

BondSenti: BERT-based Sentiment Analysis for Bond Default Risk

Captain: Liu Yiwei

Team: Mechanized Heart

Sichuan University, Nanjing University

Abstract

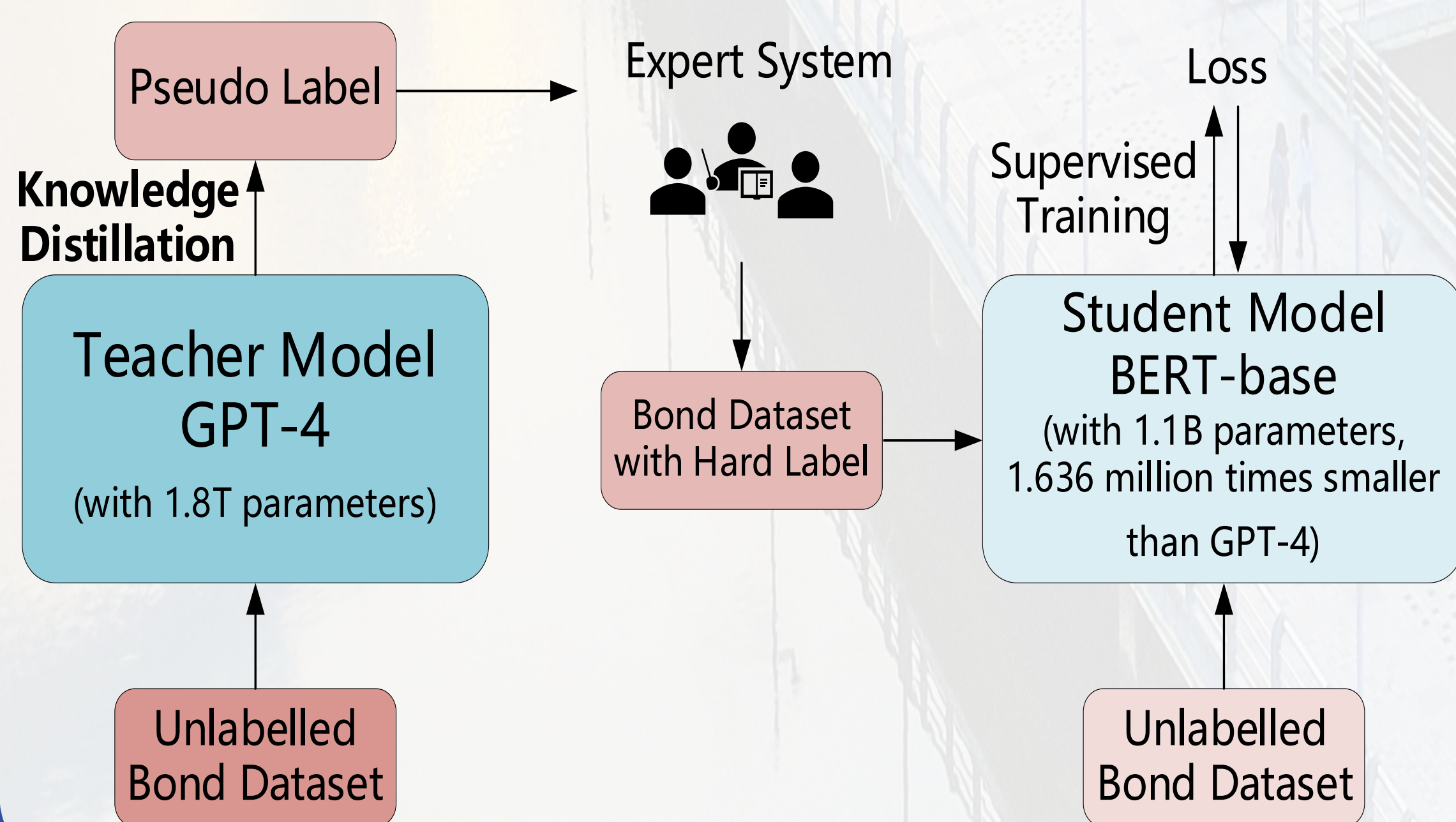
With the intensification of global economic fluctuations, the risk of bond default is also rapidly surging. However, the current methods of quantitatively predicting bond default risk only rely on structured data, ignoring the information embedded in unstructured text, which is sometimes quite significant. Therefore, we propose a BERT-based aspect-level sentiment feature extraction model for this very vertical domain. Specifically, we first utilize bond news sourced from RESSET to construct a dataset regarding bond entities and corresponding sentiments. Next, we employ BERT to perform named entity recognition tasks, following to utilize MLP to conduct sentiment analysis on the extracted individual entities. Finally, we conducted an empirical analysis of sentiment factors using Random Forest model. The experimental results indicate that incorporating sentiment factors on top of traditional structured indicators can improve the accuracy of default probability prediction by 3.2%.

