for biases in ranking systems, such as off-policy biases, i.e. we only have data for what we've recommended in the past [47], position bias, i.e. users are disproportionately more likely to click on items that were ranked highly [4, 19, 59], and popularity bias, i.e. more popular items are disproportionately likely to be recommended [3, 15, 22, 32, 50].

This prior work is valuable in highlighting the difficulty in measuring and accounting for the myriad of biases that arise in designing recommendation and search systems, but they also don't directly account for possible fairness concerns that could arise with respect to different types of items or groups of item producers (like female authors relative to male authors, as above). Researchers have studied how societal biases of different kinds creep into systems through the amplification of user input [40, 41]. Because of this, there has been significant recent attention on developing methods for measurement and mitigation for different types of "fairness" concerns. This has ranged from localized accuracy [10, 56], ranking exposure [9, 11, 48, 49], and blindness [60]. Further, similar fairness principles have been extended to studying the satisfaction of groups of users [30, 37].

Diversity, Filter Bubbles, & Polarization. In addition to ensuring that different groups of items rank well, another line of fairness work and responsibility concerns focus on the distribution of items seen by different users. While general diversity goals have been long-studied in recommendation and search [13, 18, 35, 45, 52, 54], a more recent line of work has focused on measuring and achieving diversity to different item groups [5, 16, 28, 48, 51, 51, 55, 58]. From

, ,
2020.

a societal perspective, it may be a nice to have for movie recommendations to cover diverse genres, but is particularly crucial for applications like hiring [28].

Aside from the fairness risks, lack of diversity has been a consistent concern with respect to the diversity of information users consume, and how this evolves over time. In particular, one concern is that of "filter bubbles" [25, 42] where users are only shown or consume content on one side of a controversial topic. Measuring if this effect exists and how to deal with it has been highly debated [6, 39]. Beyond individual filter bubbles is the concern of polarization: do these filter bubbles effect users' perspective [8, 29] and cause divergent consumption patterns across groups of users [20, 46]? As these concerns are strongly tied to understanding user behavior over time, an increasing amount of work is exploring the temporal dynamics of these systems, e.g., how diversity evolves over time [12, 17, 33].

Ecosystem. Production recommendation and search systems often include a variety of stakeholders, including content consumers, content producers, matchmakers, and policymakers. System designers need to consider multiple–sometimes competing–objectives when supporting information access in production. Platforms such as online dating systems and marketplaces such as Airbnb and Uber explicitly consider both consumers and producers as first class users to retain [24, 43]. Even when not explicitly satisfying consumer and producer populations, there may be a social responsibility or implicit objective of balancing the satisfaction of these two groups [38]. As such, the more general problem of multi-sided fairness has developed into a research area [2, 14] strictly more complex than the fairness scenarios considered in classification tasks. We will include in our discussion platform-level objectives such as the health and viability of the marketplace [1].

REFERENCES

[1] H. Abdollahpouri, G. Adomavicius, R. Burke, I. Guy, D. Jannach, T. Kamishima, J. Krasnodebski, and L. Pizzato. Beyond personalization: Research directions in multistakeholder recommendation. *arXiv preprint arXiv:1905.01986*, 2019.

[2] H. Abdollahpouri and R. Burke. Multi-stakeholder recommendation and its connection to multi-sided fairness. *arXiv preprint arXiv:1907.13158*, 2019.

[3] H. Abdollahpouri, M. Mansoury, R. Burke, and B. Mobasher. The impact of popularity bias on fairness and calibration in recommendation. *arXiv preprint arXiv:1910.05755*, 2019.

[4] A. Agarwal, I. Zaitsev, X. Wang, C. Li, M. Najork, and T. Joachims. Estimating position bias without intrusive interventions. *arXiv preprint arXiv:1812.05161*, 2018.

[5] A. Asudeh, H. Jagadish, J. Stoyanovich, and G. Das. Designing fair ranking schemes. In *Proceedings of the 2019 International Conference on Management of Data*, pages 1259–1276, 2019.

[6] E. Bakshy, S. Messing, and L. A. Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132, 2015.

[7] N. Barbieri, F. Silvestri, and M. Lalmas. Improving post-click user engagement on native ads via survival analysis. In *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016*, pages 761–770, 2016.

[8] M. A. Beam. Automating the news: How personalized news recommender system design choices impact news reception. *Communication Research*, 41(8):1019–1041, 2014.

[9] A. Beutel, J. Chen, T. Doshi, H. Qian, L. Wei, Y. Wu, L. Heldt, Z. Zhao, L. Hong, E. H. Chi, and C. Goodrow. Fairness in recommendation ranking through pairwise comparisons. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019.*, pages 2212–2220, 2019.

[10] A. Beutel, E. H. Chi, Z. Cheng, H. Pham, and J. Anderson. Beyond globally optimal: Focused learning for improved recommendations. In *WWW*, pages 203–212, 2017.

