Neural Gaze Pattern Prediction

Ekta Sood, Andreas Bulling

Human-Computer Interaction - Cognitive Systems Group, University of Stuttgart, Germany.

Contact: ekta.sood@vis.uni-stuttgart.de

Motivation

- Previous models which predict text eye movement during reading tasks = rule-based, biased towards the features and the domain.
- Neural based models fail to accurately predict fixations across various domains.
- Robust evaluation techniques = lacking as gaze data collection is expensive.
- Text saliency to deal with varying semantic contexts for cognitively motivated machine based understanding.

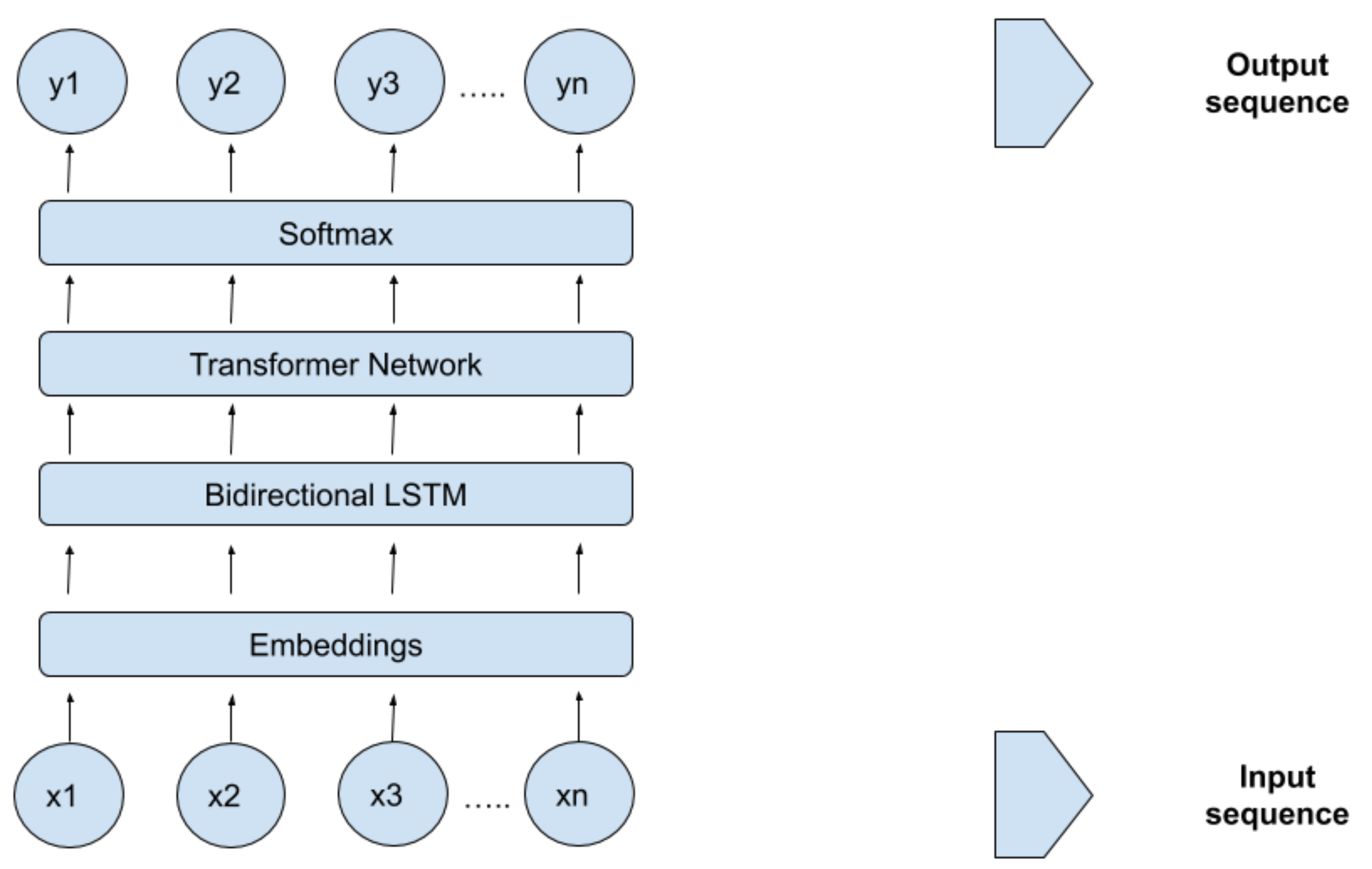
Methods

- BiLSTM with stacked multi-headed self-attention network to learn cross domain gaze patterns
- Binary **classification task to predict fixations or skips** for each token in the input sequence.
- Each token (word) in input sequence has corresponding labels: 0 for skip or 1 for fixation.
- W2V word embeddings

Data

- Training: Provo and Geco Corpus = 65547 sentences (61.8% fixated).
- Val: Provo and Geco Corpus = 7284 sentences (53.6% fixated).
- Test: MQA-RC Corpus = 1581 sentences (50.1% fixated).
- Model is **trained on combined corpus** and **tested on a different out of domain corpus**.

