

Sequential Sentence Classification in Medical Abstracts

ME 781 Course Project

by

Krishanu Das

Third Year Undergraduate

Roll No: 20D170020

*Under the guidance of
Prof. Asim Tiwari*



Indian Institute of Technology, Bombay
Mumbai 400 076

2nd November 2022

Replicating the deep learning model behind the 2017 paper [PubMed 200k RCT: a Dataset for Sequential Sentence Classification in Medical Abstracts](#).

When it was released, the paper presented a new dataset called PubMed 200k RCT which consists of ~200,000 labelled Randomized Controlled Trial (RCT) abstracts.

The goal of the dataset was to explore the ability for NLP models to classify sentences which appear in sequential order.

In other words, given the abstract of a RCT, what role does each sentence serve in the abstract?

Nutritional psychiatry: the present state of the evidence

Wolfgang Marx¹, Genevieve Moseley², Michael Berk², Felice Jacka²

Affiliations + expand

PMID: 28942748 DOI: 10.1017/S0029665117002026

Abstract

Mental illness, including depression, anxiety and bipolar disorder, accounts for a significant proportion of global disability and poses a substantial social, economic and health burden. Treatment is presently dominated by pharmacotherapy, such as antidepressants, and psychotherapy, such as cognitive behavioural therapy; however, such treatments avert less than half of the disease burden, suggesting that additional strategies are needed to prevent and treat mental disorders. There are now consistent mechanistic, observational and interventional data to suggest diet quality may be a modifiable risk factor for mental illness. This review provides an overview of the nutritional psychiatry field. It includes a discussion of the neurobiological mechanisms likely modulated by diet, the use of dietary and nutraceutical interventions in mental disorders, and recommendations for further research. Potential biological pathways related to mental disorders include inflammation, oxidative stress, the gut microbiome, epigenetic modifications and neuroplasticity. Consistent epidemiological evidence, particularly for depression, suggests an association between measures of diet quality and mental health, across multiple populations and age groups; these do not appear to be explained by other demographic, lifestyle factors or reverse causality. Our recently published intervention trial provides preliminary clinical evidence that dietary interventions in clinically diagnosed populations are feasible and can provide significant clinical benefit. Furthermore, nutraceuticals including n-3 fatty acids, folate, S-adenosylmethionine, N-acetyl cysteine and probiotics, among others, are promising avenues for future research. Continued research is now required to investigate the efficacy of intervention studies in large cohorts and within clinically relevant populations, particularly in patients with schizophrenia, bipolar and anxiety disorders.

Source: <https://pubmed.ncbi.nlm.nih.gov/28942748/>

Harder to read

Model

Considerations for a surgical RCT for diffuse low-grade glioma: a survey

Alireza Mansouri¹, Karanbir Brar², Michael D Cusimano³

Affiliations + expand

PMID: 32537182 PMCID: PMC7274180 (available on 2021-06-01) DOI: 10.1093/nop/npz058

Abstract

Background: Diffuse low-grade gliomas (DLGGs) are heterogeneous tumors that inevitably differentiate into malignant entities, leading to disability and death. Recently, a shift toward up-front maximal safe resection of DLGGs has been favored. However, this transition is not supported by randomized controlled trial (RCT) data. Here, we sought to survey the neuro-oncology community on considerations for a surgical RCT for DLGGs.

Methods: A 21-question survey focusing on a surgical RCT for DLGGs was developed and validated by 2 neurosurgeons. A sample case of a patient for whom management might be debatable was presented to gather additional insight. The survey was disseminated to members of the Society for Neuro-Oncology (SNO) and responses were collected from March 16 to July 10, 2018.

Results: A total of 131 responses were collected. Sixty-three of 117 (54%) respondents thought an RCT would not be ethical, 39 of 117 (33%) would consider participating, and 56 of 117 (48%) believed an RCT would be valuable for determining the differing roles of biopsy, surgery, and observation. This was exemplified by an evenly distributed selection of the latter management options for our sample case. Eighty-three of 120 (69.2%) respondents did not believe in equipoise for DLGG patients. Quality of life and overall survival were deemed equally important end points for a putative RCT.

Conclusions: Based on our survey, it is evident that management of certain DLGG patients is not well defined and an RCT may be justified. As with any surgical RCT, logistic challenges are anticipated. Robust patient-relevant end points and standardization of perioperative adjuncts are necessary if a surgical RCT is undertaken.

