Modeling Fixation Behavior in Reading with Character-level Neural Attention

Songpeng Yan[1], Michael Hahn[2], and Frank Keller[3]

[1]School of Informatics, University of Edinburgh, S.Yan-13@sms.ed.ac.uk
[2]Department of Linguistics, Stanford University; SFB 1102, Saarland University, mhahn2@stanford.edu
[3]School of Informatics, University of Edinburgh, keller@inf.ed.ac.uk

Abstract

Humans read text in a sequence of fixations connected by saccades spanning 7–9 characters. While most words are fixated, some are skipped, and sometimes there are reverse saccades. Previous work has explained this behavior in terms of a tradeoff between the accuracy of text comprehension and the efficiency of reading, and modeled this using attention-based neural networks. We extend this line of work by modeling the locations of individual fixations down to the character level. We evaluate our model on an eye-tracking corpus and demonstrate that it reproduces human reading patterns, both quantitatively and qualitatively. It achieves good performance in predicting fixation positions and also captures lexical effects on fixation rate and landing position effects.

Keywords: Computational linguistics; eye-tracking and reading; Cognitive modeling

Introduction

Human readers normally fixate some words and skip others while reading a text, rather than simply reading it word by word. The eyes remain fairly static for 200–250 ms in a fixation, before making a saccade to the next fixation position (Rayner, 1998). A range of models account for many properties of fixation behavior, based on a sophisticated analysis of saccade generation and word recognition (e.g., Reichle et al., 1998; Engbert et al., 2002; Snell and Theeuwes, 2020). Here, we ask whether key properties of fixation behavior can be recovered using a simple rational modeling approach, which derives predictions from an objective function balancing efficiency of attention allocation and accuracy of information extraction. One reason to be interested in such an approach is that it naturally lends itself to accounting for task variation, which is known to substantially modulate reading behavior (e.g., Kaakinen and Hyönä, 2010; Schotter et al., 2014b). Such a modeling approach has been instantiated in the Neural Attention Tradeoff (NEAT) model (Hahn and Keller, 2016, 2018), which accounts for task effects by optimizing reading behavior for task-specific objective functions. This contrasts with most prior models of reading behavior, which focus on the process of word identification. However, the NEAT model so far only accounts for word-level summary statistics (fixations and reading times), but does not model saccades at the level of characters. Here, we propose an extension of NEAT that models character-level fixation decisions.

Our aim is to build a first-principles rational model based on the general assumption that human readers optimize a tradeoff between efficiency and accuracy while reading, maximizing the identifiability of the full input from the observed characters, while minimizing the number of fixations (e.g. Legge et al., 1997; Hahn and Keller, 2016, 2018). We train the model using a large corpus of unannotated English text, and test it on the Dundee eye-tracking corpus (Kennedy, 2003). We evaluate the model on quantitative fit of fixation positions, and reproducing well-documented effects of word length, word frequency, and part-of-speech. Throughout, we compare with the predictions of a widely used prior model of saccade generation with an openly available implementation, E-Z Reader (Reichle et al., 2003).

Our model will make a set of simplifying assumptions. First, we will make the simplifying assumption that the aim of reading is to recognize the words read, even though the modeling framework is compatible with more high-level reading tasks. Second, we will focus on fixation locations, and leave modeling of fixation durations within this framework to future work. Third, we only model forward saccades, whereas a substantial minority of human saccades goes backwards. We discuss prospects for relaxing these assumptions in the Conclusions.

Related Work

A range of computational cognitive models have been developed to simulate various aspects of human reading (Rayner, 2009). E-Z Reader (Reichle et al., 1998, 2003, 2009) and SWIFT (Engbert et al., 2002, 2005) are two representative computational reading models accounting for saccade generation.

E-Z Reader assumes two visual processing stages to determine when and where the eyes move during reading. The first one is called "familiarity check". It takes place during lexical access up to the point when the word can be reliably identified. When the point is reached, the model initiates a saccade to the next word. The second stage of lexical access begins, called "saccade programming". It is divided into two stages: the initial labile stage and the non-labile stage, which depends on whether a saccade can still be canceled or it has become obligatory. Another key assumption of the model is that human readers tend to fixate at the center of the word, but they are also subject to random overshoot and undershoot errors.

