International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 07, July-2022

Customized Name Entity Recognition for Medical Data using BERT

Vinayak Krishan Prasad
Information Science Department
R.V College of Engineering
Bangalore, India

Vikram Shenoy
Information Science Department
R.V College of Engineering
Bangalore, India

Merin Meleet
Information Science Department
R.V College of Engineering
Bangalore, India

***Abstract*—Electronic Health Records contain large amounts of useful information about patient health which cannot be capitalized on due to inconsistencies in natural language writing. This makes it very difficult to obtain useful data as one has to understand the different formats of writing in a record, and even larger amounts of useless information before obtaining the required information. The advent of Google's Bidirectional Encoder Representations from Transformers (BERT) which is a transformer-based learning technique is a solution towards the extraction of annotated words and structures from unstructured text. However, to annotate medical terms in an electronic health record with high accuracy, further training and fine tuning of the model is required. The fine-tuned BERT model created is intended to increase annotation accuracy, such that useful data from unstructured EHRs can be extracted easily. The results show that our model has an accuracy of 99% indicating it can be used to extract data from unstructured EHRs. This allows doctors, researchers and patients to obtain useful information that could be vital for diagnoses or creation of new datasets for further research in a medical domain.**

Keywords—Electronic Health Records; Natural language processing; Entity Recognition; Tokenization; BERT

I. INTRODUCTION

BERT is a free and open-source machine learning framework for dealing with natural language (NLP). BERT uses the surrounding text to provide context in order to help computers understand the meaning of ambiguous words in text. With the help of question-and-answer datasets, the BERT framework can be adjusted after being pre-trained on text from Wikipedia.

The use of the transformer's bidirectional training, an established attention model, to language modeling is the main technological achievement of BERT. However, past studies either examined text sequences from a left-to-right training viewpoint or a mixed left-to-right and right-to-left training perspective.

Google initially unveiled Transformers in 2017. When language models were first developed, NLP tasks were mostly handled by recurrent neural networks (RNN) and convolutional neural networks (CNN).

Following the introduction of transformers, Google unveiled BERT (open-source) in 2018. (i)Sentiment analysis, (ii)semantic role labeling, (iii)sentence categorization, and (iv)the disambiguation of polysemous words, or words with various meanings, were among the 11 natural language comprehension tasks for which the framework produced ground-breaking results during its research stages.

The results of the study show that bidirectionally trained language models are better able to understand context and language flow than single-direction language models. In their study, the authors provide a novel approach, Masked LM (MLM), which enables bidirectional training in models where it was previously impractical.

By successfully completing the 4 tasks listed above, BERT set itself apart from earlier language models like word2vec and GloVe, which are constrained in their ability to grasp context and polysemous words. According to research scientists in the field, ambiguity is the biggest problem for natural language processing and is successfully addressed by BERT. It has the ability to parse text using "common sense" that is largely human-like.

NER is a task in NLP that involves finding and extracting entities—meaningful information—from sentences or other material. A word or even a group of words that all refer to the same category can be considered an entity. By fine tuning BERT for the Electronic health record data given, we can create a powerful entity extraction tool, which helps in tabulation of data and structuring of information which can further be applied in machine learning tasks such as CDSS (Clinical Decision Support Systems).

II. LITERATURE SURVEY

An article titled "BERT" (Bidirectional Encoder Representations from Transformers) was published [5] by Google AI Language researchers. The Transformer encoder reads the full sequence of words at once, in contrast to directional models, which read the text input sequentially (from right to left or left to right). Although it would be more appropriate to describe it as non-directional, it is thus thought of as bidirectional. The largest model of its sort is called BERT large, and it has 345 million parameters.

A contextualized embedding with pre-training on large-scale structured health records is called MED-BERT [1]. They enhance illness prediction studies using modest local training datasets, lower data collecting costs, and quicken the speed of healthcare using artificial intelligence. But only medications were noted in the study; no external influences were.

The purpose of the author's [2] investigation on the BERT model's efficacy for clinical or biological entity normalization. The enormous number of unlabeled EHR notes they used to fine-tune BioBERT produced a BERT-based model that was

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 07, July-2022

trained using 1.5 million notes from electronic health records (EHR-BERT). They discovered that several cutting-edge systems were outperformed by BERT-based normalization models. Furthermore, pretraining our models on extensive EHR notes might enhance performance even more.

