Proceedings of Machine Learning Research 1:1–6, 2022 NeurIPS 2022 Gaze Meets ML Workshop

Time-to-Saccade Metrics for Real-World Evaluation

Tim Rolff TIM.ROLFF@UNI-HAMBURG.DE *Universität Hamburg*
Niklas Stein NIKLAS.STEIN@WWU.DE *University of Münster*
Markus Lappe MLAPPE@UNI-MUENSTER.DE *University of Münster*
Frank Steinicke FRANK.STEINICKE@UNI-HAMBURG.DE *Universität Hamburg*
Simone Frintrop SIMONE.FRINTROP@UNI-HAMBURG.DE *Universität Hamburg*

Abstract

In this paper, we explore metrics for the evaluation of time-to-saccade problems. We define a new sampling strategy that takes the temporal nature of gaze data and time-to-saccade problems into account, avoiding samples of the same event in different datasets. This allows us to define novel error metrics for a more intuitive evaluation of predicted durations. The metrics are defined to evaluate the consistency of a predictor and the evaluation of the error over time. We evaluate our method using a state-of-the-art method for time-to-saccade prediction along with an average baseline on three different datasets.

Keywords: time-to-saccade prediction, gaze classification, time-to-event, metrics

1. Extended evaluation

To provide a more detailed evaluation on the proposed metrics of our paper, we also evaluate against five different baseline predictors, namely:

- Zero predictor (zero): Predicts an event for every step by reporting a time-to-event of zero. This should result in high undershot and low overshot rates. Furthermore, it should have a good but not zero consistency.
- Maximum predictor (max): This predicts the maximum time-to-event. This should result in high overshot and low undershot rates. Similar to the zero predictor, it should also have a good consistency. However, we also expect this to perform the worst out of all predictors on the mean square and mean absolute error.
- Random predictor (rand): Predicts a uniformly distributed random event length as the start of a sequence and consistently reduces the time-to-saccade by the update rate of the eye-tracker. Once, we predict a time-to-event below zero, we just report that the event is going to happen every step. We expect this predictor to have an overshot and undershot rate of 0.5 and an excellent consistency, due to its definition.

Using those predictors, we measure how the proposed metrics behave on the DGaze Hu et al. (2020), FixationNet Hu et al. (2021) and EGTEA Gaze+ Li et al. (2018) datasets.

First, Tab. 1 shows the evaluation of the zero predictor, which will report a time-to-event of zero. As expected, it is evident that the predictor undershoots every prediction, which is also shown through the undershot rate. This also results in a high undershot error, as the full sequence undershots the actual target. Second, Tab. 2 shows the evaluation of the random predictor. As predicted, this predictor has a much lower undershot rate and very high consistency. However, it does not reach a 0.5 overshot and undershot rate. This might be due to the uniform sampling that does not reflect the general distribution of the data. At last, Tab. 3 shows the maximum predictor. Here, it is shown that the predictor does not produce any undershots and thus has an excellent undershot error. However, this also results in the highest average time-to-saccade errors, meaning that it does not well in its prediction. Moreover, Fig. 1 and Fig. 2 show the overshot and undershot rate using 10 sections to visualize the behavior of the overshot and undershot over time. As expected, the zero and maximum predictors have the highest over- and undershot rate across the sequence lengths. Whereas, the random predictor is consistently at a 0.6 overshot and 0.4 undershot rate. It can also be infered that the mean and SGD predictors tend to overshoot as the sequence reaches the event.

Table 1: Results of the zero predictor using the metrics described in Sec. 3 of the main paper and the mean square error (mse) and mean absolute error (mae). A lower error is preferred in all cases.

Metric	DGaze	FixationNet	EGTEA
mse↓	0.5168 s^2	0.6695 s^2	0.5168 s^2
mae↓	0.5434 s	0.6408 s	0.5434 s
avg. tts mse↓	0.2089 s^2	0.3088 s^2	0.2000 s^2
avg. tts mae↓	0.4066 s	0.3831 s	0.3733 s
undershot mse↓	0.2089 s^2	0.3088 s^2	0.2000 s^2
undershot mae↓	0.4066 s	0.3831 s	0.3733 s
overshot rate↓	0.0	0.0	0.0
undershot rate↓	1.0	1.0	1.0
consistency↓	1.0	1.0	1.0

Table 2: Results of the random predictor using the metrics described in Sec. 3 of the main paper and the mean square error (mse) and mean absolute error (mae). A lower error is preferred in all cases.

