Predicting Gender via Eye Movements

Rishabh Vallabh Varsha Haria, Sahar Mahdie Klim Al Zaidawi, and Sebastian Maneth

Department of Informatics in University of Bremen, Germany

haria@uni-bremen.de, saharmah@uni-bremen.de, maneth@uni-bremen.de

In this paper, we report the first stable results on gender prediction via eye movements. We use a dataset with images of faces as stimuli and with a large number of 370 participants. Stability has two meanings for us: first that we are able to estimate the standard deviation (SD) of a single prediction experiment (it is around 4.1%); this is achieved by varying the number of participants. And second, we are able to provide a mean accuracy with a very low standard error (SEM): our accuracy is 65.2%, and the SEM is 0.80%; this is achieved through many runs of randomly selecting training and test sets for the prediction. Our study shows that two particular classifiers achieve the best accuracies: Random Forests and Logistic Regression. Our results reconfirm previous findings that females are more biased towards the left eyes of the stimuli.

Keywords: Gender Prediction · Eye Movements Data · Machine Learning

1 Introduction

A lot of research has been spawned by the advent of new eye tracking devices. Recordings of eye movements can be used for a variety of purposes; for instance, to detect particular diseases, to identify persons (biometrics), or to study cognitive developments in children. Previous studies on using machine learning classifiers for gender prediction via eye movements have achieved accuracies of 64% [13] and 70% [22], respectively. One drawback of these results, however, is their instability in these two respects: either an unknown or large standard deviation (SD), or a low number of participants. The first study does not mention the SD while the second states an SD of 13.22% and these studies use 52 and 45 users, respectively. Here, we present the first stable results for gender prediction: our employed dataset features 370 users. Our results show a mean accuracy of 65.2% based on 50 runs, with a standard error (SEM) of 0.80%, and we estimate the SD of a single prediction experiment to be around 4.1% by varying the number of participants.

Our method segments eye movement trajectories into fixations and saccades by using the I-VT algorithm [3,8,14]. A small number of statistical features are then computed. We use two separate classifiers of the same kind, one for fixation’s and another for saccade’s. Different ML classification algorithms were examined, and we found that Logistic Regression [12] and Random Forest [15,16] achieved the best accuracies.

* R.Haria and S. Al Zaidawi contributed equally to the work.
performed best. We find that the best accuracy is obtained when we use the Nelder-Mead algorithm [10] to weight the fixation and the saccade classifiers. We observe increased accuracy when fixation weights are higher.

Another highlight of our results is the fact that our best accuracies are obtained by using only two features: the maximum angular velocity (this is applied to fixations) and the saccade ratio (maximum angular velocity divided by the duration; this is applied to saccades). We consider such a small number of features an advantage.

Our main contributions in this paper are:

– We present the first stable results for gender prediction using 50 runs and we achieve an accuracy of 65.2% with SEM of 0.80%.
– We compare our approach to the best previously known accuracy of 70% with an SD of 13.22% [22] and we show that in their setting (45 users and 5 runs) we achieve an accuracy of 77.5% with an SD of only 4.00%.
– We are also able to confirm previous findings that females are biased towards the left eye [17, 21, 28] and have a more explorative gaze behavior when compared to males [2, 3, 22] (the stimuli of our dataset are pictures of human faces) and we show that these differences are statistically significant.

2 Proposed System

This section describes our gender prediction architecture including the employed dataset, data preprocessing and segmentation methods, feature extraction, machine learning classifiers, and the accuracy metric used in this study.

2.1 Dataset

In our study we use the gaze on faces (GOF) dataset. This dataset [20] comprises eye movement recordings of 405 participants aged 18–72 years. There were 27 participants whose data was erratic or absent so they were eliminated. After eliminating the participants the age group changed to 20–72 years. The GOF dataset comprised of 370 users with 185 females and 185 males. The stimuli employed in this dataset are faces of actors. The recordings were produced in 2016 at the Science Museum of London, UK. The eye tracker was an EyeLink 1000 eye-tracker running at 250 Hz. The distance of the seated participants was 57 cm from an LCD screen of 19 inch (1280 × 1024 pixels). The stimuli’s height and width was 429 × 720 pixels. Overall, eight actors were used as stimuli consisting of four females and four males. During the start and end of the video clip, the actor gazed towards the bottom of the screen for half a second.

