Research on racial bias in Twitter's image cropping algorithm

Ruida He (22762872)
Word Count : 1011

Executive summary

In order to improve user's experience on image preview, twitter develop an image cropping algorithm and use neural network to collect user's eye-tracking data. However, The data it collected and the image cropping algorithm cause twitter in awkward situation. Many twitter users found out Twitter is favouring white face over black faces. When a white person and a black person showed in the image at the same time, Twitter's image cropping algorithm tends to show the face of the white person in the image review. Beside, it also works on the animal or even fictional character.

Actually, the reason why twitter "favours" white faces over black faces is related to the data they collected and the algorithm they used. Twitter tried to predict the most interesting part of an image that the user is likely to look at and crop this part of the image roughly as the image preview, so they need to use eye trackers to record the pixels people fixated with their eyes.

Misused data can reinforce the discrimination against black people. Additionally, some researches show the racism against black people is highly related to their psychological and physical health. Concerning what can be done, we provide several options here. Such as taking sensitive attributes of data like gender or race as into the account while collecting data or using data.

Background:

Photo sharing has always been an integral part of the twitter experience. Twitter needs to process millions uploaded images with different formats and sizes and confine them into the preview frames with same size every days.

In order to improve user's experience on their image preview, Twitter developed a neural network to automatically crop photo previews to show their most interesting part, like the face of a person. And this happens to trigger public discussion around racial bias.

How the issue occurred:

The issue began after a twitter user posted a problem that zoom removed the face of his black college when he was using the virtual background. Since then, the other twitter users found out twitter was doing the same thing using its image cropping algorithm.

Some of them posted some experiments on twitter with two long images, one of which shows a white person on top and a black person on bottom (They even conducted the experiment with animals and fictional characters). And they exchange their position in the other image. Both of the image previews only showed the face of the white person, which happened to prove Twitter's image preview was racially biased, in particular by favouring white faces over black faces.

Role of data:

Before we get into the role that data plays in this issue, an explanation of how Twitter's neural network works needs to be clarified firstly.

From the beginning, twitter tried to use face detection to crop the image, but not every image contains faces. Then they developed a better way

to crop, which is to have a neural network focus on the salient image regions. A region with higher visual saliency means a person is more likely to look at it while freely viewing the image.

In order to obtain what people are more interested in, twitter needs to use eye trackers to record the pixels people fixated with their eyes, which is the data twitter used. The eye tracking data can be used to train their neural network and other image cropping algorithms to predict the rough area that people might be interested in.

Due to the biased data set, twitter's image cropping algorithm picked whiter face more often when there are multiple persons, animals or even fictional characters different colour in one image.

Key player and misused data impact

Key players in this issue include The twitter, white and non-white people (such as black or Asian). The issue can be dated back to the collected eye-tracking data. What we need to know is data can't be racist or sexist, but the way it's used would reinforce discrimination.

To a large extent, that data they collected actually reflects unfair social biases that already existed in our society. Twitter's collected data and their algorithm is actually reinforce racism. Based on what Reuben Binns claimed in his article, the way we misused data could lead to discrimination. If the data demonstrates white people would get more attention than black ones, then twitter's image cropping algorithm is more likely to focus more their limited resources on those white people (It's the face of white people that will put into the photo preview in this case).

Furthermore, Some researchs even illustrate the discrimination will have impact on the psychological and physical health of black people. As mentioned in April Thames' s research, due to black's exposure to chronic stress from the discrimination, the functions of brain regions of black people is more likely to be altered, such as hippocampus, that are targeted in brain diseases like Alzheimer’s disease.

Therefore, misused data being applied to image cropping algorithm could cause discrimination against black and even jeopardise black people's health to some extent.

Policy options (what can be done):

Sometime serious consequence (just like the discrimination against black people) will happen if the data or algorithm don't get supervised. Necessary precautions are required. Here are some options we provide to prevent data in neural network being misused.

- Twitter can allow user to customise their own image preview cropping, which can avoid racism in some way.
- The image cropping algorithm still needs to be improved for avoiding unintentional data discrimination.
- Ethic measurement and management can be included as one of the corporate governance, which could be helpful to avoid racism.
- Sensitive attributes, such as gender or race, could be excluded from the factors that determines what area would be cropped.
- Data prediction are often accurate, but they are not flawless. The test result of algorithm in the lab environment should go public and data prediction needs to be continuously tested. Only in this way can users evaluate whether the efforts that made by twitter are enough to ensure algorithms are mitigating racism.

Reference:

A. Hern. "Twitter apologises for 'racist' image-cropping algorithm" . The Guardian. https://www.theguardian.com/technology/2020/sep/21/twitter-apologises-for-racist-image-cropping-algorithm

L. Theis, Z. Wang. "Speedy Neural Networks for Smart Auto-Cropping of Images". Twitter Engineering. https://blog.twitter.com/engineering/en_us/topics/infrastructure/2018/Smart-Auto-Cropping-of-Images.html

R. Binns. "It’s not big data that discriminates – it’s the people that use it". The conversation. https://theconversation.com/its-not-big-data-that-discriminates-its-the-people-that-use-it-55591

A. Thames. "Study: Racism shortens lives and hurts health of blacks by promoting genes that lead to inflammation and illness". The conversation. https://theconversation.com/study-racism-shortens-lives-and-hurts-health-of-blacks-by-promoting-genes-that-lead-to-inflammation-and-illness-122027