A BERT-based Firm-level Risk Measurement on Epidemic Diseases

Zhou Xing†

†University of Chicago

Introduction

- Existing literature in firm-level risk measurement relies on simple related words or n-grams counting in earning call transcripts, ignoring the sementics of sentences and context. (Hanssen et al. 2017,2020[1],2020[2])
- We adapt a front-end natural language pretraining model, BERT(Devlin et al., 2018), to construct a measurement of firm-level risk on epidemic diseases, and focus on the impact of COVID-19 since Q4,2019 to Q2,2020.

Research Question

- How to develop a measurement of firm-level risk on epidemic disease using semantic information from context of earning call transcripts?
- How well does our NLP model in classifying sentences into epidemic-disease related or unrelated?
- What is the distribution of such risk among industries?

Data

- Main Analysis Data: Earnings conference call transcripts of 3597 companies of nine industries from Q3,2019 to Q2, 2020 from *seekalpha.com*
- Training Corpus: reports from **Bloomberg** and **Financial Times.**, with 200 articles with epidemic-disease-related title and 600 articles published in non-epidemic period.
- Pretrained Data for BERT: BookCorpus (800 million words) and English Wikipedia (2500 million words), which is used to perform masked language modeling (MLM) and next sentence prediction (NSP) in pretraining the BERT model.

BERT: Bidirectional Encoder Representations from Transformers

- **Transformer** neural network architecture: a seq2seq model with encoder-decoder structure, where encoder's inputs first flow through a self-attention layer – a layer that helps the encoder look at other words in the input sentence as it encodes a specific word and learn dependencies between distant positions. Outputs of the self-attention layer are fed to a feed-forward neural network. The decoder part has both those layers, but the self-attention layer is only allowed to attend to earlier positions in the output sequence.
- Since BERT's goal is to generate a language model, only the encoder mechanism in Transformer is necessary. BERT applies the bidirectional training of Transformer to language modelling by masked language modeling (MLM), whose goal is to predict the original value of the masked words, based on the context provided by the other, non-masked, words in the sequence, and next sentence prediction (NSP).
- Fine-tune for BERT: classification tasks are done similarly to Next Sentence Prediction, by adding a classification layer on top of the Transformer output for the [CLS] token in Fig 1.