Background

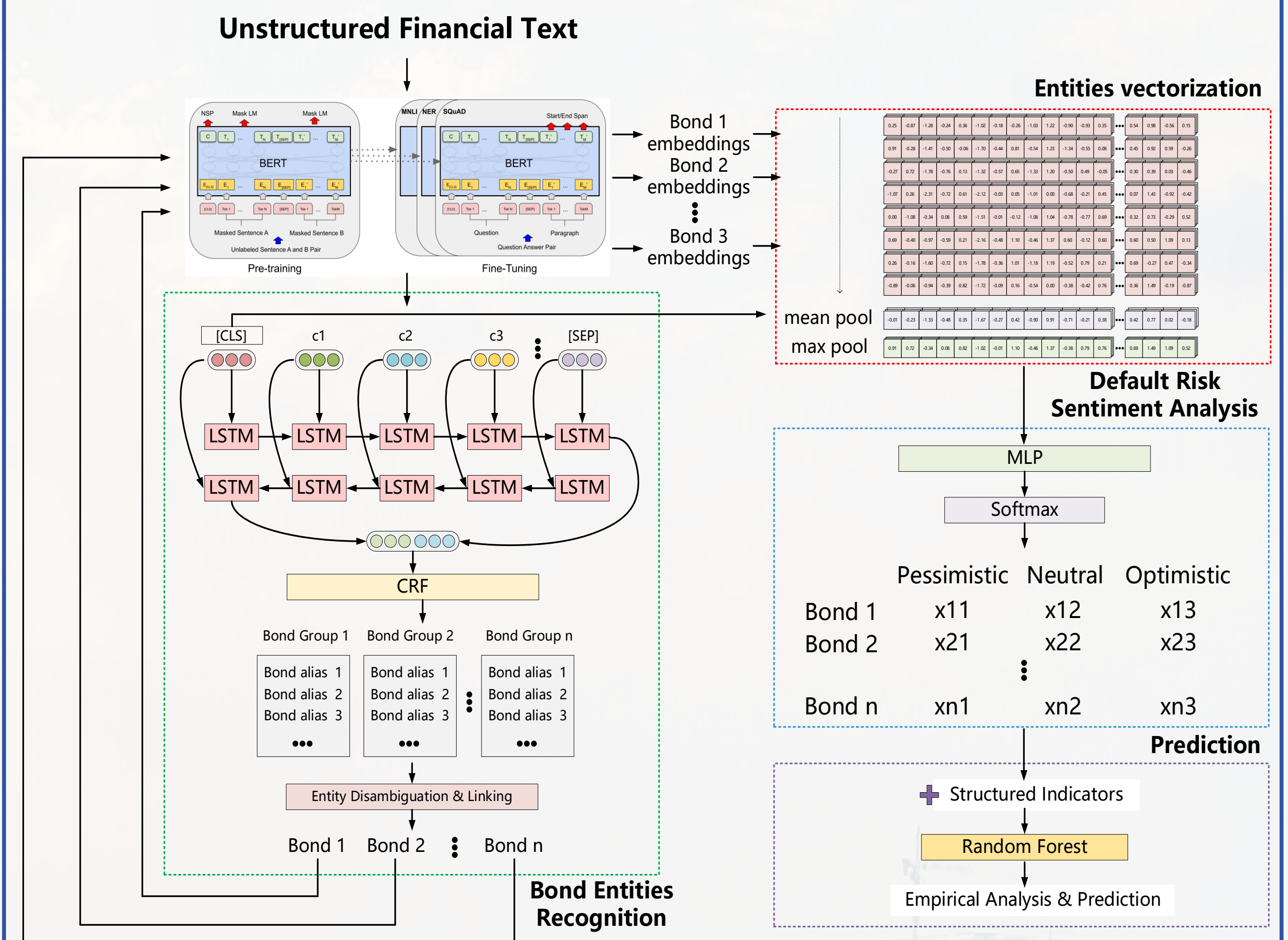
Since the release of ChatGPT, LLMs have sparked a new wave of technological revolution. The powerful representational capabilities of these LLMs give them significant advantages in the financial domain, such as swiftly and accurately extracting information. However, current LLMs both domestically and internationally mainly empower financial domain using generative dialogue. Their capability of addressing subtasks like sentiment analysis in vertical fields such as the bond market is lacking. Therefore, this project aims to fill the gap in the application of large models in bond default risk research through techniques like transfer learning.

Project Goals

1. Establish a high-quality sentiment analysis training text dataset in financial area from RESSET, iFind, etc.
2. Fine-tune a discriminative model based on BERT-base-Chinese to accomplish entity recognition for multiple bonds and corresponding bond default risk sentiment analysis.
3. Combine sentiment indicators with multi-source heterogeneous data to construct a prediction model for bond default risk.



Methods



1. Using BERT, BiLSTM, and CRF to complete the task of named entity recognition (NER), enabling the model to extract bond entities from unstructured text.
2. Using the extracted bond entities to locate their positions in the unstructured text after tokenizing, applying mean-max pooling to each entity separately to represent them as a individual tensor, and inputting each entity's tensor into MLP to accomplish aspect-based sentiment analysis (ABSA).
3. Building multi-source heterogeneous data by combining structured indicators of bonds from 2013 to 2023 with news text. Conducting empirical analysis on sentiment factor output from our model and utilizing these factors to enhance the probability of predicting bond default risks.

Results

The sentiment prediction accuracy of the ABSA model on the validation set is 90.6%. The empirical analysis demonstrates that incorporating sentiment factors, compared to using only traditional structured indicators like Shibor, improves the accuracy of predicting bond default risks by 3.2%.

User Segmentation

1. Financial institutions: Banks, securities companies, and asset management institutions requiring timely and accurate financial sentiment analysis and risk predictions to optimize portfolios and reduce transaction risks.
2. Investors: Individual and institutional investors seeking to use sentiment indicators and prediction models to guide investment decisions for higher returns and lower risks.
3. Financial media: Financial news websites and data platforms needing competitive news analysis and sentiment indicators to attract user traffic and increase market influence.