SWIFT is a parallel eye-movement reading model in which

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

activation is conceptualized as a gradient. The model spreads activation across several words and computes lexical access for these words at the same time. Furthermore, the time when people move to the next view point in reading is regarded as a random decision, determined by a SWIFT's random timer. For saccade programming, the assumptions are similar to the ones made by E-Z Reader.

Bicknell and Levy (2010) model reading as Bayesian inference on the identity of a sequence. The model generates saccades using a control policy that aims to decrease uncertainty about the identity of the sequence. Our approach makes similar assumptions about the tradeoff between economy of attention and the accuracy of text comprehension. However, our approach differs in a number of ways. First, by using reinforcement learning, our model provides a learning algorithm and is applicable to a range of different reading tasks. Second, our approach draws on modern neural network methods, making it scalable to arbitrary input, whereas Bicknell and Levy (2010) only ran their model on sentences covered by a fixed vocabulary and had to assume a very simple model of language statistics.

Lewis et al. (2013) propose a model of saccade control that combines Bayesian optimization and bounded optimal control. Their model also assumes a speed-accuracy tradeoff which enables eye-movements to adapt to task conditions, and is able to capture eye-movements behavior in a list lexical decision task. Presumably, the Lewis et al. model could be extended to other tasks, including normal reading.

Model

Our model is based on the NEAT model (Hahn and Keller, 2016, 2018), which derives human reading behavior from the *Tradeoff Hypothesis*: the assumption that human readers rationally optimize a tradeoff between successfully extracting information and the economy of attention while reading. NEAT as implemented models the allocation of attention across the words in a text, but does not account for the detailed behavior of skipping and fixation at the level of individual characters. We extend the tradeoff hypothesis to the level of individual fixations at the character level, proposing that readers optimize a tradeoff between successfully reading the text (e.g., recognizing and memorizing the words) and making as few fixations as possible. We follow much prior work on reading (e.g. Legge et al., 1997; Bicknell and Levy, 2010; Hahn and Keller, 2016) in assuming that the aim of reading is simply to recognize the words read, though the modeling framework can be flexibly adapted to other reading tasks (see Conclusions for more discussion).

Model Architecture

We illustrate our model in Figure 1. It is based on a standard neural sequence-to-sequence architecture and consists of three modules, *Reader*, *Decoder*, and *Attention*. *Reader* and *Decoder* are realized as a one-layer Long Short-term Memory neural network with 1,024 memory units. At each time step, the reader model takes a fixed-window character sequence as input and encodes it into a series of hidden vectors. At the end of a text, the activations produced by the reader model are provided to the decoder model, which attempts to reconstruct the whole sequence.

For each fixation, the *Attention* component decides how far to jump in the next saccade based on the information available from a fixed-size window around the current fixation point. Saccade length is assumed to be bounded by the length of the window. Human visual acuity is highest in a small window around the fixation point (the fovea), while further away (in the parafovea), readers only receive partial visual input. In order to account for this fact, only the closest four characters to the right of the fixation point are made fully available to the *Attention* module making the decision; for the other characters to the right of the fixation point and up to the end of the window, the module only receives an encoding distinguishing characters from whitespace. This design is clearly a simplification: visual acuity actually decays more continuously (as accounted for by Bicknell and Levy 2010). On a technical level, the input character sequence is encoded into a series of embedding vectors $v_1, \ldots, v_K$, which are passed through a linear transformation into a softmax function, which outputs a probability distribution over the number n of characters to skip over in the next saccade:

$$P(\omega_i = n|\omega_{1\ldots i-1}, s) = \text{softmax}(\mu + w^T[v_1, \ldots, v_K]) \quad (1)$$

where $\hat{v_i} \in \mathbb{R}^{300}$ is the character embedding of the i-th character and the bias vector μ and matrix w are the parameters of the *Attention* component. Here, we make the simplifying assumption that all saccades go forward (i.e., $n > 0$, see also Conclusions).

