9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Predicting when information will be remembered: A machine learning and eye-tracking approach to the study of misinformation

Keywords: misinformation, memory, eye tracking, machine learning, prediction

Extended Abstract

Civic organizations and scholars often use corrective messages to combat the spread of misinformation. However, one important challenge with this approach is that the content of corrective messages can be forgotten, which often leads people to revert to their initial misperceptions [1]. This study addresses this challenge in a unique way by combining eye tracking with machine learning to predict whether an individual will accurately remember information contained in a message. The ability to predict, at the time of message exposure, whether an individual will remember information contained in a corrective message can allow actors engaged in combating misinformation to tailor their messaging strategies for a specific person (e.g., showing the messages multiple times to some people) – increasing the likelihood that the individual remembers information in the corrective message.

We use eye-tracking technology because there is a large body of work linking eye movements to stimuli (e.g., text, images) with memory for those stimuli [3]. Furthermore, eye-tracking technology is ubiquitous and will likely be part of people's everyday lives in the very near future as eye-tracking devices become a standard feature of smartphones, computers, and virtual reality headsets. In this study, our central finding is that a computational model trained on eye-movement data from one group of participants can robustly predict whether a different group of participants will succeed or fail at remembering critical information contained in a message.

Research Design. Participants read short paragraphs about social groups that contained two numbers. The numeric relationship of these numbers was either consistent or inconsistent with many people's beliefs based on our previous studies (i.e., belief-consistent or belief-inconsistent; citation blinded for review). For example, one topic concerned Mexican immigration in the U.S.; given prominent discussion of immigration in U.S. discourse, people may incorrectly think that the number of Mexican immigrants increases every year. Yet, reputable sources identify that Mexican immigration has sometimes decreased. Thus, we showed participants factual information that said there were 12.8 million Mexican immigrants in the United States in 2007 and 11.7 million in 2014. We showed half of the participants factually-accurate, belief-inconsistent information and the other half factually-inaccurate, belief-consistent information (e.g., 11.7 million in 2007 and 12.8 million in 2014). Our critical stimuli consisted of eight paragraphs representing four issues (see Table 1).

We recorded people's eye movements as they read the paragraphs. After reading all the paragraphs, we gave people a recall test. Participants were asked to remember the numerical facts we previously showed them (e.g., "how many millions of Mexican immigrants lived in the United States in 2007?"). Our eye-movement measures consisted of several types of fixations and fixation durations. While reading a passage of text, people's eyes make a series of rapid jumps separated by discrete pauses. The pauses are called fixations, and one of their functions is to place information in the environment, such as a word, within the part of the eye called the fovea, where visual acuity is the highest. Fixation duration corresponds to the amount of time that the fovea is directed at a specific location in the visual environment. Some of our critical eye-movement features include (averaged and summed): first-fixation duration, first-pass fixations, first-pass fixation duration, regression fixations, regression fixations duration, total fixations, and total fixation duration. Researchers have linked these eye-movement responses to stimuli to memory for that stimuli [3].

Modeling Approach. We used a random forest classifier to predict whether participants' recollection of the number contained in the message was correct. Random forests are one of the more transparent approaches [2], allowing researchers to understand the complex feature space in a way that will inform investigations of other applications of eye movements and memory recall tasks. Our approach involved splitting the data by participants (i.e., each of a participant's eight responses will appear together in the same set) into three groups: training ($\sim 60\%$; 66 participants), validation ($\sim 20\%$; 22 participants), and test ($\sim 20\%$; 22 participants). We optimized the hyperparameters of the random forest (e.g., number of trees, tree length, etc.) on the training set with a Bayesian search of the available space using Weights and Biases (see Figure 1). This procedure used the class-average accuracy (*CAA*) of predictions made on the validation set to assess the performance of each combination of hyperparameter values tried (higher is better). Finally, once the hyperparameter values were determined from this process, one prediction was made by the model on the held-out test set.

Results and Conclusion. Our core finding is that by training a machine-learned computational model on eye-movement features *alone* from one set of participants, we were able to accurately predict the memory performance of a different set of participants ($CAA_{val} = 0.76$, $CAA_{test} = 0.71$). The ROC curve visualizes the discrimination ability of the model (see Figure 2). That is, the rates in which the model can correctly classify a remembered item as one that will be remembered (true positive rate) and incorrectly classify a forgotten item as one that will be remembered (false positive rate). An area under the curve analysis, $AUC_{test} = 0.81$ ($AUC = .5$ is chance discrimination), indicates that the model can discriminate between numeric information that people will remember from numeric information that they will forget. This "proof-of-concept" study establishes that we can use a machine-learned model and eye movements to predict, at the individual level, people's memory for the content of messages. This work lays the foundation for future efforts at improving people's ability to combat misinformation via corrective messages.

References

[1] N. M. Brashier, G. Pennycook, A. J. Berinsky, and D. G. Rand. Timing matters when correcting fake news. *Proceedings of the National Academy of Sciences*, 118(5):e2020043118, 2021.

