Faster Image Rating Using a CAPTCHA Inspired Task

Siyi Fan & Marc Berman
The Department of Psychology, University of Chicago

Introduction

- Machine learning requires a massive dataset.
- The bottleneck to building a supervised image dataset is the high cost of manual labeling.
- Traditionally, manual image ratings have only been carried out on Likert scales rating a single image or ranking scales comparing two images at the same time.
- Image-based CAPTCHA (Completely Automated Public Turing Test to Tell Computers and Humans Apart) requires users to perform an image recognition task, such as: click all of the images that meet a requirement.
- Inspired by the format of CAPTCHA, we propose a new image rating method with participants who view 12 images simultaneously and click on 4 images they prefer most. Once an image gets clicked on, it is assigned 1 point. Otherwise, it is assigned 0 points.
- **We expect that CAPTCHA-style image rating remains a consistency to the Likert scale rating but with higher speed.**

Research Questions

- Will CAPTCHA-style image rating be faster than conventional Likert scale rating in terms of preference?
- Will CAPTCHA-style image ratings positively correlate with the Likert scale ratings in terms of preference?
- Will CAPTCHA-style preference rating be as reliable as Likert scale preference rating?

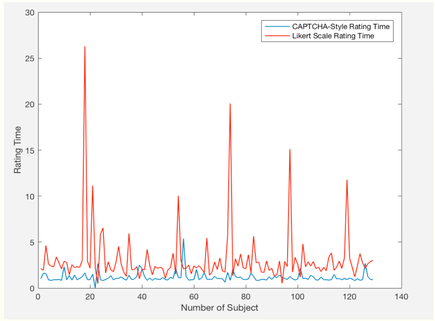
Methods

- Ran simulation to determine the ideal number of subjects (120) to recruit and images to be clicked on (4) during each trial;
- Participants (n=126) were recruited from Amazon Mechanical Turk;
- 1030 scene images in total during 87 trials, including 14 attention check trials;
- 12 different scene images arranged in 3 discrete rows with random item placement (Figure 1);
- Click on the “top 4” images in terms of preference.

Result 1

Rating Time

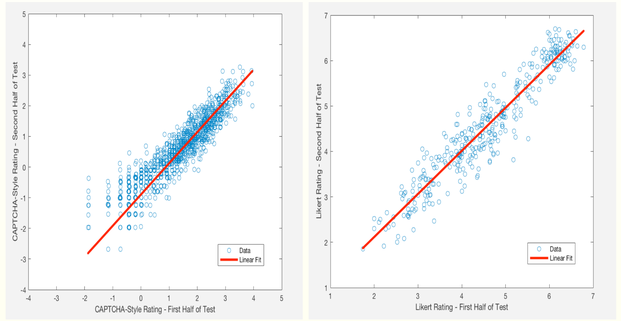
- A paired t-test indicated that CAPTCHA-style rating (M = 1.1705, SD = 0.5445) is faster than Likert scale rating (M = 2.8944, SD = 2.0413; $t(256)$ = -9.2677, $p < 0.05$).



Result 3

Reliability

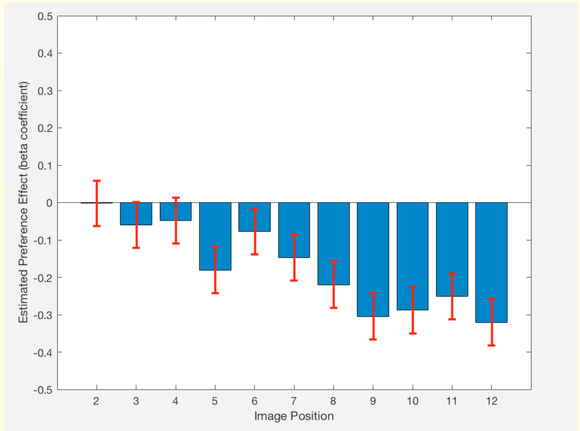
- Split-half reliability indicated that CAPTCHA-style rating (r = 0.9309, 95% CI [0.9237, 0.9374], $p < 0.05$) is comparable to the Likert scale rating (r = 0.9311, 95% CI [0.9189, 0.9416], $p < 0.05$).

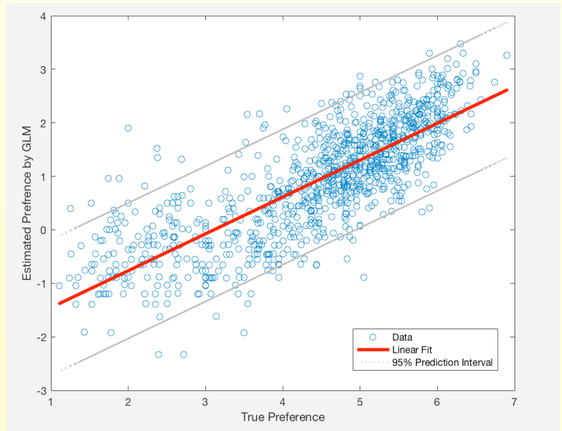


Result 2

Correlation

- Generalized Linear Model (GLM): Click (0 or 1) ~ Image(1, 2, 3,…, 1030) + Image Position (1, 2, 3…, 12)
- Position bias: subjects tended to click on the image displayed on the second row and the second column (position 6) no matter what that image was.
- Correlation: correlation value between the estimated ratings (i.e., beta coefficients) generated by GLM and ratings collected previously by Likert scale was 0.7955 ($p < 0.001$).





Conclusions

- CAPTCHA-style image rating is faster than conventional Likert scale rating in terms of preference.
- CAPTCHA-style image ratings positively correlate with the Likert scale ratings in terms of preference.
- CAPTCHA-style preference rating is as reliable as Likert scale preference rating.
- The probability of an image being clicked depends not only on its preference, but on its position on the screen.
- Together, the CAPTCHA-style image rating method yields perfect consistency to the Likert scale rating**. Therefore, it could be used as an alternative to the Likert scale for image rating.**

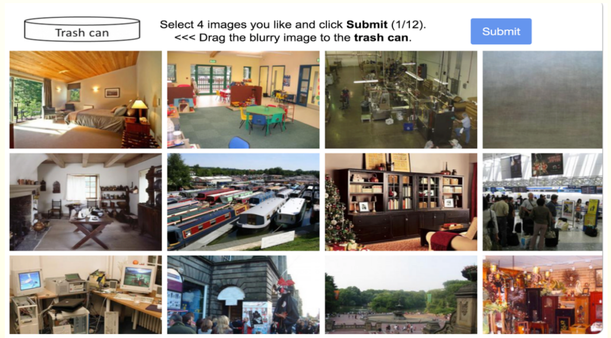


Figure1. ***An example of an attention check trial on preference rating.***

Future Directions

- Future studies may conduct eye-tracking experiments to get a comprehensive record of the regions of the image that are being selected and explore the underlying mechanisms responsible for this tendency. Understanding the nature of this bias will be an important step to study how to model and correct for it.
- Future studies may also examine the CAPTCHA-style image ratings in terms of naturalness, habituality, etc.

Contact

Siyi Fan
T: (217) 979-1560
E: siyifan1@uchicago.edu