Embedding Electronic Health Records [3] to Learn BERT based Models for Diagnostic Decision Support is covered by Rui Tang et al. The innovative diagnostic prediction framework for illness categorization based on textual clinical notes and age information from EHR data is proposed. It is built on Bidirectional Encoder Representations from Transformers (BERT). The outcomes show that our models perform better than baselines trained using sophisticated text categorization techniques.

A study using patient electronic health records [4] and BERT-based neural networks to predict clinical diagnosis. On the problem of multiclass classification for 265 illness subsets of ICD-10, they conducted a variety of comparison tests with alternative text representation methods. The trials show enhanced performance of the models in comparison to other baselines, and a tweaked Russian BERT (RuBERT) variation

demonstrates the system may be able to minimize misdiagnosis.

RoBERTa - an optimized BERT pretraining model [6], exceeded the performance of BERT achieving good results on

the GLUE, RACE, SquAD datasets respectively. The model was modified to train longer, in larger batches, on more data, remove the purpose of predicting the next sentence, train on longer sequences, and dynamically change the masking pattern used on the training data. To more effectively control for training set size effects, they also gather a sizable new dataset (CC-NEWS) that is comparable in size to existing privately used datasets.

III. METHODOLOGY

The Groningen Meaning Bank (GMB) collection is made up of multi-sentence texts that also include annotations for named entities, parts of speech, lexical categories, and other natural language structural phenomena. We used a portion of the dataset [10] that IBM had annotated with POS and IOB tags.

A class known as Sentence Getter is created. This defines a method which converts the data into separate sentences with the delimiter being a full stop '.'. Following which preprocessing and tokenization was done.

Tokenization involved importing transformers library and Bert-base-cased NLP model. Each sentence was passed through a tokenization function which used the tokenize() method of bert-tokenizer. The tokenized texts were appended with their respective labels and stored in separate memory i.e., a 2-D list.

This 2-D list was then separated into tokens and labels for tokenized text and their corresponding labels respectively. The maximum length of each tokenized sentence was 75 characters long. Padding was added to whichever sentence that required in.

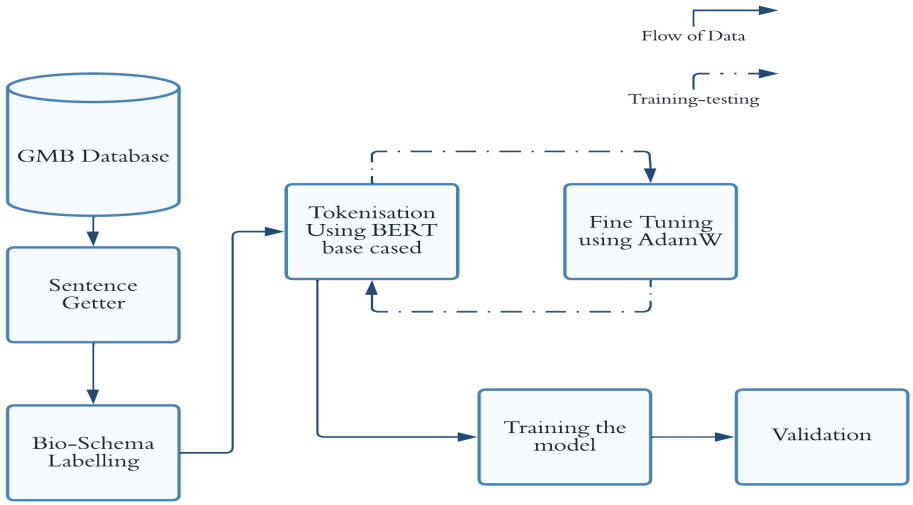

Fig. 1. Loss Curve for Entity Recognition

The training and test split was chosen to be 90% training data and 10% validation data. Then we convert inputs and masks to torch tensors. Data shuffling was done; Training set - random shuffling, Validation Set - sequential shuffling.

The AdamW optimizer was used for parameter optimization and fine tuning. The optimizer AdamW features a better weight decay implementation than the original Adam. The weight decay is implicitly correlated with the learning rate in the Adam optimizer's common weight decay implementation. However, when each training data point is entered into this implementation, a new optimum weight decay for each learning rate is allocated.

The weight decay is separated from the optimization process by the AdamW optimizer. As a result, the learning rate may be changed without affecting the ideal weight decay, allowing for the optimization of both variables independently. This patch significantly enhances the generalization performance.

