Improving prediction of visual outcomes from the electronic health record in ophthalmology using BERT and time-aware LSTMs

Bryan Gopal[1], Brian Soetikno[2]
[1]Department of Computer Science, [2]Department of Ophthalmology

Purpose

To investigate whether BERT-based models and time-aware LSTM models could improve prediction performance of low-vision from unstructured clinical notes in the electronic health record.

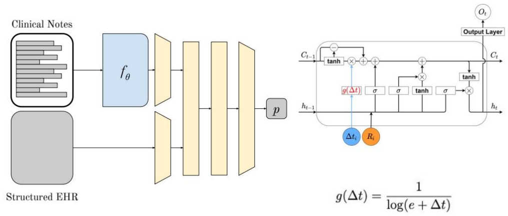
Background

- The **electronic health record (EHR)** possesses a wealth of unstructured clinical data that could be unlocked for the prediction of health risk.
- In ophthalmology, an important measure of a patient's vision is their visual acuity, usually measured by their best line read on a standardized eye chart.
- Patients with **low vision** - defined as <20/40 visual acuity - are at increased risk of falls, depression and anxiety, and mortality.
- Importantly, the direct and timely intervention of providing low vision services to visually impaired patients could maximize their quality of life.
- However, these services are unfortunately underutilized, and patients that could benefit from low vision services receive treatment do not receive this care.
- The goal of this project is to develop algorithms that can accurately predict patients with continued low vision from HER data.

Dataset

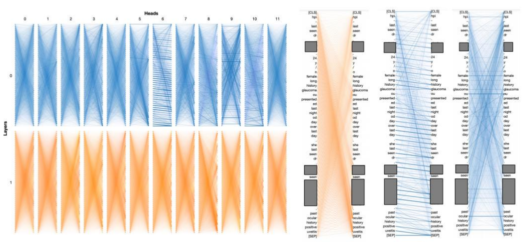
- Ophthalmology notes from **Stanford EHR data warehouse** between 2009 to 2018 (curated by collaborator Dr. Sophia Wang, MD)
- **Structured data fields** – 556 features including age, race, gender, visual acuity
- **Free-text clinical notes** included physician notes, surgical operative reports, and imaging interpretation
- **Prediction label:** a binary prediction task where 1 represents persistent low vision (<20/40 measured visual acuity) in the next 12 months and 0 return to normal visual acuity in the next 12 months.
- The dataset contains 15,409 notes related to 2258 patients that have low vision, and 52,267 notes where 3290 patients have normal visual acuity

Models



Model Architectures. Our **Structure-Fused Fine-tuned BERT** model combined the final layer outputs of a biomedical BERT model and a projection of the structured EHR data and passed the resulting embedding through an MLP in order to derive predictions. A **Time-aware LSTM (TLSTM)** cell includes a weighting factor g to more heavily attend to recent notes in a sequence. In our experiments, we added an TLSTM cell to the end of the BERT model in order to learn time-aware predictions.

Qualitative Results



Examples of attention visualization. We qualitatively analyze the behavior of our best transformer by visualizing its attention heads for a given example. Some heads paid attention to the next token in the sequence. Other heads paid attention to medically relevant keyworks such as 'glaucoma' and 'uveitis'. Many heads showed uninterpretable behavior, with seemingly uniform attention paid to all words.

Quantitative Results

Model Type	AUROC
FNN	0.784
CNN	0.805
FNN+CNN	0.826
Chunked Pretrained BERT Embeddings FNN	0.702
Fine-tuned BERT	**0.843**
Structure-Fused Chunked Pretrained BERT Embeddings FNN	0.712
Structure-Fused Fine-tuned BERT	**0.844**
BERT w/ LSTM	0.553
BERT w/ TLSTM	0.538

- BERT-only approaches outperformed baseline approaches.
- Structured fusion only marginally helped BERT-only methods.
- Sequence-based approaches involving LSTMs and TLSTMs heavily underperformed baselines. We suspect that this is due to a lack of data.
  - Our training dataset needed to be heavily reduced to create the longitudinal dataset required to train a TLSTM.
  - Further experiments on more naturally longitudinal datasets would need to be done in order to assess the value of these methods.

Conclusions

- Using a single-center EHR database with ophthalmology notes, we successfully implemented BERT-based prediction models.
- We qualitatively analyzed our models and found distinct attention patterns in the visualizations.
- Time-aware LSTMs did not improve the performance of BERT based models on this dataset. Other methods of incorporating temporal information into models should be explored in the future