Model Architecture



Evaluation

- Against predictions from two baseline systems: the E-Z Reader 10 model (rule based system) and our simple BiLSTM network without attention.
- Compare our network against the pertained BERT transformer network.
- **Accuracy** to measure predictions against human gaze data = ratio of correct predictions (compared to gold standard).
- **Normalized Mutual Information** to measure similarity of distribution from E-Z reader token fixation durations to humans (closer to 1 = more similarity)

Results

	Val	Test
EZ	-	54%
BiLSTM	65.1%	54.2%
BiLSTM+Att	68%	62%
Bert	65.6%	62.8%

Table 1: All Model Accuracy Results

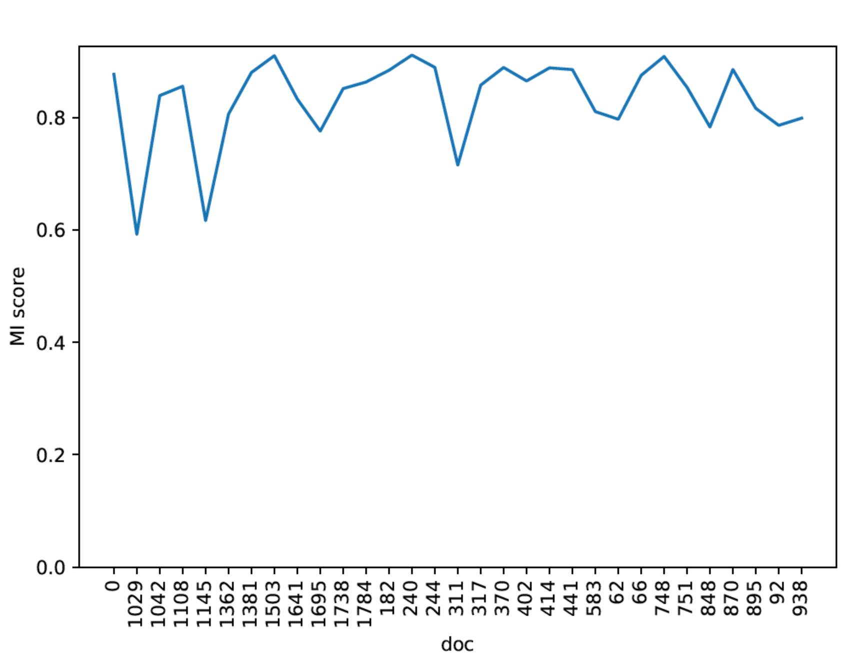


Figure 1: Mutual information score between E-Z Reader and humans on test set.

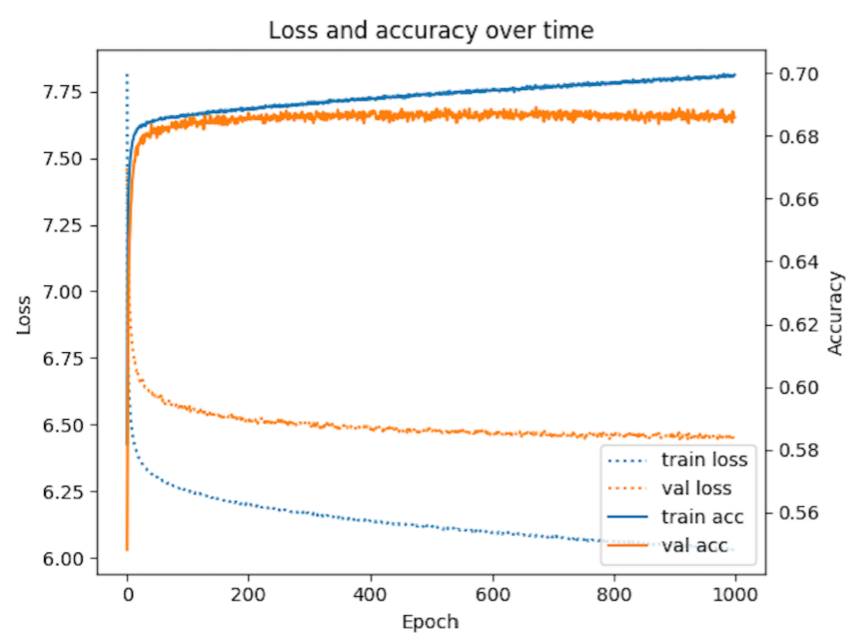


Figure 2: BiLSTM with self attention

Our model is comparable to BERT (pertained on out-of-domain corpus), resulting in 62% accuracy. The E-Z Reader model accuracy is lower, yet the distribution of fixation durations shares similar information (0.6-0.8) observed in the human data —indicating that the E-Z reader model is successful in predicting token level fixation **durations**.

Conclusion

- Successfully trained classifier to predict reading patterns
- Our attention based model = increased performance against both baselines. We show comparable performance to BERT transformer network.
- Future work = change task to a regression task. The model objective will be to predict token level fixation **durations**. We aim to evaluate the distribution of predicted durations as well as the token level attention weights against humans.

References

[1] Reichle et al. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. Behavioral and brain sciences 26.4.
[2] Hahn & Keller (2016). Modeling human reading with neural attention. arXiv preprint arXiv:1608.05604.
[3] Devlin, Jacob, et al. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805