ACL PAPER

PAPER CHOSEN

GPT-D: Inducing Dementia-related Linguistic Anomalies by Deliberate Degradation of Artificial Neural Language Models

Authors:

Primary:
Changye Li- Institute of Health Informatics, University of Minnesota

Co-Authors:
David Knopman- The Mayo Clinic, Rochester, MN
Weizhe Xu – Biomedical Health Informatics, University of Washington
Trevor Cohen - Biomedical Health Informatics, University of Washington
Serguei Pakmahov – Pharmaceutical Care and Health Systems, University of Minnesota

PROBLEM ADDRESSED

Deep Learning (DL) and Deep Learning models are currently being used in detection of several disorders, and specifically in use for Alzheimer's disease research. Usually, DL models are used for correlating the difference and distinguish language between healthy and Alzheimer's patients. Although there have been many improvements, ultimately, ML model rely on parameter settings of the model so as not to either underfit or overfit (which is usually the case). Under the premise that paired perplexity can be the defining parameter by which a language model can be built for dementia models, a ML system can focus solely on this parameter to tune Language model hyperparameters. However, there is plenty of text to base attribute to distinguish a "healthy class" there is little to no text for the dementia class. To overcome this, they utilized a GPT-2 deep model and compromised the semantic processing capabilities to produce artificially generated dementia type responses.

PRIOR WORK

Cohen and Pakmahov previously had shown that DL models for Language detecting of healthy and dementia models is reliant on “paired perplexity”. As defined in the paper, “paired perplexity (PPL) is a measurement of how well a language sample fits a trained Language model (LM). The premise is that a ML would be “surprised” by the language produced by a dementia participant from a healthy participant, and vice versa.

Cohen and Parkmahov (2020) have suggest the that language models react or are sensitive to lexical frequency or repetitiveness.

CONTRIBUTIONS

This work was originally set up to find a means of producing dementia like text for the automated detection of dementia related linguistics. The GPT-D or dementia Transformer was a novel use of the GPT Transformer to produce text and “cognitive task to conversational data”. This work would also provide the basis of future used of the GPT-D transformer by impairing 50% , as they produced studies at 25%, 50% 75% and 100%The model produces language anomalies seen in dementia.

As noted by the authors, their findings are consistent with the findings that Transformer models are able to predict responses during language comprehension. This helps in the understanding of generative artificial neural language models with that of those seen in in human language production and predictions.

RESEARCH EVALUATION

It would seem that with their groundbreaking model the author attempted to stress various factors of cross validation. They first evaluated performance by utilizing a 12-layer BERT model to compare against results produced by the GPT- 2 model and the GPT-D model utilizing established dataset (ADReSS, DB and CCC). It should be noticed that the datasets used are disproportionately have more healthy individual transcripts over dementia transcripts.
The model outperformed the BERT/Distilbert slightly (85% in comparison to 83% accuracy).

For Language generation they used a known prompt used in Dementia and Alzheimer’s research (the “Cookie Thief”). The text produced by the GPT model and GPT-D model support that with degeneration of layers of the model produced visibly more repetitiveness, than the GPT-2 model.

However, as noted by the authors, the results were not statistically significant. However, their patterns reflect to that similarly seen in Alzheimer's patients with dementia.

The authors realize that there are limitation of the study (specifically with that of the dataset sizes). There is also limitation that the work was solely done in English and that that the pretrained language model is gender-biased.

Overall, the work is inspiring and still can be used to predict language patterns in dementia stricken patients. Also, this was an interesting take on the use of ML models.

CITATIONS BY AUTHORS

Changye Li – 20 citations total from 7 total publications, with a h-index of 2 and i10-index of 1.
Weizhe Xu – This author has no citations listed in Google Scholar; however, he is a graduate student with Dr. Trevor Cohen at the University of Washington
David Knopman – 7030 total citations.
Trevor Cohen – 3939 total citations, with a h-index of 23 and i10-index of 81.
Sergei Pakhomov – 6948 total citations, with an h-index of 38 i10-index of 104.

Analysis- Although Dr. Li has only 20 citations, one of them landed with an impact factor of 10, which indicates that her work is fairly groundbreaking to merit a high impact publication. Most likely it seems she is the graduate student of Dr. Pakhomov, whose citations are high, along with current works (h-factor) high as well as numerous impact publications. Dr. Cohen tends to co-publish man works with Dr. Pakhomov which could indicate an extended friendship or good collaboration between the labs. Dr. Knopman has many publications, however, after a little research, he is medical doctor who seems to collaborate with anyone who joins his interests in Alzheimer's Disease and mental degradation. He is most likely a collaborator who provides more of the clinical approaches and treatment options to this group. Finally, Weizhe Hu at first came back with many citations. However, after further review most of the citations were not his works, but rather someone with a similar name or initials. I reviewed the University of Washington and saw he is listed as a graduate student of Dr. Cohen. He may not have many publications now, but this publication might help his graduate thesis and further work. In all, this indicates that the work produced by this group is solid scientifically and in the field of Bioinformatics. It seems that with the exception of Li, they are mainly concerned more in informatics and not so much in the computer science.