[11] A. J. Biega, K. P. Gummadi, and G. Weikum. Equity of attention: Amortizing individual fairness in rankings. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 405–414. ACM, 2018.

[12] D. Bountouridis, J. Harambam, M. Makhortykh, M. Marrero, N. Tintarev, and C. Hauff. Siren: A simulation framework for understanding the effects of recommender systems in online news environments. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 150–159. ACM, 2019.

[13] K. Bradley and B. Smyth. Improving recommendation diversity. In *Proceedings of the Twelfth Irish Conference on Artificial Intelligence and Cognitive Science, Maynooth, Ireland*, pages 85–94. Citeseer, 2001.

[14] R. Burke. Multisided fairness for recommendation. *arXiv preprint arXiv:1707.00093*, 2017.

[15] R. Cañamares and P. Castells. Should i follow the crowd?: A probabilistic analysis of the effectiveness of popularity in recommender systems. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '18, pages 415–424, New York, NY, USA, 2018. ACM.

[16] L. E. Celis, D. Straszak, and N. K. Vishnoi. Ranking with fairness constraints. *arXiv preprint arXiv:1704.06840*, 2017.

[17] A. J. Chaney, B. M. Stewart, and B. E. Engelhardt. How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In *Proceedings of the 12th ACM Conference on Recommender Systems*, pages 224–232. ACM, 2018.

[18] S. Channamsetty and M. D. Ekstrand. Recommender response to diversity and popularity bias in user profiles. In *The Thirtieth International Flairs Conference*, 2017.

[19] N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position-bias models. In *Proceedings of the 2008 international conference on web search and data mining*, pages 87–94. ACM, 2008.

[20] P. Dandekar, A. Goel, and D. T. Lee. Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences*, 110(15):5791–5796, 2013.

[21] J. Donovan and danah boyd. Stop the presses? moving from strategic silence to strategic amplification in a networked media ecosystem. *American Behavioral Scientist*, September 2019.

[22] M. D. Ekstrand, M. Tian, I. M. Azpiazu, J. D. Ekstrand, O. Anuyah, D. McNeill, and M. S. Pera. All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness. In *Conference on Fairness, Accountability and Transparency*, pages 172–186, 2018.

[23] M. D. Ekstrand, M. Tian, M. R. I. Kazi, H. Mehrpouyan, and D. Kluver. Exploring author gender in book rating and recommendation. In *RecSys*, pages 242–250, 2018.

[24] D. S. Evans and R. Schmalensee. *Matchmakers: The New Economics of Multisided Platforms.* Harvard Business Review Press, 2016.

[25] S. Flaxman, S. Goel, and J. M. Rao. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly*, 80(S1):298–320, 2016.

[26] J. Garcia-Gathright, C. Hosey, B. S. Thomas, B. Carterette, and F. Diaz. Mixed methods for evaluating user satisfaction. In *Proceedings of the 12th ACM Conference on Recommender Systems*, pages 541–542. ACM, 2018.

[27] J. Garcia-Gathright, B. St Thomas, C. Hosey, Z. Nazari, and F. Diaz. Understanding and evaluating user satisfaction with music discovery. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 55–64. ACM, 2018.

[28] S. C. Geyik, S. Ambler, and K. Kenthapadi. Fairness-aware ranking in search & recommendation systems with application to linkedin talent search. *arXiv preprint arXiv:1905.01989*, 2019.

[29] M. R. Hasan, A. K. Jha, and Y. Liu. Excessive use of online video streaming services: Impact of recommender system use, psychological factors, and motives. *Computers in Human Behavior*, 80:220–228, 2018.

[30] T. B. Hashimoto, M. Srivastava, H. Namkoong, and P. Liang. Fairness without demographics in repeated loss minimization. In *Proceedings of the 36th International Conference on Machine Learning (ICML 2018)*, 2018. Best Paper Runner-Up Award.

[31] X. He, J. Pan, O. Jin, T. Xu, B. Liu, T. Xu, Y. Shi, A. Atallah, R. Herbrich, S. Bowers, et al. Practical lessons from predicting clicks on ads at facebook. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*, pages 1–9. ACM, 2014.

[32] D. Jannach, L. Lerche, I. Kamehkhosh, and M. Jugovac. What recommenders recommend: An analysis of recommendation biases and possible countermeasures. *User Modeling and User-Adapted Interaction*, 25(5):427–491, Dec. 2015.

[33] R. Jiang, S. Chiappa, T. Lattimore, A. Agyorgy, and P. Kohli. Degenerate feedback loops in recommender systems. *arXiv preprint arXiv:1902.10730*, 2019.

[34] S. Kumar and N. Shah. False information on web and social media: A survey. *arXiv preprint arXiv:1804.08559*, 2018.