[2] S. K. D'Mello, R. Southwell, and J. Gregg. Machine-learned computational models can enhance the study of text and discourse: A case study using eye tracking to model reading comprehension. *Discourse processes*, 57(5-6):420–440, 2020.

[3] D. E. Hannula, R. R. Althoff, D. E. Warren, L. Riggs, N. J. Cohen, and J. D. Ryan. Worth a glance: using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in human neuroscience*, 4:166, 2010.

Issue Domain	Consistent with Many People's Beliefs	Inconsistent with Many People's Beliefs
Level of support and opposition for same-sex marriage	Same sex marriage is a contested topic among Americans. In a poll conducted by the Pew Research Center, 57% of respondents reported **favoring** same-sex marriage. 39% reported **opposing** same-sex marriage.	Same sex marriage is a contested topic among Americans. In a poll conducted by the Pew Research Center, 39% of respondents reported **favoring** same-sex marriage. 57% reported **opposing** same-sex marriage.
Gender preference for bosses	According to a recent poll, 33% of Americans say that they would prefer working under a **male** boss. 22% of Americans would prefer to work under a **female** boss.	According to a recent poll, 22% of Americans say that they would prefer working under a **male** boss. 33% of Americans would prefer to work under a **female** boss.
Number of immigrants from Mexico in 2007 and 2014	In **2007**, 11.7 million Mexican immigrants lived in the United States. In **2014**, 12.8 million Mexican immigrants lived in the US. Mexican immigrants have been at the center of one of the largest mass migrations in modern history.	In **2007**, 12.8 million Mexican immigrants lived in the United States. In **2014**, 11.7 million Mexican immigrants lived in the US. Mexican immigrants have been at the center of one of the largest mass migrations in modern history.
Number of white and Black individuals shot by police officers in 2016	According to a database compiled by The Washington Post, 963 individuals were shot and killed by police in 2016. Of those shot, 233 individuals were **white** and 465 individuals were **Black**.	According to a database compiled by The Washington Post, 963 individuals were shot and killed by police in 2016. Of those shot, 465 individuals were **white** and 233 individuals were **Black**.

Table 1: Critical stimuli. Based on our previous norming study, we classified paragraphs as either consistent or inconsistent with many people's beliefs. Paragraphs that are shaded in grey are factually accurate; unshaded paragraphs were rewritten to flip the order of the high-low referent relationships. Bolded green and red text are shown here to emphasize the differences between number and reference pairs; all text shown to participants appeared in a uniform black font. Stimuli version pairs were counterbalanced across participants such that an equal number of participants read each version.