Model Objective Function

As in the original NEAT model, we postulate that reading rationally optimizes a tradeoff between comprehension accuracy and economy of reading. In line with much prior work modeling reading (e.g. Legge et al., 1997; Bicknell and Levy, 2010), we assume the comprehension task of recognizing the input. The overall model objective is to minimize the expected loss:

$$Q(\theta) := E_{w,\omega}\left[\text{Loss}(\omega|s,\theta) + \alpha \cdot \frac{\|\omega\|}{N}\right] \quad (2)$$

Here w is the input, ω denotes saccades, $\text{Loss}(\omega|s,\theta)$ is the cross-entropy loss for correctly identifying the words in the input, $\|\omega\|$ is the distance jumped (this models the economy of reading – longer saccades are more efficient), N is the sequence length, α is a parameter trading off the two factors, and θ denotes all parameters of the reader, decoder, and attention networks.

Model Training

We train the *Reader* and *Decoder* components using stochastic gradient descent (Equation 3), while the *Attention* component is trained using reinforcement learning (Williams, 2004)

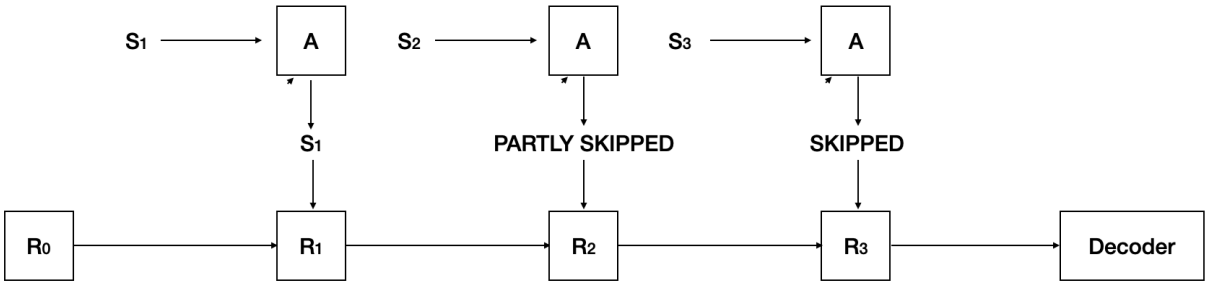

Figure 1: The architecture of our model. The model reads a sequence s_1, s_2, s_3 of words. The *Attention* module (A) decides which parts of different words to skip or fixate on. Fixated parts are fed into the *Reader* module. At the end, the *Decoder* module attempts to reconstruct the input based on the parts of the input that were made available to the *Reader*.

(Equation 4). We calculate the gradients of our network using the backpropagation algorithm and the parameters θ are updated as follows:

$$\Delta_1 := \partial_\theta Loss(\omega|s,\theta) \quad (3)$$

$$\Delta_2 := (\partial_\theta log P_A(\omega|s_{1\ldots N};\theta)) \cdot [Loss(\omega|s,\theta) + \alpha \cdot ||\omega||] \quad (4)$$

$$\theta \leftarrow \theta - \lambda \cdot (\Delta_1 + \Delta_2) \quad (5)$$

where λ stands for the learning rate. We trained our model using the Daily Mail corpus (Hermann et al., 2015), which is composed of 195,462 articles and approximately 200 million tokens from the Daily Mail newspaper. The recurrent neural networks and attention network were each trained for one epoch.

Experiments

We test the effectiveness of our models in simulating human fixations against the Dundee eye-tracking corpus. As a point of comparison, we also report results from E-Z Reader (Reichle et al., 2003), and a simple baseline that randomly selects a fraction of words to fixate and places a fixation at their center ("Random Word + Center Fixation"). We first evaluate the model on fixating location prediction.

Setup

We used the Dundee eye-tracking corpus developed by Kennedy (2003). The corpus was collected using a Dr. Bouis Oculometer eye-tracker (Barrett et al., 2015) in an experiment where 10 native English-speaking participants read newspaper articles from The Independent newspaper. The corpus contains 20 texts with 51,502 tokens across 2,368 sentences in total. We split the corpus into a development partition and test partition, using the former (texts 1–3) for setting model parameters and the latter (texts 4–20) for evaluating the model. For evaluation, we removed the words at the beginning or the end of sequences to avoid incomplete words (and remove return sweeps, which our model is not able to capture).

