BERT-based Firm Risk Measurement and Stock Return Volatility

by Zhou Xing

Apr 2020

Motivation

- The lack of measurement for firm-level risks, especially in specific categories such as politics, international trade, educational issues.
- Firms in the same industry are exposed to heterogeneous risks
- Effects of risk on behavior of households and firms might outweigh potential upside of well-meaning reforms
- Existing measurements are mostly based on "counting of words" without syntax and semantic embedding.

Introduction

- Develop a novel, firm-level, measure of political risk based on textual analysis of annual reports and conference call transcripts.
- Applied **BERT** model to capture syntax and semantic of sentences.
- Study association with cross-sectional variation of stock market volatility.

Former models and Contribution of this paper

- *Hassan et al.(2019)* extract all two-word combinations from training libraries that are indicative of discussion of political topics and count the number of occurrences of political "bigrams"
- *Gao (2016)* applies bag-of-words model to represent a document in vector space, and learn the relation of single word and risk according to stock market volatility changes.
- **This paper:** uses more comprehensive text analysis techniques to classify sentences, providing risk measure of documents that can be decomposed by topic.

BERT: a Pre-training Language Representation Model

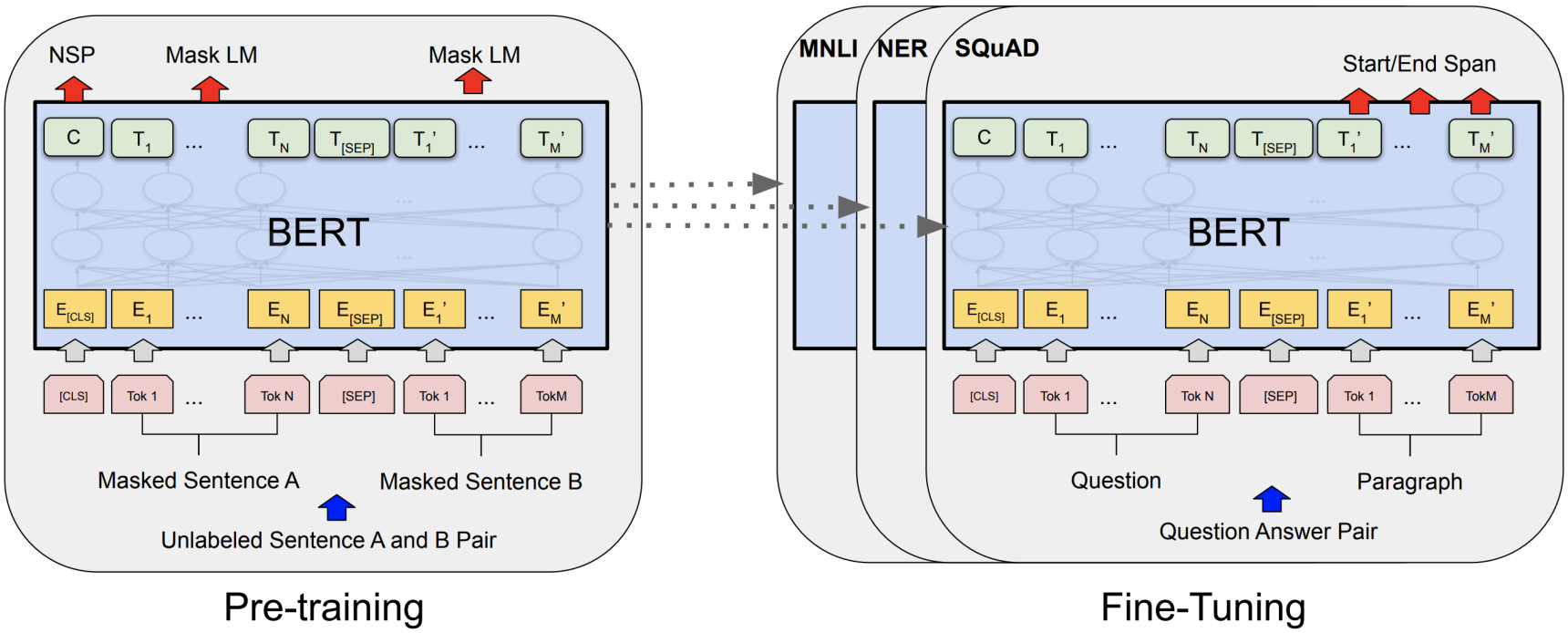

Figure 1: Overall pre-training and fine-tuning procedures for BERT. Apart from output layers, the same architectures are used in both pre-training and fine-tuning. The same pre-trained model parameters are used to initialize models for different down-stream tasks. During fine-tuning, all parameters are fine-tuned. `[CLS]` is a special symbol added in front of every input example, and `[SEP]` is a special separator token (e.g. separating questions/answers).

Training Libraries

- **Non-Risk Sentences:** Textbook on financial accounting
- **Risk-related Sentences:**
  - T extbook on American Politics
  - T ext contained in 8 topics from OnTheIssues.org
  - C ontains snippets from newspapers, speeches, press releases, books, voting records, and bill sponsorships identifying where candidates for political office stand on each topic (health care, environment, defense, ...)

Data

- Conference Call Transcripts: Complete transcripts of 175,797 earnings conference calls of US listed firms 2002-16 from Thomson-Reuters.
- SEC MDA annual (10-K) report

Beta Pricing Model

$$r^i_{t+1} = a^i + \beta^i_\tau f_{\tau,t+1} + \sum_{k=1}^{K} \beta^i_k f_{k,t+1} + \epsilon^i_{t+1}$$

- r^i_{t+1}: the excess return of stock i from time t to $t+1$.
- $f_{,t+1}$: the risk factor constructed using textual data
- $f_{k,t+1}$: the K controlling factors
- a^i : the industry control factor

The End