

Algorithmic recommendations, anyone?



A recent data competition steers clear from leaderboard chasing and promotes the use of a diverse range of metrics to develop rounded, practical algorithms.

Most of us regularly interact with recommender systems as they are embedded in many of the digital platforms we use daily. Recommender systems offer a way to filter the overwhelming amount of content we are exposed to, such as news, social media posts, and video or music streaming. Designing a good recommender algorithm is a challenge: recommendations should be sufficiently interesting and different from what individuals have already chosen, viewed or bought to keep them engaged, but not so off the mark that users ignore them and lose confidence in the algorithm.

On the flip side, these same benefits can also give rise to challenging ethical issues^{1,2}. Successful algorithms for recommender systems can have a powerful impact on consumption patterns, and users can get trapped in filter bubbles where they are exposed to a narrow range of content. This effect is particularly concerning with regards to exposure to political and societal views via news and social media channels, where polarization, manipulation and the spread of false information is an ongoing risk for democratic discourse³. Other concerns with recommender systems are related to their encroachment upon personal autonomy as users are nudged towards certain choices and entrenchment of societal biases.

To encourage the development of more rounded recommender systems, Tagliabue et al. report in this issue on the

EvalRS recommender systems competition (EvalRS) at the 31st ACM International Conference on Information and Knowledge Management. While data competitions can be a good way to accelerate developments in algorithms or their applications, leaderboard chasing and overfitting can lead to unsuitable solutions. For recommender systems in particular, marginally improving the accuracy in recommending popular items can lead to higher performance, but it does not necessarily lead to a better algorithm for practical purposes. To address such issues, algorithms in this competition were not just scored on performance in terms of item ranking (that is, the predicted quality of a given recommendation), but also on more diverse metrics.

The EvalRS competition invited participants to develop an algorithm for recommending songs to individual users based on a dataset from the music streaming platform Last.fm⁴. The organizers introduced metrics to assess robustness or accurate recommendations over various subgroups of users and songs, as good recommendation systems should perform similarly for different subgroups; such fairness measures are designed to prevent discrimination or systematic disadvantages against members of certain groups⁵. The authors additionally included behavioural tests to emphasize the importance of being ‘less wrong’ – that is, making recommendations more relevant even when predictions are inaccurate, for instance by recommending songs from the same genre as the ground truth. Furthermore, impractical or overly complex solutions were discouraged as participants could only use a fixed compute budget for developing their algorithms.

Overall, the organizers of EvalRS intended to develop a unique data challenge for the machine learning community, by encouraging

participants to build models that were not just accurate according to one metric, but also fair, robust and of practical use for real-world scenarios. Moreover, the competition was designed to be fully open source to ensure reusability of the developed approaches.

Data challenges and competitions can have a substantial impact on science and engineering communities, in particular by accelerating advances in algorithmic approaches and developing new standards such as in sharing data and code. They can also offer a great opportunity for early-career scientists to demonstrate and develop their skills. We introduced the article format of Challenge Accepted in *Nature Machine Intelligence* to highlight these benefits in short, accessible articles that are written by organizers, winners or runners-up. For instance, a recent Challenge Accepted reported on a competition in object detection with aerial vehicles and also combined several metrics, rewarding efficient and practical solutions rather than ones that narrowly focused on achieving high accuracy⁶. We welcome proposals for contributions to this section of the journal highlighting competitions that are similarly forward-looking and stimulating for the community.

Published online: 23 February 2023

References

1. Milano, S., Taddeo, M. & Floridi, L. *AI Soc.* **35**, 957–967 (2020).
2. *Nat. Mach. Intell.* **3**, 1007 (2021).
3. Lorenz-Spreen, P. et al. *Nat. Hum. Behav.* **4**, 1102–1109 (2020).
4. Schedl, M. The LFM-1b dataset for music retrieval and recommendation. In *ICMR '16: Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval* 103–110 (2016).
5. Yang, K. & Stoyanovich, J. Preprint at arXiv <https://doi.org/10.48550/arXiv.1610.08559> (2016).
6. Jia, Z., Xu, X., Hu, J. & Shi, Y. *Mat. Mach. Intell.* **4**, 1265–1266 (2022).