Sentence Classifier using Pretrained BERT model

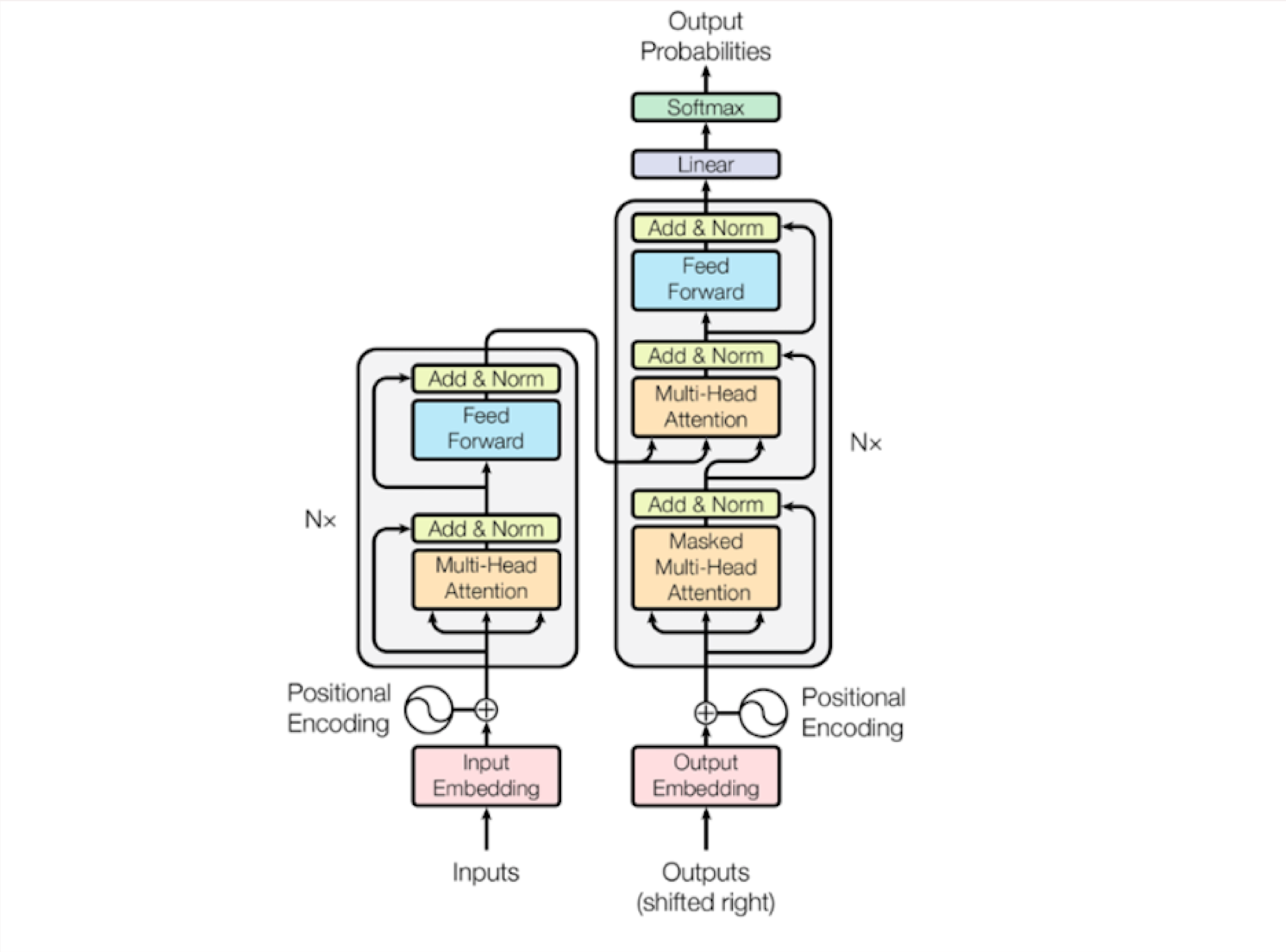

Fig. 1: Transformer Architecture

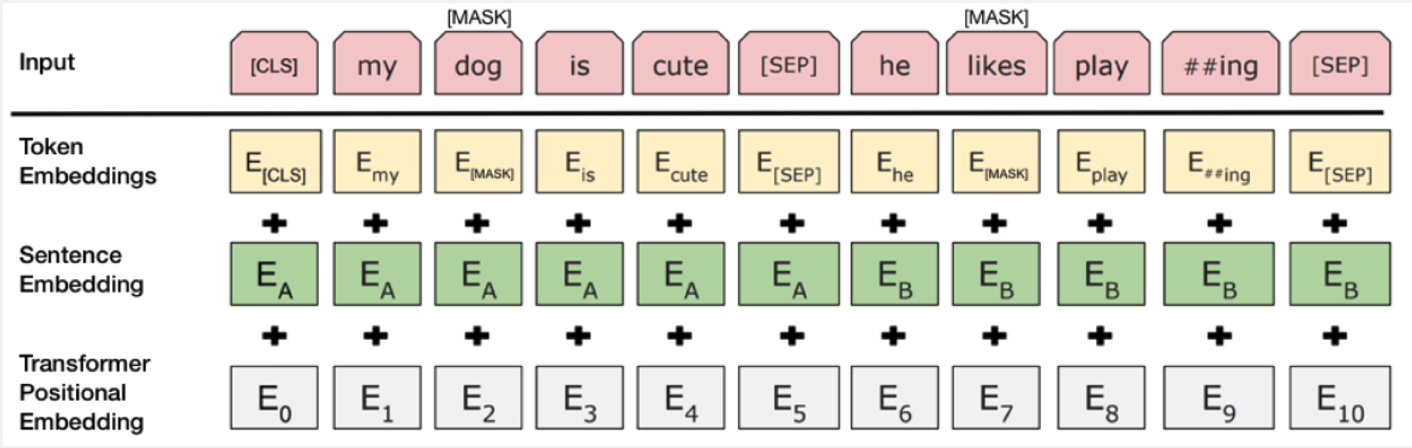

Fig. 2: BERT [Devlin et al.,2018

Our goal is to determine if a sentence is epidemic related in firm earnings call conference transcript. We adapt the following algorithms:

- Parse the news articles into sentences, removing the stop words.
- Split the dataset into Training, Validation and Test sets (20 : 1 : 10)
- Use the Transformers package **Simple Transformers**, which wraps the complex architecture of all BERT related models, to initialize, train and evaluate the model and obtain ready-to-go transformers.
- We experiment with three particularly prominent BERT-based models, RoBERTa, DistilBERT and ALBERT, each of which is specialized on a different task.
- Benchmark: applied two non-transformer model in classification: simple bidirectional neural network classifier with LSTM and SVM classifier, and compare performaces.

Firm-level Epidemic Risk Measurement

- Applied the trained classifier to parsed sentences in a firm's earning call conference transcripts.
- Denote the total number of sentences in transcript of firm i at time t as $S_{i,j}$, and the number of sentences labeled as epidemic-disease-related as $D_{i,t}$, then we take the measure of related risk as

$$Risk_{i,t} = \frac{D_{i,t}}{S_{i,t}}$$

- For industry-wise risk measurement, we sum up the nominator and denominator across all listed firms in a certain industry J, and take the division as risk measure:

$$Risk_{J,t} = \frac{\sum_{i \in J} D_{i,t}}{\sum_{i \in J} S_{i,t}}$$

Preliminary Results

We evaluate the performace of models based on three metrics:

- **Area Under Curve (AUC)** is a measure that takes into account the true and false positives. It measures the area under the (ROC) curve, which plots the false positive rate (specificity) against the true positive rate (sensitivity).
- **F1-Score (F1)** balances between precision (= True positives /(True positives + False positives)) and recall (= True positives /(True positives + False negatives))
- **Matthew's Correlation Coefficient (MCC)** uses all four measures - true positives, true negatives, false positives and false negatives - in its calculation.

The following result table was calculated based on the standard binary cutoff of 0.5. If the classifier's output probability greater than 0.5, the sentence is labeled as related to epidemic disease and marked as "1", otherwise it is marked as "0".

Evaluation	BERT	RoBERTa	DistilBERT	ALBERT	Bidirectional NN	SVM
AUC	0.8763	0.9011	0.8761	0.8757	0.8541	0.8561
F1-Score	0.425	0.4539	0.434	0.4717	0.37	0.47
MCC	0.4086	0.4528	0.4141	0.4456	0.29	0.41
Runtimes	18min	18min	16min	15min	20min	5min

Fig. 3: Classifier Result

The BERT-family models all perform better than the benchmark models. Their runtimes were similar around 18 minutes, with DistilBERT and ALBERT being slightly faster, and the SVM taking by far the least time with only 5 minutes runtime.

To be continued (by June 5th)

Apply the trained classifier on transcript data and get firm-level risk measure and industry risk measure.