We first determined the window size and the tradeoff parameter α using random search (Bergstra and Bengio, 2012), varying window size from 5 to 11, and α from -1.0 to 1.0 in steps of 0.01. We chose two sets of parameter values resulting in overall fixation rates on the development set that best matched either the human fixation rate in the Dundee corpus (52.8%), or the fixation rate of E-Z Reader on the same data (77.7%), respectively. Learning rate λ was set at 0.5. The same hyperparameters were used in all experiments, as described in Table 1.

Fixation Rate	Parameter	Value
52.8%	Window size	11
	α	0.12
77.7%	Window size	6
	α	-0.79

Table 1: Parameter values of our model.

Accuracy of Fixation Location Prediction

Figure 2 shows a visualization of the fixation points of one reader on a sample text, as well as the fixation decisions predicted by our model.

Previous studies quantitatively evaluating models for predicting fixations mostly operated at the word level (Nilsson and Nivre, 2009; Matthies and Søgaard, 2013; Hahn and Keller, 2016). They evaluated by measuring the overlap between the fixated words predicted by the models and those in the human eye-tracking data. The design of our model is different, as we assume that human readers operate over sequences of characters while reading.

We therefore use a simple evaluation metric that computes, for each word, the Euclidean distance between each predicted fixation point and the closest human fixation point, normalized by the word length. We calculate this separately for each participant in the Dundee corpus, and average over those.

We compared our character-based version of NEAT to two other models. The first one is a baseline that randomly selects a fraction of words to fixate and places a fixation at the center of each fixated word ("Random Word + Center Fixation", abbreviated "RW+CF"). Second, we compared with E-Z Reader (Reichle et al., 2003), a well-established model of saccade generation with an openly available implementa-

Human (A)	Politicians, under pressure to do something about crime, pass laws and write guidelines requiring more criminals to be imprisoned for more crimes for longer. Every now and then, however, there is a crisis as the walls of our prisons begin to bow under the weight of numbers contained therein.
Model	Politicians, under pressure to do something about crime, pass laws and write guidelines requiring more criminals to be imprisoned for more crimes for longer. Every now and then, however, there is a crisis as the walls of our prisons begin to bow under the weight of numbers contained therein.

Figure 2: Visualization of fixation decisions from human readers and our model. We select as example one participant (A) in the Dundee Corpus.

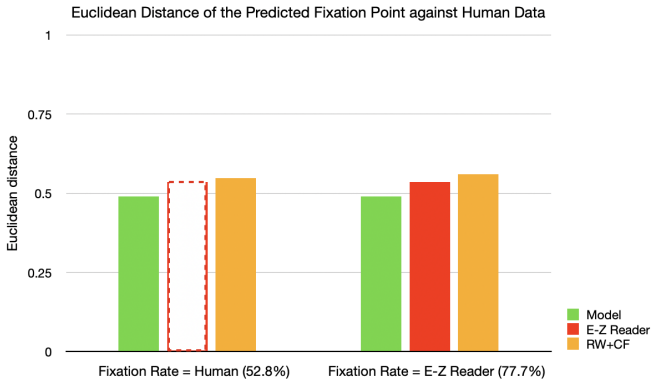

Figure 3: Euclidean distance (in characters) of the predicted fixation point at the character level against human data. Smaller Euclidean distance means a better fit to the correct human fixation point from. The dotted line bar represents E-Z Reader, whose fixation rate is static and cannot be adjusted to match the human fixation rate. RW+CF refers to the baseline "Random Word + Center Fixation", abbreviated "RW+CF". Model means the current model.

tion.[1] We provided E-Z Reader with word frequency and predictability metrics estimated from the same corpus we used for training our model; predictability metrics were estimated using an LSTM language model trained on the same corpus. In order to ensure meaningful comparison, we include a version of our model whose fixation rate matched that of E-Z Reader (77.7% of words).

Figure 3 illustrates that our model shows lower Euclidean distance compared to the baseline Random Word + Center Fixation, and also compared to E-Z Reader. The latter results holds even when matching our model's fixation rate with that of E-Z Reader. This is remarkable, as our model is derived in a way quite different from E-Z Reader: Whereas E-Z Reader is a sophisticated eye-movement model specifically designed for modeling saccade generation, our proposed model is instead derived from general assumptions about a rational tradeoff underlying reading, plus training on large-scale unlabeled text.

