Algorithmic Curation of News and Transparency of News Automation

Heikki Rantala

Automatic generation of news is a thing that can be currently done only in very simple cases. However, perhaps far more importantly, algorithmic curation, the automatic selection, organizing and presenting, of news is already an existing thing that affects in important ways how the public view the news [1]. Popular social media sites, such as Twitter and Facebook, use algorithms to determine what content they show to their individual users and this content is often news. For example a study published in 2010 found that over 85 percent of trending topics on Twitter were news in nature [2].

There is somewhat conflicting research about how well users of social media understand that their news feed is automatically curated. One study from 2015 found that over 60 percent of Facebook users were unaware that their news feed was curated [3]. Other study from the same year but this number to only 8 percent [1]. What can be said is that at least a significant number of social media users do not realize that an algorithm is selecting and presenting their news. Also, those users, that understand that their news feed is determined by an algorithm, have very much differing theories of how the algorithm works [1].

Power to determine the algorithm that decides what news people read can be great power indeed, and so is understanding such algorithms. However algorithms such as those for automatic news generation tend to be very complicated and often secret [4]. There are good reasons for secrecy about news algorithms. Of course newspapers are usually private firms that want to protected their intellectual property. A completely open algorithm might also make manipulation and gaming of system easier. Thus algorithms are often black boxes whose inner workings are a mystery. A secret algorithm still has input and output and if those are known the algorithm can at least in theory be reverse engineered. Reverse engineering is easiest if both input, the source data, and output are known and the input can be freely manipulated. Often however the source data is at least partially secret and can't be controlled. This obviously makes reverse engineering more difficult and may for example require interviewing insiders to find out about the input. Algorithms can also change or have some randomness built in to them, and this can increase the difficulty of reverse engineering the algorithm. Algorithm need not be fully understood to make meaningful statements about it. Statistical analysis can reveal correlations between input and date. This should however be done carefully because correlation does not imply causation.

References:

[1] Rader, E., & Gray, R. (2015). Understanding user beliefs about algorithmic curation in the facebook news feed. In Conference on Human Factors in Computing Systems - Proceedings. (Vol. 2015-April, pp. 173-182). Association for Computing Machinery.

[2] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In Proceedings of the 19th international conference on World wide web, pages 591600, 2010.

[3] Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., ... Sandvig, C. (2015). I always assumed that I wasn't really that close to [her]": Reasoning about invisible algorithms in news feeds. In CHI 2015 - Proceedings of the 33rd Annual CHI Conference on Human Factors in Computing Systems: Crossings. (Vol. 2015-April, pp. 153-162). Association for Computing Machinery.

[4] Diakopoulos, N. (2014). Algorithmic Accountability Journalistic investigation of computational power structures
In Digital Journalism Volume 3, 2015 - Issue 3: Journalism in an Era of Big Data: Cases, Concepts, and Critiques Pages 398-415 | Published online: 07 Nov 2014

More material:

Mapping the field of Algorithmic Journalism
Konstantin Nicholas Dörr

Algorithmic Transparency in the News Media
Nicholas Diakopoulos & Michael Koliska

Enter the Robot Journalist
Users' perceptions of automated content
Christer Clerwall

From Mr. and Mrs. Outlier To Central Tendencies
Computational journalism and crime reporting at the Los Angeles Times