

Relations between political preferences of users and recommendation algorithms on social medias

Tim FAVERJON^{1,2}

¹Médialab, Sciences Po & ²LPI, Université Paris Cité
`tim.faverjon@sciencespo.fr`

Abstract

Recommender systems in social platforms attract attention in part because of their potential impact over political phenomena, such as polarization or fragmentation of online communities. These research topics are also important because of the need for understanding systemic effects in view of upcoming risk-oriented AI regulation in the EU and the US. A common approach leverages outcomes of recommendations to audit recommender systems. A different approach is that of explainability, seeking to render recommendation mechanisms intelligible to humans, potentially enabling both auditing and actionable design tools. This second approach is particularly challenging in the context of online systems of political opinions because of the intrinsic unobservability of opinions.

In this article we leverage multi-dimensional political opinion estimation of large online populations (along a left-right dimension but also along other political dimensions) to investigate latent spaces in representation learning computed by recommender systems. We train a recommender based on ubiquitous collaborative filtering principles using data on content sharing on Twitter by a large population, evaluating accuracy and extracting a latent space representation leveraged by the recommender. On the other hand, we leverage multi-dimensional political opinion inference to position users in political spaces representing their opinions. We then show for the first time the relation between latent representations leveraged by a recommender system and the spatial representation of users. We show that some dimensions learned by the recommender capture ideological positions of users, bridging politics and algorithmics in our social and algorithmic system, opening a path towards political explainability of AI.

In order to test our method, we leverage other important socio-demographic features such as profession and language of the users and we show that the dimensions identified as political by our method were in fact related only to political information and not to other socio-demographic features. We then compare the factor analysis of the input data with the one of the representation space of the algorithm, and we show that, while professions and languages are the main determinant in the input data, political opinions became a significant determinant in the representations leveraged by the recommender. We argue that this not only proves our method to be reliable in explaining the mechanisms of recommendation with politics, but also opens the possibility to assess the potential responsibility of algorithms in political polarization and fragmentation.¹

¹This work is part of the research project AI political machines and part of the first year of work for my PhD thesis : *The socio-political lives of AI systems - An investigation into how socio-demographic features and political attitudes guide the algorithms in social medias.*