The minimum recording duration for each participant is two minutes. We capped all recordings to two minutes (a few recordings were slightly longer). During the dataset recording each participant had 32 individual trials. The participants looked at multiple images of a single actor gazing towards them for

\[ \text{GOF dataset can be found here} \]
varying durations (0.1 s to 10.3 s) in all the 32 different trials. This two minute trajectory is the concatenation of all the 32 trails of each participant. We always merge the last point of the first trial with the first point of the next trial. Figure 1 shows an example stimulus and the trajectories of one female and one male participant.

Fig. 1: Example stimulus with gaze trajectories of a female and a male participant.

2.2 Data Processing and Segmentation

There is noise in the raw data. To reduce this, a Savitzky-Golay filter [6, 19] is used with polynomial order of 6 and a frame size of 15. We use the same parameters for the Savitzky-Golay filter as used in [8, 24].

For several prediction tasks, segmentation has been shown to be beneficial [3, 8, 22]. Hence, we segment the eye movement trajectories into fixations and saccades. We use the version of the identification by velocity threshold (I-VT) algorithm used in [8] and [3] to segment eye movements into fixations and saccades. It uses two parameters: the velocity threshold (VT) and the minimum fixation duration (MFD). Any segment of the trajectory for which the velocity is below the VT, and the length is longer than the MFD is classified as a fixation; all other segments are classified as saccades. The standard MFD value of 100 ms is chosen and changing this value (e.g., to 0 ms or 200 ms) has no impact on the accuracy in our study. We choose a VT value of 20 degrees/second for the dataset. By selecting the VT parameter, it is ensured that each participant has a non-zero number of fixations which is around the mean of the maximum number of fixations over all participants. In previous experiments of [14], this way of choosing the VT has been shown to give the best accuracies of certain classification tasks.
2.3 Feature Extraction

In the work of [3], they show that, contrary to previous research, gender prediction is possible in prepubescent children with 64% of accuracy using machine learning. As a result of this promising approach, we calculate the sixty-seven features mentioned in [3].

The work of [27] mentions that a rich amount of information is present in saccades about the dynamics of the oculomotor plant. Hence, we extract the saccade amplitude and saccadic ratio. The temporal properties are leveraged by using the distance and the angle with the previous fixation and saccade as features. Also, global features like duration, dispersion, and path length are computed. The derivatives, such as velocity and acceleration are computed using the “forward difference method”. Earlier work [18] suggests that different areas of the brain generate fixations and saccades in horizontal and vertical directions. Therefore, we generate versions in $x$ and $y$ directions of our features. Various statistical features (M3S2K) mean, median, maximum, SD (standard deviation), skewness, and kurtosis of the trajectories are computed for velocities and accelerations.

We pick the top eight features (see Table 1) in order of their ANOVA scores [7, 26]. We only consider a maximum of eight features since the accuracy started to drop when we increased the number of features beyond eight.

| Fixation features | Saccade Features |
|-------------------|------------------|
| 1) Maximum angular velocity | 1) Saccade ratio |
| 2) SD angular velocity | 2) Mean angular velocity |
| 3) Minimum angular acceleration | 3) Mean angular acceleration |
| 4) Mean angular velocity | 4) SD angular velocity |

2.4 Machine Learning Classifiers and Performance Metrics

According to our study on gender prediction, the following classifiers are the most effective:

– Logistic Regression (LogReg) [12]
– Random Forest (RF) [15, 16]

We also tried Support Vector Machines (SVM) [9], Radial Basis Function Networks (RBFN) [8, 25], and Naïve Bayes (NB) [5], but their results were worse than LogReg and RF. During the experiments, the ratio of male to female is always balanced. We perform hyper-parameter tuning for RF using grid search.

For every experiment, we perform cross-validation using 50 runs. Here, cross-validation refers to selecting training and test sets for the prediction randomly.
for each new run. In our work, we use the ratio of 80:20 meaning we use 80 % participants for training and the remaining 20 % for testing.