Source: <https://pubmed.ncbi.nlm.nih.gov/32537182/>

Easier to read

Overview

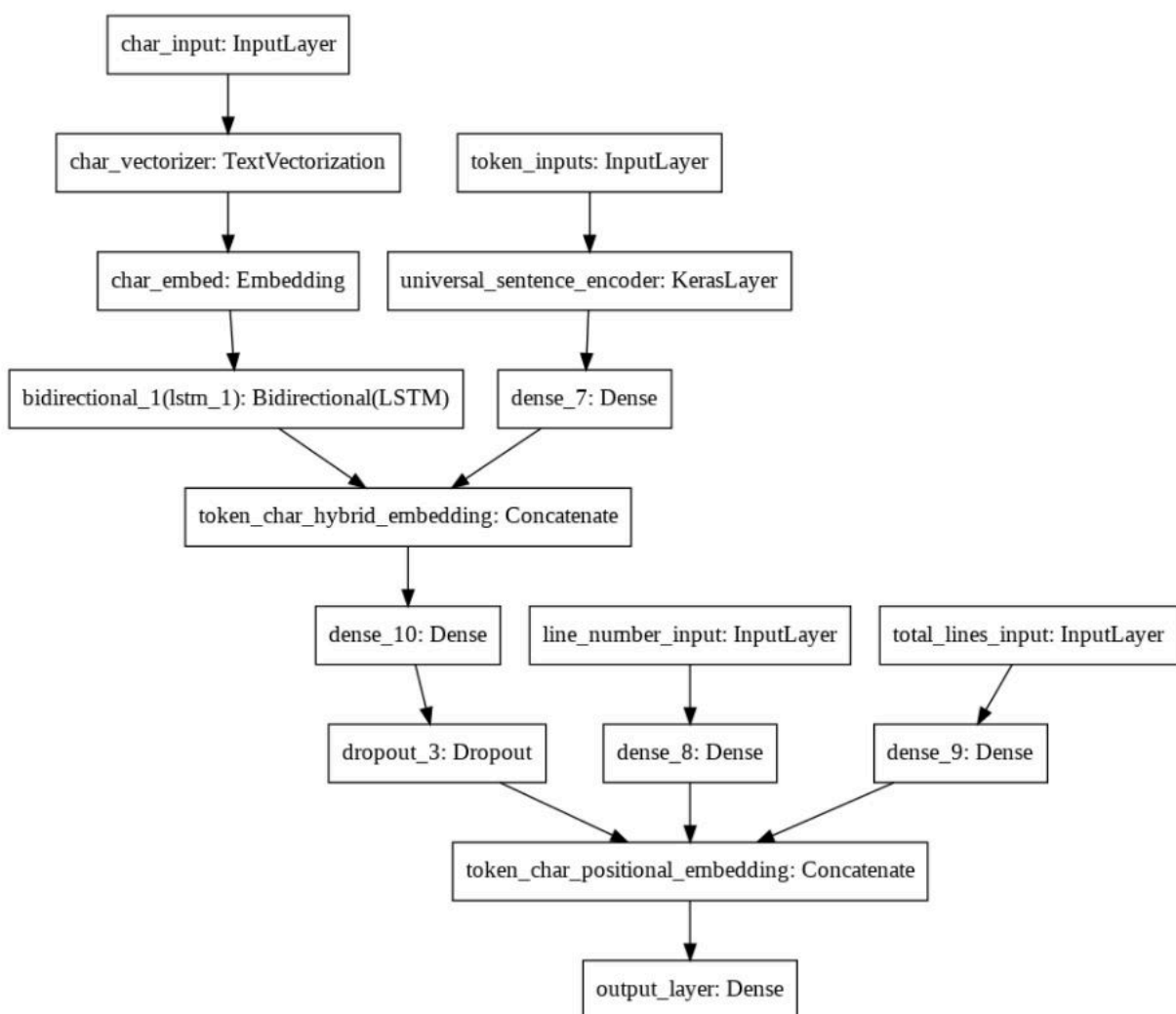
1. *Classify a Randomized clinical trials (RCTs) abstract to subclasses for easier to read and understand.*
2. *Basically convert a medical abstract to chunks of sentences of particular classes like "Background", "Methods", "Results" and "Conclusion".*
3. *It's a Many to One Text Classification problem. Where we categorize a sequence to a particular class.*

Models

1. Model-0 : Baseline (TF-IDF Multinomial Naive Bayes)
2. Model-1 : Conv1D (token embeddings)
3. Model-2 : USE (Universal Sentence Encoder Pretrained embeddings)

4. Model-3 : Conv1D (character embeddings)
5. Model-4 : Pretrained token embeddings(USE) + character embeddings (Hybrid embedding)
6. Model-5 : Pretrained token embeddings(USE) + character embeddings + positional embeddings(Tribrid embedding)

Visualize Tribrid Model (Pretrained token embeddings(USE) + character embeddings + positional embeddings)