The fine Tuning of parameters involved weight decay rate set to 0.01 if the default optimizer were found to be either 'bias', 'gamma' or 'beta' otherwise, weight decay rate was set to

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 07, July-2022

0.0. The weight optimization was grouped and passed to AdamW optimizer with learning rate 3e-5 and eps 1e-8.

IV. RESULTS

The model was run over 3 epochs using AdamW as the optimizer and varying the parameters as done in the Fine-Tuning Stage. Overfitting is prevented by adjusting the weight decay rate and beta and gamma batch normalization is added to ensure the data at each hidden layer is normalized, increasing accuracy of the recognition.

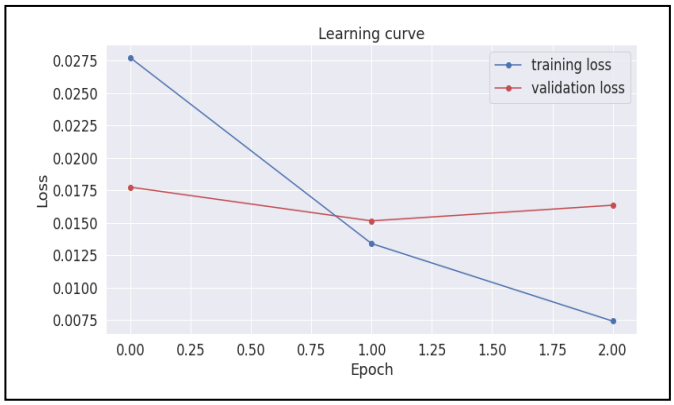

Fig. 2. Loss Curve for Entity Recognition

The validation size was 20% of the total dataset and produced an accuracy of 99% and a validation loss of 0.016. The high accuracy indicates that the fine-tuning technique used produced better results as opposed to directly using the BERT model used in earlier works.

V. CONCLUSION

The work done shows that a name entity system, customized and fine-tuned on the BERT model, to recognize medical terms from unstructured EHRs with high accuracy can be used for the purpose of identifying and extracting useful information from large cohorts of text.

The work done is a step ahead in the field of entity recognition for medical related data, however, there are a few issues and a larger future scope to build on. Firstly, further fine-tuning and an increased number of epochs can be used. Furthermore, the annotation of data, although useful in the identification of useful data, cannot be directly used for research purposes. The data could be formed into separate medical entities consisting of diagnoses, disease, medication of a patient. This data could then be further formed into tabulated data that can be used directly by doctors and research.

REFERENCES

[1] Rasmy, L., Xiang, Y., Xie, Z. et al. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. npj Digit. Med. 4, 86 (2021).
[2] Li F, Jin Y, Liu W, Rawat BPS, Cai P, Yu H. Fine-Tuning Bidirectional Encoder Representations From Transformers (BERT)-Based Models on Large-Scale Electronic Health Record Notes: An Empirical Study. JMIR Med Inform. 2019 Sep 12.
[3] R. Tang et al., "Embedding Electronic Health Records to Learn BERT-based Models for Diagnostic Decision Support," 2021 IEEE 9th International Conference on Healthcare Informatics (ICHI), 2021, pp. 311-319.
[4] Blinov, P., Avetisian, M., Kokh, V., Umerenkov, D., Tuzhilin, A. (2020). Predicting Clinical Diagnosis from Patients Electronic Health Records Using BERT-Based Neural Networks. In: Michalowski, M., Moskovitch, R. (eds) Artificial Intelligence in Medicine. AIME 2020.
[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186.
[6] Liu, Yinhan & Ott, Myle & Goyal, Naman & Du, Jingfei & Joshi, Mandar & Chen, Danqi & Levy, Omer & Lewis, Mike & Zettlemoyer, Luke & Stoyanov, Veselin. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach.
[7] Choi, E. et al. RETAIN: An Interpretable Predictive Model for Healthcare using Reverse Time Attention Mechanism. Adv. Neural Inf. Process. Syst. 29, 3504–3512 (2016).
[8] Choi, E., Bahadori, M. T., Schuetz, A., Stewart, W. F. & Sun, J. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. In Machine Learning for Healthcare Conference, 301–318 (MLHC, 2016).
[9] Rajkomar, A. et al. Scalable and accurate deep learning with electronic health records. NPJ Digital Med. 1, 18 (2018).
[10] IBM developer website last accessed 2022/7/4.
[11] Esteva, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118 (2017).