CSCE 580- Artificial Intelligence
Shruti Jadhav
Graduate Paper Presentation Report

When Are Two Lists Better Than One?: Benefits and Harms in Joint Decision-Making
by
Kate Donahue, Sreenivas Gollapudi, Kostas Kollias
The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)

This paper offers a clear and rigorous analysis of a simple and powerful model of human–algorithm collaboration. The authors examine a setting where an algorithm first ranks a set of items and presents its top-k choices to a human, who then selects the final option. Both agents operate under noisy information, and the goal is to determine when collaboration improves the probability of selecting the truly best item. The paper addresses the main question: under what conditions is human–algorithm collaboration genuinely beneficial, rather than harmful?

The work's strongest contribution is its identification of anchoring as a central barrier to effective collaboration. The authors prove that if a human's decision-making is anchored or completely tied to the algorithm's ranking, meaning the human is influenced by the order in which the algorithm presents items, then performance always deteriorates relative to using the algorithm alone. This is a meaningful result due to the examples of anchoring we see in real world such as any automated quiz system suggesting products to a user and user picking solely based on that. This risks unintentionally pushing users toward algorithmic mistakes. The paper gives a strong theoretical base using mathematical proofs for each claim.

The second major contribution is that in the absence of anchoring, human and algorithm performance can become complementary. When human and algorithm rankings are independent and equally accurate, aka the unanchored case, the system performs strictly better at exactly k = 2. This occurs because the algorithm reduces the choice set to a manageable pair, while the human's independent perspective helps correct algorithmic errors. The authors support this with simulations using the Random Utility Model, showing that the results generalize beyond the Mallows model.

Another insight from the paper is the asymmetry in collaboration benefits when humans and algorithms have different accuracy levels. A more accurate human often benefits from teaming with a less accurate algorithm, but the reverse is not true. This asymmetry reflects the fact that the human makes the final decision.

Despite its strengths, the paper also has limitations. First, its modeling of anchoring is overly rigid and a more nuanced human behavioral model could reveal richer dynamics between the two agents. Second, the human is modeled as a simple noisy ranker, which overlooks real cognitive factors such as heuristics, bias, and confidence. Third, the paper provides satisfying proof for k=2. While it is an efficient choice to prove the system works, it is unsure if k=2 is the

optimal choice. More exploration of the k set can provide a holistic view on understanding dynamics of the system.

Overall, this paper provides a valuable theoretical foundation for understanding when human–algorithm collaboration succeeds or fails. Its insights are practically relevant at offering design principles for building systems that truly leverage human expertise rather than unintentionally undermining it.