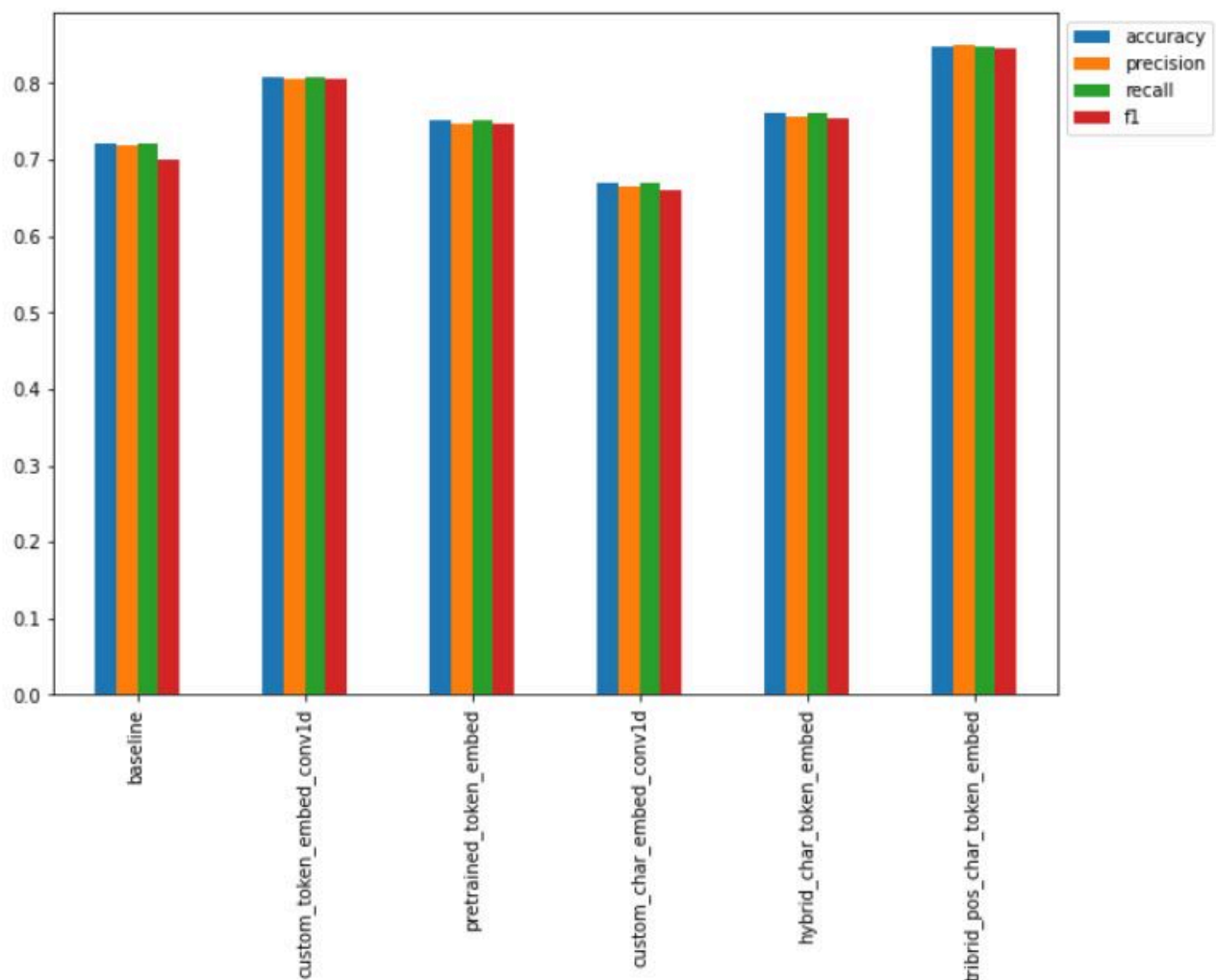
Results

Since the [PubMed 200k RCT: a Dataset for Sequential Sentence Classification in Medical Abstracts](#) paper compares their tested model's F1-scores on the test dataset, let's take at our model's F1-scores.

Though, in comparison to the results reported in Table 3 of the [PubMed 200k RCT: a Dataset for Sequential Sentence Classification in Medical Abstracts](#) paper, our model's F1-score is still underperforming (the authors model achieves an F1-score of 90.0 on the 20k RCT dataset versus our F1-score of ~84.4).

There are some things to note about this difference:

- Our models (with an exception for the baseline) have been trained on ~18,000 (10% of batches) samples of sequences and labels rather than the full ~180,000 in the 20k RCT dataset.
 - This is often the case in machine learning experiments though, make sure training works on a smaller number of samples, then upscale when needed (an extension to this project will be training a model on the full dataset).



The table shows the Metrics of all the model:

Model	accuracy	precision	recall	f1
baseline	0.721832	0.718647	0.721832	0.698925
custom_token_emb ed_conv1d	0.808288	0.80478	0.808288	0.805473
pretrained_token_e mbed	0.751059	0.745737	0.751059	0.746384
custom_char_embe d_conv1d	0.669337	0.66386	0.669337	0.659734
hybrid_char_token_ embed	0.759831	0.756189	0.759831	0.754541
tribrid_pos_char_tok en_embed	0.848007	0.849911	0.848007	0.844585