[35] M. Kunaver and T. Požrl. Diversity in recommender systems–a survey. *Knowledge-Based Systems*, 123:154–162, 2017.

[36] H. Lamba and N. Shah. Modeling dwell time engagement on visual multimedia. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1104–1113. ACM, 2019.

[37] R. Mehrotra, A. Anderson, F. Diaz, A. Sharma, H. Wallach, and E. Yilmaz. Auditing search engines for differential satisfaction across demographics. In *Proceedings of the 26th international conference on World Wide Web companion*, pages 626–633. International World Wide Web Conferences Steering Committee, 2017.

[38] R. Mehrotra, J. McInerney, H. Bouchard, M. Lalmas, and F. Diaz. Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems. In *CIKM*, pages 2243–2251, 2018.

[39] T. T. Nguyen, P.-M. Hui, F. M. Harper, L. Terveen, and J. A. Konstan. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web*, pages 677–686. ACM, 2014.

[40] S. U. Noble. *Algorithms of oppression: How search engines reinforce racism.* nyu Press, 2018.

[41] C. O'neil. *Weapons of math destruction: How big data increases inequality and threatens democracy.* Broadway Books, 2016.

[42] E. Pariser. *The filter bubble: What the Internet is hiding from you.* Penguin UK, 2011.

[43] L. Pizzato, T. Rej, J. Akehurst, I. Koprinska, K. Yacef, and J. Kay. Recommending people to people: the nature of reciprocal recommenders with a case study in online dating. *User Modeling and User-Adapted Interaction*, 23(5):447–488, 2013.

[44] M. Potthast, S. Köpsel, B. Stein, and M. Hagen. Clickbait detection. In *European Conference on Information Retrieval*, pages 810–817. Springer, 2016.

[45] F. Radlinski, P. N. Bennett, B. Carterette, and T. Joachims. Redundancy, diversity and interdependent document relevance. In *ACM SIGIR Forum*, volume 43, pages 46–52. ACM New York, NY, USA, 2009.

[46] B. Rastegarpanah, K. P. Gummadi, and M. Crovella. Fighting fire with fire: Using antidote data to improve polarization and fairness of recommender systems. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, WSDM '19, pages 231–239, New York, NY, USA, 2019. ACM.

[47] T. Schnabel, A. Swaminathan, and T. Joachims. Unbiased ranking evaluation on a budget. In *Proceedings of the 24th International Conference on World Wide Web*, pages 935–937. ACM, 2015.

[48] A. Singh and T. Joachims. Fairness of exposure in rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pages 2219–2228, 2018.

[49] A. Singh and T. Joachims. Policy Learning for Fairness in Ranking. *arXiv preprint arXiv:1902.04056*, 2019.

[50] H. Steck. Item popularity and recommendation accuracy. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 125–132. ACM, 2011.

[51] J. Stoyanovich, K. Yang, and H. Jagadish. Online set selection with fairness and diversity constraints. In *EDBT*, 2018.

[52] Z. Szlávik, W. Kowalczyk, and M. Schut. Diversity measurement of recommender systems under different user choice models. In *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[53] R. Warlop, A. Lazaric, and J. Mary. Fighting boredom in recommender systems with linear reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 1757–1768, 2018.

[54] M. Wilhelm, A. Ramanathan, A. Bonomo, S. Jain, E. H. Chi, and J. Gillenwater. Practical diversified recommendations on youtube with determinantal point processes. In *CIKM*, pages 2165–2173, 2018.

[55] K. Yang and J. Stoyanovich. Measuring fairness in ranked outputs. In *Proceedings of the 29th International Conference on Scientific and Statistical Database Management*, pages 1–6, 2017.

[56] S. Yao and B. Huang. Beyond parity: Fairness objectives for collaborative filtering. In *Advances in Neural Information Processing Systems*, 2017.

[57] X. Yi, L. Hong, E. Zhong, N. N. Liu, and S. Rajan. Beyond clicks: dwell time for personalization. In *Proceedings of the 8th ACM Conference on Recommender systems*, pages 113–120. ACM, 2014.

[58] M. Zehlike, F. Bonchi, C. Castillo, S. Hajian, M. Megahed, and R. Baeza-Yates. Fa*ir: A fair top-k ranking algorithm. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 1569–1578. ACM, 2017.

[59] Z. Zhao, L. Hong, L. Wei, J. Chen, A. Nath, S. Andrews, A. Kumthekar, M. Sathiamoorthy, X. Yi, and E. H. Chi. Recommending what video to watch next: a multitask ranking system. In *Proceedings of the 13th ACM Conference on Recommender Systems, RecSys 2019, Copenhagen, Denmark, September 16-20, 2019.*, pages 43–51, 2019.

[60] Z. Zhu, X. Hu, and J. Caverlee. Fairness-aware tensor-based recommendation. In *CIKM*, pages 1153–1162, 2018.