Bernstein Center for Leadership and Ethics Research Series

Fast *and* Ethical: Breaking the Speed Limit on Responsible Content Recommendations

Ethical Implications

– Digital platforms may need to adjust their content recommendations in line with ethical considerations.

– Using classical algorithms for this purpose introduces a substantial time delay, which hurts the experience of platform users who are accustomed to lightning-fast page-load speeds.

– A new, predictive approach enables online platforms to generate recommendations that satisfy ethical constraints without perceptible slowdowns.

Digital media platforms such as Netflix, Facebook, and TikTok are under increasing scrutiny regarding the ethical implications of their personalized content recommendations. To combat bias and avoid skewed content suggestions, sophisticated algorithms can perform additional layers of analysis to ensure that recommendations give space to topics such as racial equity, sexuality, and political persuasion. However, doing this in real time with the conventional algorithmic approach would greatly increase page-load times and create a frustrating user experience. New research affiliated with Sanford C. Bernstein & Co. Center for Leadership and Ethics sets out a new, faster method for applying ethical constraints to produce responsible content recommendations.

In "Scaling up Ranking under Constraints for Live Recommendations by Replacing Optimization with Prediction," Columbia Business School PhD candidate and Bernstein Research Grant recipient Yegor Tkachenko, with co-authors Wassim Dhaouadi (Stanford University) and Kamel Jedidi (Columbia University), develop and test a computational method for faster generation of personalized content rankings that satisfy ethical constraints. The innovative method maintains critical speed performance by predicting in advance the solution to an otherwise time-consuming optimization procedure.

Testing the Application

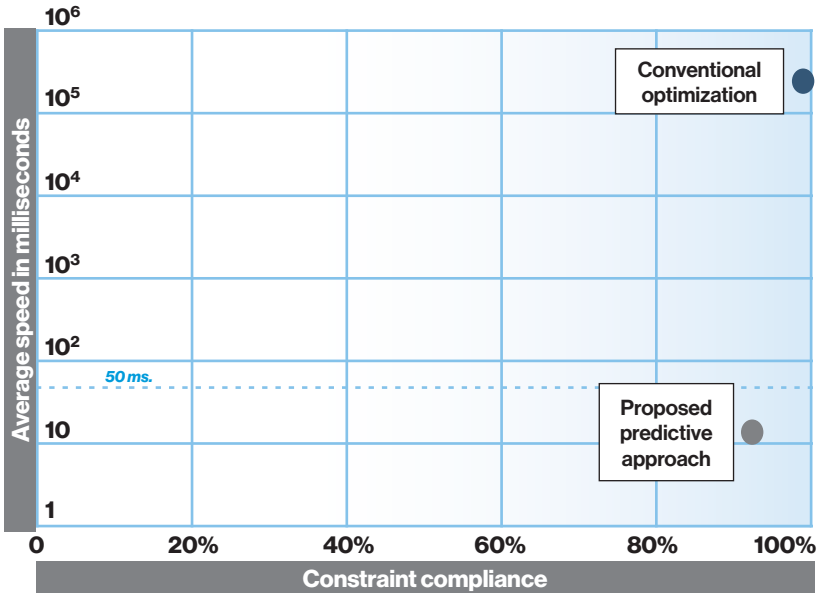
Modern recommendation systems rank hundreds or thousands of content options before displaying a prioritized handful to the limited screen space and attention span of an individual user. Primary business objectives, such as generating more clicks, typically drive ranking schemes, but to avoid unwanted bias and suggest content that reflects a diversity of themes and perspectives, recommendation systems need to perform additional, time-intensive analysis, which takes the form of a mathematical optimization problem.

The challenge is the recommendations need to be generated in under 50 milliseconds to satisfy user expectations for page-load speed. Instead of trying to solve the complex ranking optimization problem exactly, which can take minutes or longer on real-life content ranking problems, Tkachenko's method *predicts* the optimization solution in real time using historical data, thereby allowing user-specific recommendations to be produced within the time limit.

The conventional approach was much slower than 50 milliseconds, whereas the new method performed the ranking under the time limit.

In one experiment that tested and compared the conventional method to Tkachenko's method, the researchers used a database of 25 million ratings generated by 162,000 users at MovieLens.org. The evaluated task was to generate a personalized ranking of 1,000 movies to maximize user satisfaction for genre and recent releases, while also satisfying a set of constraints for the inclusion of a gay character and the mention of race and freedom of speech issues. While both Tkachenko's method and the conventional approach generated recommendations that maximized user satisfaction and complied with the constraints almost perfectly, the conventional approach was much slower than 50 milliseconds, whereas the new method performed the ranking under the time limit. To the best of the researchers' knowledge, this was the first report of content ranking problems of this size being solved live in under 50 milliseconds.

Comparing algorithmic approaches



Conclusions

The application of the prediction-based ranking method developed by Tkachenko and his colleagues is not limited to online content recommendations and ethical considerations. For example, it could also be used to assign time-sensitive tasks to a swarm of robots under resource constraints in warehousing, shipping, or agriculture settings. Ultimately, the value of this approach is the ability to deploy constrained ranking or assignment to new, large-scale problems, where speed matters. The code to reproduce this new methodology is available on GitHub.

Read the full working paper:

Scaling up Ranking under Constraints for Live Recommendations by Replacing Optimization with Prediction

RELATED ARTICLES:

Deadline: "House Democrats Introduce Bill to Remove Some Section 230 Liability Protections For Tech Platforms"

Wall Street Journal: "How TikTok's Algorithm Figures Out Your Deepest Desires"

Yegor Tkachenko
PhD Candidate, Marketing Division
Columbia Business School

Learn more about The Sanford C. Bernstein & Co. Center for Leadership and Ethics