CSC2552: Review 7, Paper 2

Due on March 13

Ashton Anderson

496 words

Thomas Hollis

Paper 2

This paper by Sweeney, a Harvard CS professor, is a *digital observational study* which leverages searches on both Google and Reuters' websites, using Google AdSense, to ascertain racial discrimination in ad delivery. The results of this paper conclude that the null hypothesis, that ad delivery was not discriminatory on Google and Reuters, can be rejected with a p-value of 0.001.

The main weakness of this paper stems from its weak *external validity* which results in significantly reduced *generalisation*. Indeed, no attempt has been made to suggest that the same effects would hold true for other websites not using Google AdSense or for other advertisements than the four investigated. To make matters worse, the name database generation and name searches were all done in parallel in an *ad-hoc* manner where first names were taken from a few studies and last names inferred from "peekyou.com" [1]. This highly unrigorous approach not only results in *sample bias* but also required over one month of Prof. Sweeney's time. This is however a compromise between simplicity, which comes at the cost of *reproducibility* and *scalability*. A straightforward alternative would be to use automated methods or to delegate the work to multiple people to avoid any possibility of cherry-picking accusations. Another notable weakness of this paper is the lack of follow-through into some of the questions raised. Most notably, while Prof. Sweeney received a gift from Google for this study, there was no comment of how the data is handled by them internally to help avoid such ad delivery discrimination. An alternative *partner-with-the-powerful* approach could have helped to answer some of these questions but this would be a compromise with the trustworthiness and perceived bias of the paper.

Conversely, a significant strength of this paper arises from its *transparency* [2]. Since the methodology implemented was very simple, this paper can detail step by step how each piece of data was acquired. This methodology also helps alleviate many ethical concerns behind designing and releasing scripts that enable users to log information about people automatically using only their first or last names. Another strength of this paper is the wide range of different people, devices, times and locations used in this study. While this certainly came with a high time cost burden, such a *custom-made* approach allows the author to rule out many possible *confounds*.

While the implications of this paper are at first glance extremely concerning, a lot more work needs to be done to investigate a technical cause of such bias and to check if it *generalises* to other platforms. While some of this work has been done in a followup paper in [3], the field remains a very active area of research. In addition, my interpretation of these results are similar to those of Prof. Sweeney, but I think more should have been undertaken in this study to pressure a response from the platforms regarding this particular issue of *algorithmic fairness*.

[1] Levitt, S., Dubner, S. (2005) Freakonomics: A rogue economist explores the hidden side of everything. New York: William Morrow.

[2] Salganik, M. J. (2017). Bit By Bit: social research in the digital age. Princeton University Press.

[3] Kay, M., Matuszek, C., & Munson, S. A. (2015). Unequal representation and gender stereotypes in image search results for occupations. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems(pp. 3819-3828). ACM.