Our performance metric is the accuracy. Together with each accuracy, we also report the standard error of the mean (SEM), \( \sigma_{\mu} = \frac{\sigma}{\sqrt{n}} \), where \( \sigma \) is the standard deviation (SD) and \( n \) is the number of runs. This is done using the notation \( x \pm y \% \), where \( x \) is the accuracy (in percentage points) and \( y \) is the SEM. Section 4 is the only section in which we also report the standard deviation (SD).

3 Gender Prediction Experiments

In this section, we describe our experiments for gender prediction. We chose an equal number of fixation and saccade features since we have two classifiers, one for fixations and one for saccades. We perform the experiments using an incremental number of features (see Table 1) from two to eight and show the achieved accuracies with different classifiers in Table 2. We observe that both classifiers peak with two features (one “fixation” and one “saccade” feature). In our study we call these two top features as the “pipeline features”. The LogReg classifier gives us the best accuracy of 63.7 ± 0.60 % followed by the RF classifier with an accuracy of 63.4 ± 0.70 %.

Table 2: Accuracies with SEM using top two to eight features over 50 runs.

| No. of features | LogReg      | RF     |
|-----------------|-------------|--------|
| 2 (1 Sac 1 Fix) | 63.7 ± 0.60 % | 63.4 ± 0.70 % |
| 4 (2 Sac 2 Fix) | 61.8 ± 0.70 % | 62.8 ± 0.80 % |
| 6 (3 Sac 3 Fix) | 61.3 ± 0.70 % | 62.6 ± 0.80 % |
| 8 (4 Sac 4 Fix) | 59.3 ± 0.80 % | 62.8 ± 0.80 % |

3.1 Optimizing Weights for the Fixation and Saccade Classifiers

The Nelder-Mead Method [10] is used to find the optimal weights for the fixation and saccade classifiers. In a survey of black box optimization methods [11] for low-dimensional problems, it outperformed the other methods. Table 3 presents the accuracy and weights for fixations and saccades using equal weights as well as weight-optimized approaches. According to the research of [23], the fixation classifier is more important for gender prediction and they found this by optimizing the weights manually. Our study also revealed the same with the fixation classifier being more important for this task.

The weights increase the accuracy by 1.5 % (see Table 3) for the best performing experiment using two features for the LogReg classifier and by 0.6 % for the RF classifier.
Table 3: Accuracies with SEM using Nelder-Mead method over 50 runs.

| No. of features | LogReg weights | RF weights | RF |
|----------------|---------------|------------|----|
|                | Sac/Fix      |            |    |
| two (1 Sac 1 Fix) | 0.5/0.5      | 63.7 ± 0.60 % | 63.4 ± 0.70 % |
| two (1 Sac 1 Fix) | 0.304/0.696  | 65.2 ± 0.80 % | 64.0 ± 0.70 % |

4 Comparison with State-of-the-Art

Only in this section we use standard deviation (SD) with the accuracy as our metric. To the best of our knowledge, the study of [22] has the highest accuracy which uses machine learning classifiers to predict gender using eye movements. They report an accuracy of 70 % with an SD of 13.22 % using 5 runs. The demographics of the study is as follows: 20 females and 25 males aged 25–34 years. Indoor images are used as stimuli from the Change Blindness database. An Eye-link 1000 plus eye-tracker is used at 1000 Hz.

In order to compare our pipeline features with the state-of-the-art features we perform a similar experiment (SOTA) using the same number of participants and also using 5 runs. We use two types of features for the experiment. We use our pipeline features and the features of [22]. The state-of-the-art features include six features in total. These features are fixation duration, spatial density feature, RFDS (ratio fixation duration to saccade duration), number of saccades, saccade amplitude, and path length. The results are shown in Table 4. For the pipeline features, we achieve the best accuracy of 72.5 % with an SD of 9.35 % using the RF classifier. And using state-of-the-art features we achieve an accuracy of 55.0 % with an SD of 11.00 % with the RF classifier.

