Inferring Reading Comprehension from Eye Movements

David R. Reich[1,2], Paul Prasse[1], Chiara Tschirner[1,2], Patrick Haller[2], Maja Stegenwallner-Schütz[1], Frank Goldhammer[3,4], and Lena A. Jäger[1,2]

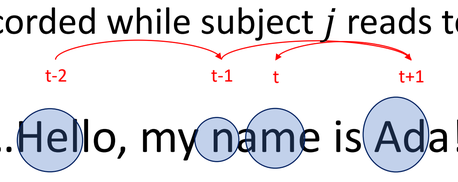
[1]University of Potsdam; [2]University of Zurich; [3]DIPF | Leibniz Institute for Research and Information in Education; [4]Centre for International Student Assessment (ZIB)

Research Question

- Predict reading comprehension from eye movements on a given text.
- **Challenge:** Generalize to readers not seen during training

Problem Settings

Given scanpath S, recorded while subject j reads text T_i, infer task-specific binary label y.



$$f(\text{...Hello, my name is Ada!...}) \rightarrow y_{i,j}$$

Goal: Find a model f that can infer label y from scanpath S and Text T_i.

Bundesministerium für Bildung und Forschung

This work was partially funded by the German Federal Ministry of Education and Research under grant 01|S20043.

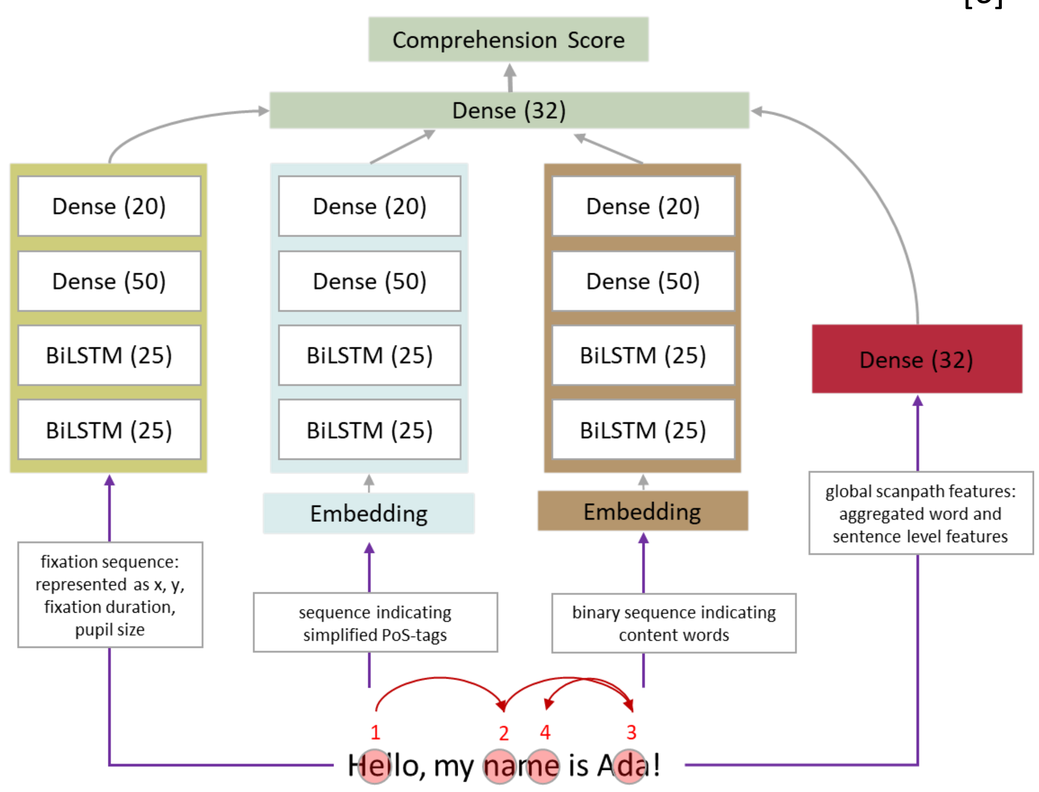
Method

Data: Publicly available data set [1]
- Four text passages from the SAT, 95 readers
- Eye movement data and scores on respective comprehension questions

Investigated tasks:
i) **Text Comprehension**: readers' **text specific** comprehension score (binarized)
ii) **General Reading Comprehension: average** of a reader's comprehension scores on all four SAT text passages (binarized)
iii) Readers' self-reported **Text Difficulty**
iv) **Native Speaker:** whether the reader is a native speaker of English

Evaluation settings – 3 different cross-validation settings:
- New Page: hold out individual pages.
- New Book: hold out entire books.
- New Reader: hold out subjects.

Model architecture [6]



Global Scanpath Features [2, 3, 4, 5, 6, 7]

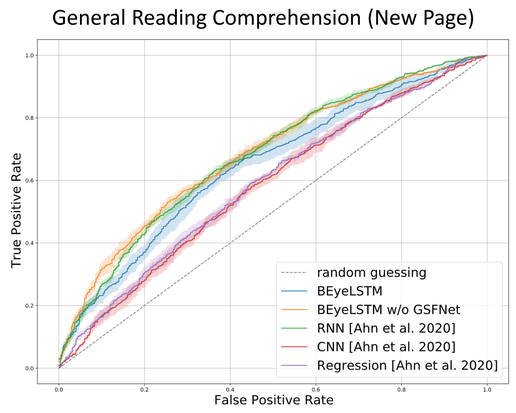
Fixation Features
- Horizontal coordinate
- Vertical coordinate
- Fixation duration
- Pupil size on fixation