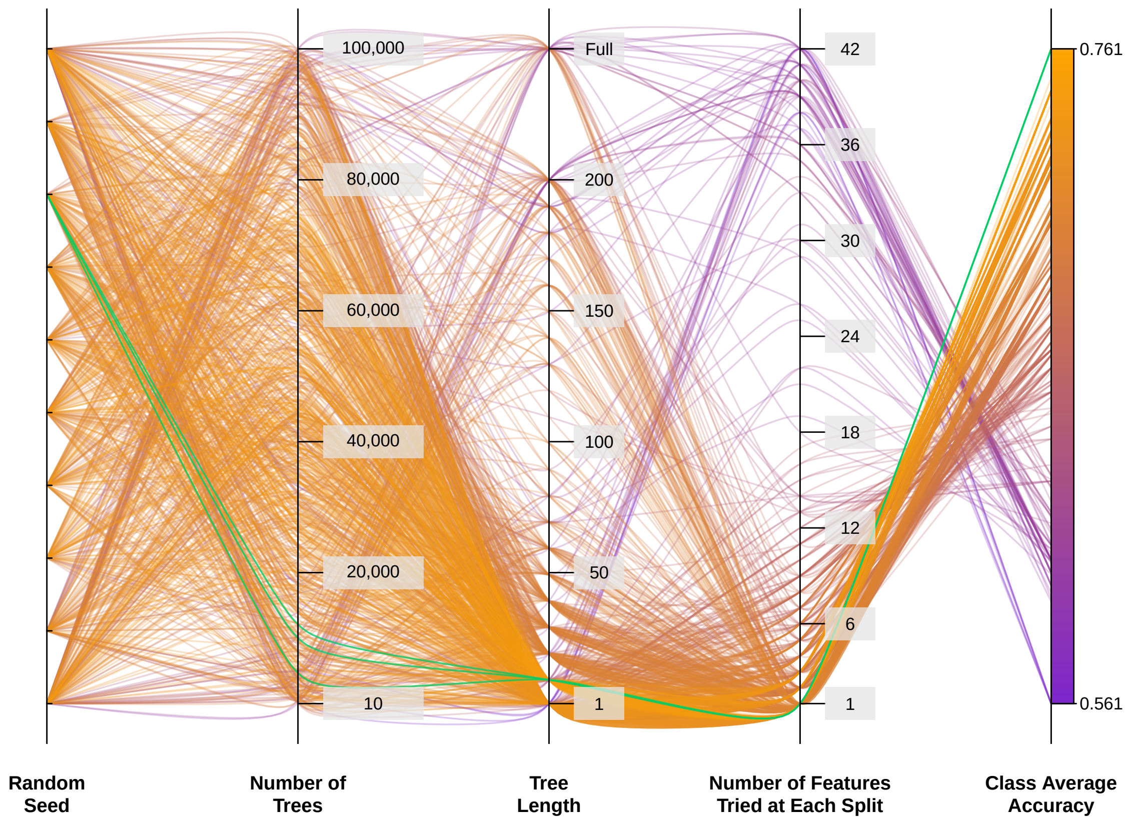

Figure 1: Spline graph of the Bayesian hyperparameter search results. Hyperparameters are ordered along the x-axis from left to right by increasing importance (i.e., relative information gain with respect to CAA). Each spline represents each of the $n = 1000$ sets of random forest parameter values tested. The color of the line corresponds to that run's class average accuracy from low (purple) to high (orange). The three runs which tied for the highest observed CAA score (0.761) are highlighted in green. Weights and Biases controlled the Bayesian sweep of the available space. Possible hyperparameter values include: Random seed – one of ten values; number of trees – uniform integer distribution between 10 and 100,000; tree length – integers between 1 and 200, stepping by 10, and "None", or full-length trees; number of features tried at each split of the tree – uniform integers between 1 and 42 (the full feature set).

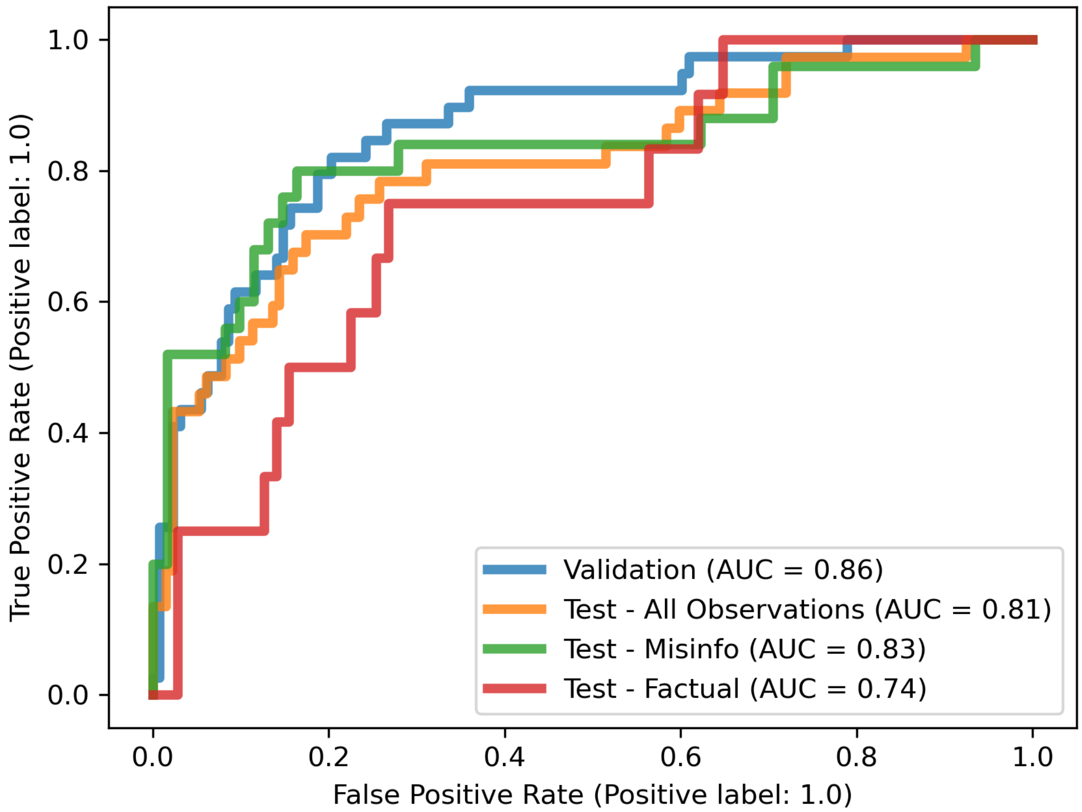

Figure 2: The ROC curve visualizes the discrimination ability of the model. That is, the rates in which the model can correctly classify a remembered item as one that will be remembered (true positive rate) and incorrectly classify a forgotten item as one that will be remembered (false positive rate). A true positive rate that equals the false positive rate suggests no discrimination ability (imagine a 45-degree diagonal line from coordinates 0,0 to 1,1 in the figure). As can be seen from the ROC curves, the true positive rate is greater than the false positive rate, which indicates that the model shows robust discrimination ability in the test set across instances in which the messages convey either factual information or misinformation. An area under the curve (AUC) analysis, $AUC_{test} = 0.81$ for all observations in the test set ($AUC = 0.5$ is chance discrimination), demonstrates that the model can discriminate between numeric information that people will remember from numeric information that they will forget. The fact that the model also displays comparable and robust discrimination ability in both the validation ($AUC_{val} = 0.86$) and test sets suggests that the model is not constrained or over-fitted to the training or validation data (i.e., results generalize well to out-of-sample individuals).