Table 4: Accuracies with SEM using pipeline and state-of-the-art features over 5 runs.

| Feature type | No. of features | LogReg         | RF              |
|--------------|----------------|----------------|-----------------|
| pipeline     | two            | 62.5 % with an SD of 13.69 % | 72.5 % with an SD of 9.35 % |
| SOTA         | six            | 52.5 % with an SD of 4.36 %  | 55.0 % with an SD of 11.00 % |

We then optimize the weights for the best accuracy experiment using the Nelder-Mead Method and we find it increases the accuracy by 5 % (see Table 5).
Table 5: Accuracies with SEM using Nelder-Mead method over 5 runs.

| Feature type | No. of features | RF weights | RF       |
|--------------|-----------------|------------|----------|
| pipeline two | 0.5/0.5         | 72.5% with an SD of 9.35% |
| pipeline two | 0.219/0.781     | 77.5% with an SD of 4.00% |

4.1 Relation between SD and Number of Users

Because we are using SD in this subsection, we wanted to know if there is a correlation between SD and the number of users. The Figure 2 illustrates how, by increasing the number of users, the SD approaches a value at around 4.10%. According to our experiments, the standard deviation is lower in an environment with more users, and it must approach a constant value as more users are added.

On the other hand, we found that when the number of runs is increased (200 runs), the SEM approaches a value of 0.31% with accuracy of 63.9% using the LogReg classifier. This means that in 95.4% of repeating our 50 runs experiment, the resulting mean accuracy will be within the range of (63.3%, 64.5%). This is a known statistical fact [4]. We consider these numbers as stable.

Fig. 2: Standard deviation (SD) for LogReg and RF classifier against number of users with two features and over 50 runs.
5 Statistics and Quantitative Observations

In this section we discuss some statistics and quantitative differences between female and male eye movements based on the existing literature \[2,3,22\].

We want to verify whether the statistics of the data as mentioned in \[20\] are the same. They found saccade amplitudes smaller and fixation duration longer for males than for females. With our findings (see Table 6) we confirm that the saccade amplitudes are smaller and total number of fixations is larger for males. We also find that the path lengths are longer for females. This indicates that females are more explorative than males.

Table 6: Mean values for females and males.

| Gender | Path length | Saccade amplitude | Left eye | Right eye |
|--------|-------------|-------------------|----------|-----------|
| Males  | 990.6 ± 1605.1 | 308.2 ± 1567.0  | 66.4 ± 5.2 | 96.2 ± 8.3 |
| Females| 2996.6 ± 14748.8 | 2064.5 ± 14741.0 | 89.5 ± 5.6 | 65.7 ± 7.2 |

The mean number of fixations (in percentage points) in eye region for females and males is shown in Table 6. It can be seen that females have more fixations on the left eye of the stimuli. Males on the other hand make more fixations on the right eye. Previous research of this kind shows very strong left eye bias during the first 250 ms of exploration in women \[21\]. The study of \[28\] found that female participants focused their attention more on the left eye of the stimulus than male participants. It is observed by the work of \[17\] that females have a greater attention to the eye region compared to males. All these previous works back our quantitative findings that women have stronger left eye bias compared to males.

6 Conclusion

In this work, we present the first stable results for gender prediction using eye movements. We trained the data on two classifiers (one fixation and one saccade) and we weighted those classifier probabilities using the Nelder-Mead method. We were able to achieve accuracies of upto 65.2% using 50 runs and a maximum of two features. We also increase the state-of-the-art accuracy by 7.5% using our set of features. Additionally, given the nature of the eye movements, we confirm the left eye bias and explorative behaviour in females in comparison to males.

Acknowledgement

We would like to thank Shubham Ajay Soukhiya for his contribution towards this work.
References