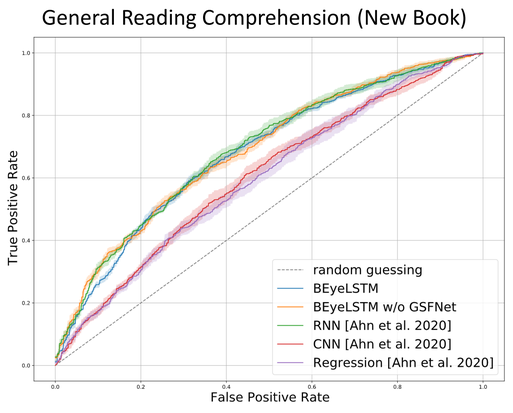
Reading Measures
- First Fixation Duration (FF)
- Total Fixation Duration (TF)
- Incoming Regression Count
- Outgoing Progressive / Regressive Saccade Count
- Averaged Horizontal / Vertical Fixation Location
- Words in Fixed Context on FF / TF
- Syntactic Clusters on FF /TF
- Information Clusters on FF / TF

Linguistic Features
- Word Length
- Lexicalized Surprisal
- PoS-tag / Simplified PoS-tag
- is_content_word
- Named Entity Type
- Right / Left dependencies count

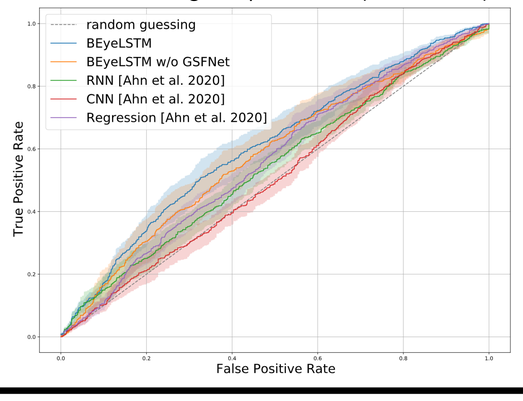
Results

	Model	New Page	New Book	New Reader
Gen. Reading Comprehension	BEyeLSTM	0.68 ± 0.006*	0.648 ± 0.023*	**0.608 ± 0.037***
	BEyeLSTM w/o GSFNet	0.687 ± 0.007*	**0.683 ± 0.009***	0.581 ± 0.045
	RNN [Ahn et al. 2020]	**0.69 ± 0.009***	0.677 ± 0.008*	0.542 ± 0.028
	CNN [Ahn et al. 2020]	0.605 ± 0.015*†	0.582 ± 0.013*†	0.513 ± 0.03
	Regression [Ahn et al. 2020]	0.599 ± 0.016*†	0.59 ± 0.007*†	0.564 ± 0.031
Text Comprehension	BEyeLSTM	0.596 ± 0.012*	0.504 ± 0.015	**0.542 ± 0.015***
	BEyeLSTM w/o GSFNet	**0.597 ± 0.013***	**0.522 ± 0.013**	0.521 ± 0.029
	RNN [Ahn et al. 2020]	0.571 ± 0.01*	0.507 ± 0.01	0.514 ± 0.024
	CNN [Ahn et al. 2020]	0.538 ± 0.006*†	0.493 ± 0.009	0.485 ± 0.016†
	Regression [Ahn et al. 2020]	0.539 ± 0.007*†	0.492 ± 0.013	0.532 ± 0.016
Text Difficulty	BEyeLSTM	**0.746 ± 0.01***	0.516 ± 0.011	**0.71 ± 0.017***
	BEyeLSTM w/o GSFNet	0.652 ± 0.008*†	**0.545 ± 0.023**	0.595 ± 0.014*†
	RNN [Ahn et al. 2020]	0.567 ± 0.013*†	0.51 ± 0.02	0.553 ± 0.032†
	CNN [Ahn et al. 2020]	0.53 ± 0.014†	0.523 ± 0.017	0.511 ± 0.016†
	Regression [Ahn et al. 2020]	0.564 ± 0.011*†	0.523 ± 0.004*	0.516 ± 0.014†
Native Speaker	BEyeLSTM	0.737 ± 0.011*	**0.7 ± 0.017***	**0.67 ± 0.025***
	BEyeLSTM w/o GSFNet	**0.744 ± 0.01***	0.696 ± 0.015*	0.612 ± 0.045*
	RNN [Ahn et al. 2020]	0.723 ± 0.013*	0.69 ± 0.014*	0.581 ± 0.015*†
	CNN [Ahn et al. 2020]	0.65 ± 0.009*†	0.628 ± 0.007*†	0.574 ± 0.022*†
	Regression [Ahn et al. 2020]	0.667 ± 0.006*†	0.664 ± 0.01*	0.599 ± 0.033*





- BEyeLSTM outperforms state-of-the-art models for each task in the New Reader setting.



Discussion

- BEyeLSTM is the first model to infer reading comprehension for new readers that are not in the training data
- The sequential information encoded in scanpaths is highly informative with respect to reading comprehension
- Classification into skilled/unskilled is easier than the prediction of text specific comprehension accuracy
- Future research: Challenge remains to generalize to texts and readers not seen during training

References

[1] Ahn et al. *ETRA* 2020 [2] Berzak et al. *NAACL* 2018 [3] Hale *NAACL* 2001 [4] Levy *Cognition* 2008 [5] Martinez-Gomez & Aizawa *IUI* 2014 [6] Reich et al. *ETRA* 2022 [7] Shiferaw et al. *Drug Alcohol Depen.* 2019