[1]We used E-Z Reader 10.2 (Reichle et al., 2012), retrieved from `http://www.erikdreichle.com/downloads.html`, December 20, 2021.

	Human	Model	E-Z Reader	RW+CF
Corr	0.966	0.880	0.886	0.268

Table 2: Correlation between word length and fixation rate.

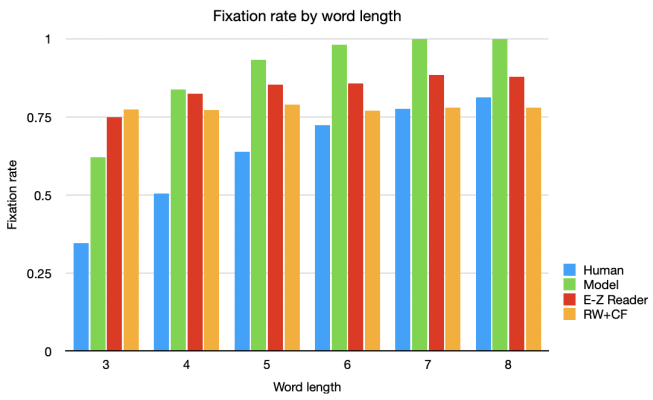

Figure 4: Fixation rate by word length.

	Human	Model	E-Z Reader	RW+CF
Corr	-0.677	-0.729	-0.178	0.431

Table 3: Correlation between log word frequency and fixation rate.

Effect of Word Length, Frequency, and Part of Speech

We furthermore analyzed if our model exhibits general features of reading, specifically the effects of lexical properties on fixation rate. The previous literature observed that lexical properties such as word length, word frequency, and part of speech (PoS) are good predictors of fixation probabilities

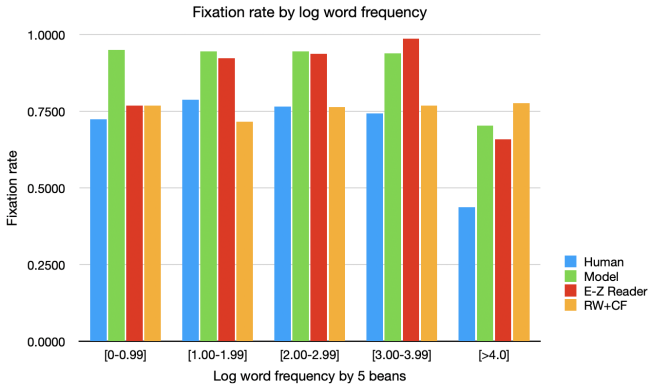

Figure 5: Fixation rate by log word frequency. We binned words into five bins according to their log word frequency.

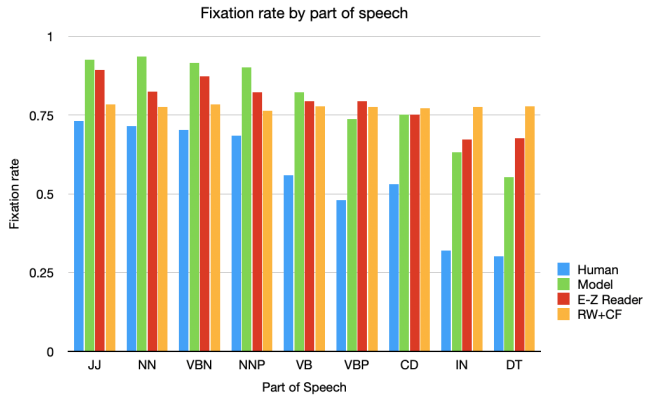

Figure 6: Fixation rate by part of speech.

(Carpenter and Just, 1983; Rayner, 1998; Hahn and Keller, 2016). For instance, human readers will not easily skip a long word because of the limitation of their visual range, and are more likely to fixate on uncommon words to aid word identification.

Fixation Rate by Word Length Figure 4 shows that E-Z Reader and our model both approximately simulate the effect of word length on fixation rate that is observed in human reading behavior (longer words have a higher fixation rate, as shown in Table 2), in contrast to the baseline (Random Word + Center Fixation).