Metric	DGaze	FixationNet	EGTEA
mse↓	0.4585 s^2	0.6805 s^2	0.7973 s^2
mae↓	0.5384 s	0.6526 s	0.7066 s
avg. tts mse↓	0.5097 s^2	0.7877 s^2	0.9966 s^2
avg. tts mae↓	0.5742 s	0.1017 s	0.8003 s
undershot mse↓	0.0653 s^2	0.1017 s^2	0.0575 s^2
undershot mae↓	0.1306 s	0.1620 s	0.1006 s
overshot rate↓	0.62	0.62	0.72
undershot rate↓	0.38	0.38	0.28
consistency↓	0.24	0.25	0.20

References

Zhiming Hu, Sheng Li, Congyi Zhang, Kangrui Yi, Guoping Wang, and Dinesh Manocha. Dgaze: Cnn-based gaze prediction in dynamic scenes. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 26(5):1902–1911, 2020.

Zhiming Hu, Andreas Bulling, Sheng Li, and Guoping Wang. Fixationnet: Forecasting eye fixations in task-oriented virtual environments. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 27(5):2681–2690, 2021.

Yin Li, Miao Liu, and James M Rehg. In the eye of beholder: Joint learning of gaze and actions in first person video. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 619–635, 2018.

Table 3: Results of the maximum predictor using the metrics described in Sec. 3 of the main paper and the mean square error (mse) and mean absolute error (mae). A lower error is preferred in all cases.

Metric	DGaze	FixationNet	EGTEA
mse↓	2.7385 s^2	3.9040 s^2	4.1814 s^2
mae↓	1.6050 s	1.9092 s	1.9899 s
avg. tts mse↓	2.9792 s^2	4.3473 s^2	4.7266 s^2
avg. tts mae↓	1.7134 s	2.0669 s	2.1601 s
undershot mse↓	0.0000 s^2	0.0000 s^2	0.0000 s^2
undershot mae↓	0.0000 s	0.0000 s	0.0000 s
overshot rate↓	1.0	1.0	1.0
undershot rate↓	0.0	0.0	0.0
consistency↓	1.0	1.0	1.0

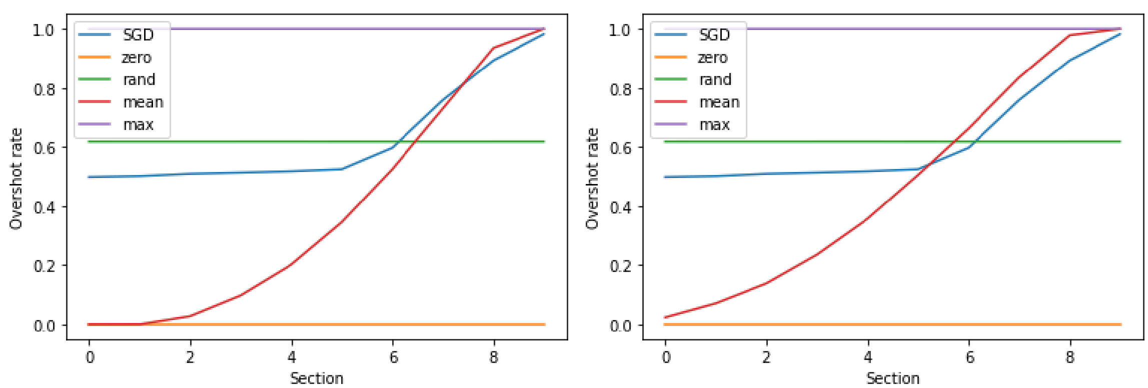

Figure 1: Overshot rate calculated over 10 sections on the DGaze Hu et al. (2020) and FixationNet Hu et al. (2021) datasets.

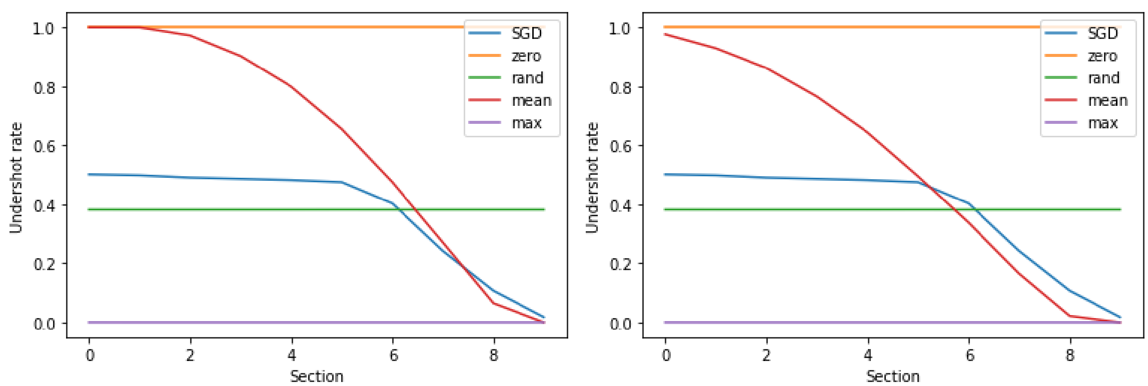

Figure 2: Undershot rate calculated over 10 sections on the DGaze Hu et al. (2020) and FixationNet Hu et al. (2021) datasets.