1. Files - uncloud. [https://uncloud.univ-nantes.fr/index.php/s/8KWh6EdyBjqpxmo](https://uncloud.univ-nantes.fr/index.php/s/8KWh6EdyBjqpxmo) (Accessed on 21/28/2020)
2. Li, C., Xue, J., Quan, C., Yue, J., Zhang, C.: Biometric recognition via texture features of eye movement trajectories in a visual searching task. PloS one 13(4), e0194475 (2018)
3. Al Zaidawi, S.M.K., Prinzler, M.H., Schröder, C., Zachmann, G., Maneth, S.: Gender classification of prepubescent children via eye movements with reading stimuli. In: Companion Publication of the 2020 International Conference on Multimodal Interaction. pp. 1–6 (2020)
4. Lee, D.K., In, J., Lee, S.: Standard deviation and standard error of the mean. Korean journal of anesthesiology 68(3), 220 (2015)
5. Bayes, T.: Naive bayes classifier. Article Sources and Contributors pp. 1–9 (1968)
6. Press, W.H., Teukolsky, S.A.: Savitzky-golay smoothing filters. Computers in Physics 4(6), 669–672 (1990)
7. de Souza Jacomini, R., do Nascimento, M.Z., Dantas, R.D., Ramos, R.P.: Comparison of pca and anova for information selection of cc and mlo views in classification of mammograms. In: International Conference on Intelligent Data Engineering and Automated Learning. pp. 117–126. Springer (2012)
8. George, A., Routray, A.: A score level fusion method for eye movement biometrics. Pattern Recognition Letters 82, 207–215 (2016)
9. Cortes, C., Vapnik, V.: Support-vector networks. Machine learning 20(3), 273–297 (1995)
10. Powell, M.J.: An efficient method for finding the minimum of a function of several variables without calculating derivatives. The computer journal 7(2), 155–162 (1964)
11. Hansen, N., Auger, A., Ros, R., Finck, S., Pošík, P.: Comparing results of 31 algorithms from the black-box optimization benchmarking bboB-2009. In: Proceedings of the 12th annual conference companion on Genetic and evolutionary computation. pp. 1689–1696 (2010)
12. Bishop, C.M.: Multiclass logistic regression. In: Pattern recognition and machine learning, chap. 4.3.4, pp. 209–210. Springer (2006)
13. Mercer Moss, F.J., Baddeley, R., Canagarajah, N.: Eye movements to natural images as a function of sex and personality. PLoS One 7(11), e47870 (2012)
14. Mahdie Klim Al Zaidawi, S., Prinzler, M.H., Lührs, J., Maneth, S.: An extensive study of user identification via eye movements across multiple datasets. arXiv e-prints pp. arXiv–2111 (2021)
15. Breiman, L.: Random forests. Machine learning 45(1), 5–32 (2001)
16. Breiman, L.: Bagging predictors. Machine learning 24(2), 123–140 (1996)
17. Sæther, L., Van Belle, W., Laeng, B., Brennen, T., Øvervoll, M.: Anchoring gaze when categorizing faces’ sex: evidence from eye-tracking data. Vision research 49(23), 2870–2880 (2009)
18. Harwood, M.R., Herman, J.P.: Optimally straight and optimally curved saccades. Journal of Neuroscience 28(30), 7455–7457 (2008)
19. Schafer, R.W.: What is a savitzky-golay filter?[lecture notes]. IEEE Signal processing magazine 28(4), 111–117 (2011)
20. Coutrot, A., Binetti, N., Harrison, C., Mareschal, I., Johnston, A.: Face exploration dynamics differentiate men and women. Journal of vision 16(14), 16–16 (2016)
21. Leonards, U., Scott-Samuel, N.E.: Idiosyncratic initiation of saccadic face exploration in humans. Vision research 45(20), 2677–2684 (2005)
22. Sargezeh, B.A., Tavakoli, N., Daliri, M.R.: Gender-based eye movement differences in passive indoor picture viewing: An eye-tracking study. Physiology & behavior 206, 43–50 (2019)
23. Soukhiya, S.: Investigation of improvement of user and gender prediction accuracy via eye-movement by using ensemble classifier and weighting fixations and saccades. (05 2021). [https://doi.org/10.13140/RG.2.2.26461.90080]
24. Schröder, C., Al Zaidawi, S.M.K., Prinzler, M.H., Maneth, S., Zachmann, G.: Robustness of eye movement biometrics against varying stimuli and varying trajectory length. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–7 (2020)
25. Broomhead, D.S., Lowe, D.: Radial basis functions, multi-variable functional interpolation and adaptive networks. Tech. rep., Royal Signals and Radar Establishment Malvern (United Kingdom) (1988)
26. Weerahandi, S.: Anova under unequal error variances. Biometrics pp. 589–599 (1995)
27. Komogortsev, O.V., Jayarathna, S., Aragon, C.R., Mahmoud, M.: Biometric identification via an oculomotor plant mathematical model. In: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications. pp. 57–60 (2010)
28. Sammaknejad, N., Pouretemad, H., Eslahchi, C., Salahirad, A., Alinejad, A.: Gender classification based on eye movements: A processing effect during passive face viewing. Advances in cognitive psychology 13(3), 232 (2017)