Fixation Rate by Word Frequency We then compared fixation rate to word frequency in the corpus used for training our model; these word frequencies were the ones provided to E-Z Reader. Figure 5 and Table 3 shows that both E-Z Reader and our model reproduce patterns of fixation rate by log word frequency that similar to the ones found in human readers, unlike the baseline. Interestingly, our proposed model shows a stronger correlation between word frequency and fixation rate than E-Z Reader, more in line with the human data.

Fixation Rate by Part of Speech We also calculated fixation rate by part of speech. It has long been documented that content words are more likely to get fixated than function words (Carpenter and Just, 1983). Figure 6 demonstrates that compared to the baseline, our model exhibits the effect of part of speech on the fixation probability of a word, so that a content word receives more fixations than a function word. In conclusion, our model reproduces the lexical effects of word length, word frequency, and part of speech on fixation rates when evaluated against the Dundee corpus.

Landing Position in a Word A range of prior work has studied the typical landing position in words during reading. It found that the landing position can be influenced by word length and by the spaces between words (Blanchard et al., 1984; O'Regan, 1979; Osaka, 1993; Rayner and Morris, 1992). According to Legge et al. (1997), readers tend to locate their first fixation at the center of the word or a bit to the left of the center. In our experiment, we replicate this result for words of length five or more on Dundee eye-tracking corpus, illustrated in Figure 7.

Conclusions

We have described a model of human fixation locations building on the Tradeoff Hypothesis, i.e., the assumption that human reading rationally trades off accuracy of information extraction with economy of reading. Our model is implemented using modern neural network-based machine learning techniques and trained using reinforcement learning with a mathematical formalization of the Tradeoff Hypothesis as the objective function.

Experimental results showed that our model reproduces basic features of human reading patterns both quantitatively and qualitatively, consistently outperforming over a random baseline. Throughout, we compared our model to E-Z Reader, a sophisticated model of saccade generation that includes substantial machinery. Even though our model is based on relatively general architectural assumptions and trained only on unannotated text, it provided a better fit to human fixation positions than E-Z Reader, as measured by the mean Euclidean distance between human fixation position and model predicted fixation positions. Our model also reproduced the relationship between word frequency and fixation rate more convincingly than E-Z Reader.

The proposed model differs from existing models of character-level fixation decisions in two main ways. First, following the NEAT model (Hahn and Keller, 2016, 2018), it aims to derive reading behavior from a general rational objective function. In this respect, it also bears resemblance to the model of Bicknell and Levy (2010), which optimizes reading behavior for word identification, but does not provide a general mechanism that could accommodate other reading tasks, such as proof-reading (Schotter et al., 2014a). Second, unlike prior models of fixations at the character level, our model draws on contemporary neural network modeling and thus scales to broad-coverage applicability on text and can utilize rich knowledge of language statistics. More broadly, our results suggest that neural networks, combined with cognitively

