# Predicting Gaze Patterns: Text Saliency for Integration into Machine Learning Tasks

Ekta Sood
ekta.sood@vis.uni-stuttgart.de

Eye tracking and reading have been a topic of interest in many research communities. In order to obtain more insight into reading comprehension and text saliency, some models have been proposed which predict eye movement during reading tasks; such as the E-Z Reader and SWIFT models (Reichle, Rayner, and Pollatsek, 2003; Engbert et al., 2005). However, these models are rule-based, biased towards the features and the domain. Recent neural network approaches have been deployed for gaze prediction (Wang, Zhao, and Ren, 2019) and modeling human reading (Hahn and Keller, 2016), however there is still much to do in this field as many models fail to accurately predict fixations across various domains and in addition robust evaluation techniques are lacking as gaze data collection is expensive.

The goal of this work would be to advance these approaches for gaze prediction during reading using attention based neural networks. We aim to generate a text saliency prediction model to simulate gaze patterns of humans. We propose two baseline models for fixation prediction, E-Z Reader 10 and a word-level bidirectional long-short-term memory neural network, in which the task of the models is to predict two classes (fixation & skip). The models are trained on the Provo and Geco Corpora (Luke and Christianson, 2018; Cop et al., 2017) and tested on a subset of the MovieQA corpus (Tapaswi et al., 2016) of which we have corresponding gaze data. We aim to evaluate how these two modeling approaches perform (rule-based and neural-based) when tested on an out-of-domain dataset: MovieQA consists of movie plot descriptions and has been used in various machine reading comprehension systems (Blohm et al., 2018; Min, M. J. Seo, and Hajishirzi, 2017; Min, M. Seo, and Hajishirzi, 2017; Xiong, Merity, and Socher, 2016)

The job of our classification systems is to predict fixations/skips on each word of the dataset. To evaluate these systems, we compute accuracy scores over fixated and skipped words in each sentence across each participant, in 21 documents total; the synthesized fixations generated on the MovieQA documents are compared to the human generated fixations and then averaged across participants. That E-Z Reader obtains an accuracy score of 54% and the Bi-LSTM obtains a 63.4%. Further analysis shows that across the 21 documents the percentage of fixated word is: 52.56% of the words in human data, 71.6% in E-Z reader predictions, and 61% in LSTM predictions. In addition, we calculated normalized mutual information over the distribution of fixations from the E-Z reader compared to fixation distribution on the human data. For this metric, we consider the distribution of fixation durations. Here we see that E-Z Reader and humans data share a high overlapping amount of information/distribution of fixations, between 0.6-0.8 (across documents). We hypothesize that while the neural model is better at classifying the two classes, when it comes to saliency of the text E-Z Reader is more similar to humans (when predicting this distribution of fixation duration).

We propose the following future work. We will extend the LSTM to predict a distribution of fixations (i.e. fixation counts on each word), to compare the mutual information score between the humans and neural approach. Subsequently, we will use both models to generate more synthesized fixations and then integrate the synthesized gaze data into the attention mechanism of SOA machine reading comprehension system (Blohm et al., 2018). We aim to investigate if our saliency model can be used for other tasks via exploring gaze assisted attention — to better model human visual attention and hypothesize enhance performance.

## References


[1] Matthias Blohm et al. "Comparing Attention-based Convolutional and Recurrent Neural Networks: Success and Limitations in Machine Reading Comprehension". In: *arXiv preprint arXiv:1808.08744* (2018).

[2] Uschi Cop et al. "Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading". In: *Behavior research methods* 49.2 (2017), pp. 602–615.

[3] Ralf Engbert et al. "SWIFT: a dynamical model of saccade generation during reading." In: *Psychological review* 112.4 (2005), p. 777.

[4] Michael Hahn and Frank Keller. "Modeling human reading with neural attention". In: *arXiv preprint arXiv:1608.05604* (2016).

[5] Steven G Luke and Kiel Christianson. "The Provo Corpus: A large eye-tracking corpus with predictability norms". In: *Behavior research methods* 50.2 (2018), pp. 826–833.

[6] Sewon Min, Min Joon Seo, and Hannaneh Hajishirzi. "Question Answering through Transfer Learning from Large Fine-grained Supervision Data". In: *ACL*. 2017.

[7] Sewon Min, Minjoon Seo, and Hannaneh Hajishirzi. "Domain Adaptation in Question Answering". In: *CoRR* abs/1702.02171 (2017). arXiv: 1702.02171. URL: http://arxiv.org/abs/1702.02171.

[8] Erik D Reichle, Keith Rayner, and Alexander Pollatsek. "The EZ Reader model of eye-movement control in reading: Comparisons to other models". In: *Behavioral and brain sciences* 26.4 (2003), pp. 445–476.

[9] Makarand Tapaswi et al. "Movieqa: Understanding stories in movies through question-answering". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 4631–4640.

[10] Xiaoming Wang, Xinbo Zhao, and Jinchang Ren. "A new type of eye movement model based on recurrent neural networks for simulating the gaze behavior of human reading". In: *Complexity* 2019 (2019).

[11] Caiming Xiong, Stephen Merity, and Richard Socher. "Dynamic memory networks for visual and textual question answering". In: *International conference on machine learning*. 2016, pp. 2397–2406.