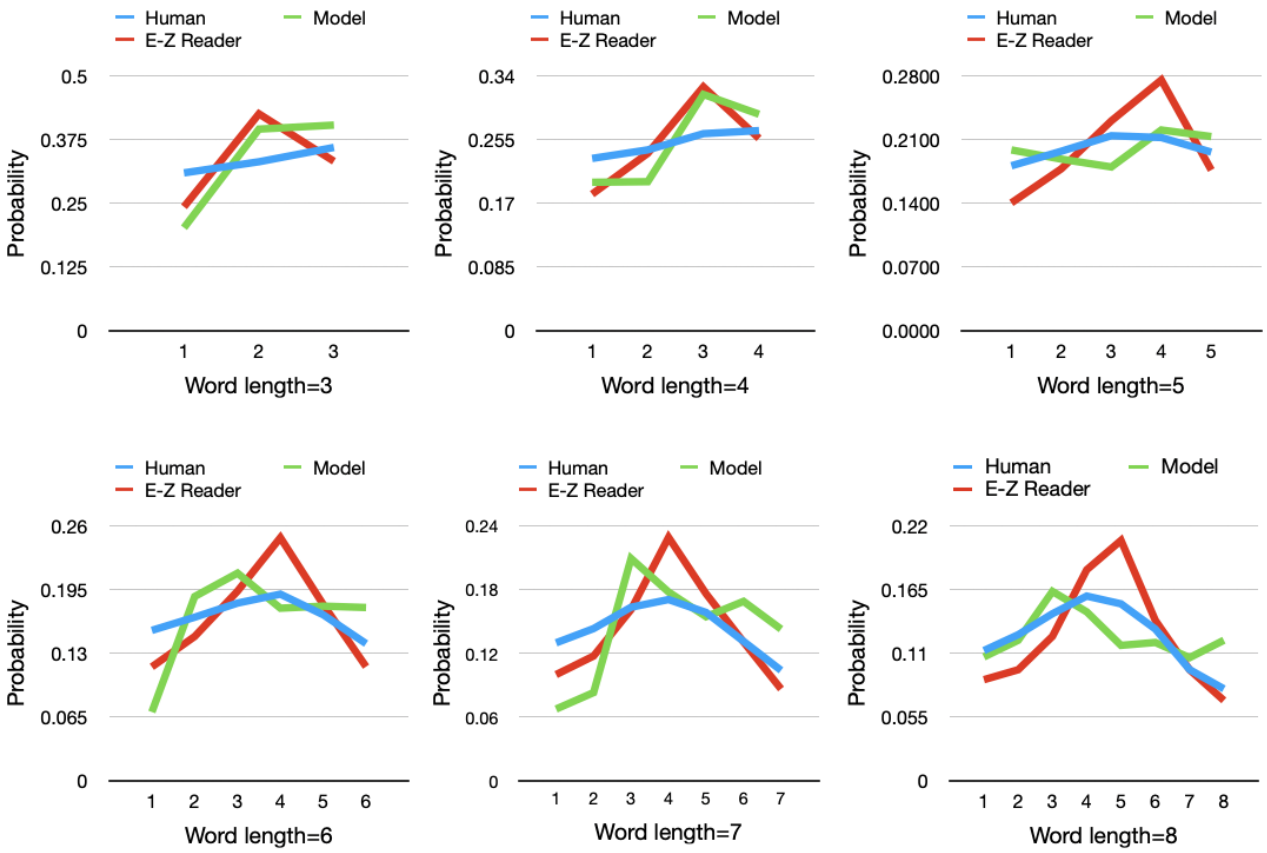

Figure 7: Landing positions for words with five or more character. The x-axis shows the character within the word that is fixated, the y-axis shows the probability of fixating on a given character.

plausible training algorithms such as reinforcement learning, can be a promising way of building scalable rational models of language processing.

In evaluating the model, we collapsed across the 10 readers in the Dundee corpus. Given that human reading exhibits substantial individual differences, accounting for those within a rational modeling framework is an interesting question for future research.

In this work, we focused on predicting fixation locations, and left modeling of the durations of those fixations to future work. Adaptations of surprisal (Hale, 2001; Levy, 2008) computed using character-level language models (e.g. Kim et al., 2016; Hahn et al., 2019) might be useful component for predicting durations for individual character-level fixations.

As described in the Introduction, our model makes a set of simplifying assumptions. First, following much prior work modeling reading, we assumed that the goal of reading was simply to recognize the words read, whereas human readers arguably aim for more high-level text comprehension (Kintsch, 1988). As the Tradeoff Hypothesis as implemented using neural networks can be equally applied to higher-level comprehension tasks (Hahn and Keller, 2018), extending the character-level modeling described here to other tasks, and to accounting for the impact of the reading task on fixations by changing the objective function, will be an interesting task for future research. Second, our model makes the simplifying assumption that saccades in reading always go forward. However, approximately 10% to 25% of saccades jump backwards (regressions, Rayner et al. 2012). A more complete reading model should capture this behavior. An important direction for future research is therefore to extend our model to also incorporate regressive saccades.

References

Barrett, M., Agic, Z., and Søgaard, A. (2015). The dundee treebank. In *Proceedings of the Fourteenth International Workshop on Treebanks and Linguistic Theories*.

Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *J. Mach. Learn. Res.*, 13:281–305.

Bicknell, K. and Levy, R. (2010). A rational model of eye movement control in reading. In *Proceedings of the 48th annual meeting of the Association for Computational Linguistics*, pages 1168–1178.

Blanchard, H., McConkie, G., Zola, D., and Wolverton, G. (1984). Time course of visual information utilization during fixations in reading. *Journal of experimental psychology. Human perception and performance*, 10 1:75–89.

Carpenter, P. A. and Just, M. A. (1983). What your eyes do while your mind is reading. In Rayner, K., editor, *Eye Movements in Reading*, pages 275–307. Academic Press, New York.

Engbert, R., Longtin, A., and Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision research*, 42(5):621–636.

Engbert, R., Nuthmann, A., Richter, E. M., and Kliegl, R. (2005). Swift: a dynamical model of saccade generation during reading. *Psychological review*, 112(4):777.

Hahn, M. and Keller, F. (2016). Modeling Human Reading with Neural Attention. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*.

Hahn, M. and Keller, F. (2018). Modeling task effects in human reading with neural attention. *CoRR*, abs/1808.00054.

Hahn, M., Keller, F., Bisk, Y., and Belinkov, Y. (2019). Character-based surprisal as a model of human reading in the presence of errors. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society (CogSci)*.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Second meeting of the north American chapter of the association for computational linguistics*.

Hermann, K. M., Kočiský, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M., and Blunsom, P. (2015). Teaching machines to read and comprehend.

Kaakinen, J. K. and Hyönä, J. (2010). Task effects on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(6):1561–1566.

Kennedy, A. (2003). The dundee corpus [cd-rom]. *Psychology Department, University of Dundee*.

Kim, Y., Jernite, Y., Sontag, D., and Rush, A. M. (2016). Character-aware neural language models. In *Thirtieth AAAI conference on artificial intelligence*.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological review*, 95(2):163.

Legge, G., Klitz, T., and Tjan, B. (1997). Mr. chips: an ideal-observer model of reading. *Psychological review*, 104 3:524–53.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177.

Lewis, R. L., Shvartsman, M., and Singh, S. (2013). The adaptive nature of eye movements in linguistic tasks: How payoff and architecture shape speed-accuracy trade-offs. *Topics in Cognitive Science*, 5:581–610.

Matthies, F. and Søgaard, A. (2013). With blinkers on: Robust prediction of eye movements across readers. In *EMNLP*.

Nilsson, M. and Nivre, J. (2009). Learning where to look: Modeling eye movements in reading. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009)*, pages 93–101.

Osaka, N. (1993). Reading vertically without a fovea. In Wright, S. F. and Groner, R., editors, *Facets of Dyslexia and its Remediation*, volume 3 of *Studies in Visual Information Processing*, pages 257–265. North-Holland.

O'Regan, K. (1979). Saccade size control in reading: Evidence for the linguistic control hypothesis. *Perception & Psychophysics*, 25(6):501–509.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372.

Rayner, K. (2009). Eye movements in reading: Models and data. *Journal of Eye Movement Research*, 2:1–10.

Rayner, K. and Morris, R. K. (1992). Eye movement control in reading: Evidence against semantic preprocessing. *Journal of Experimental Psychology: Human perception and performance*, 18(1):163.

Rayner, K., Pollatsek, A., Ashby, J., and Clifton Jr, C. (2012). *Psychology of reading*. Psychology Press.

Reichle, E. D., Pollatsek, A., Fisher, D. L., and Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological review*, 105(1):125.

Reichle, E. D., Pollatsek, A., and Rayner, K. (2012). Using e-z reader to simulate eye movements in nonreading tasks: a unified framework for understanding the eye-mind link. *Psychological review*, 119 1:155–85.

Reichle, E. D., Rayner, K., and Pollatsek, A. (2003). The ez reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and brain sciences*, 26(4):445.

Reichle, E. D., Warren, T., and McConnell, K. (2009). Using ez reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic bulletin & review*, 16(1):1–21.

Schotter, E. R., Bicknell, K., Howard, I., Levy, R., and Rayner, K. (2014a). Task effects reveal cognitive flexibility responding to frequency and predictability: Evidence from eye movements in reading and proofreading. *Cognition*, 131(1):1–27.

Schotter, E. R., Reichle, E. D., and Rayner, K. (2014b). Rethinking parafoveal processing in reading: Serial-attention models can explain semantic preview benefit and n+ 2 preview effects. *Visual Cognition*, 22(3-4):309–333.

Snell, J. and Theeuwes, J. (2020). A story about statistical learning in a story: Regularities impact eye movements during book reading. *Journal of Memory and Language*, 113:104127.

Williams, R. J. (2004). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8:229–256.