**SIG-FIN 第31回研究会メモ**

参加者：

**(07)** **T5を用いた決算短信の生成型要約**

決算短信と業績発表記事との対応を自動的に取得し、その対応を学習データとしてT5で決算短信の要約モデルを作成した。業績発表記事は決算短信を参考に日本経済新聞社が作成した記事とのこと。業績発表記事の各文と類似している文を決算短信から自動的に抽出して、決算短信の重要文を抽出するための学習データを作成した。まず、業績発表記事の文と決算短信の各文の類似度をSentence-BERTで計算し、類似度が最も高い文と最も低い文を抽出。それらを正例・負例としてBERTを学習し、そのモデルを用いて決算短信から重要文を抽出。抽出された重要文をファインチューニングしたT5で要約した。重要文抽出を行ったものとそうでないものを比較したところ、重要文抽出した場合の方の若干精度が良かったとのこと。精度評価にはBERTScoreを用いた。

（感想）

内容としては業務やっているものとの関連があり、わかりやすそうに感じた。T5のファインチューニングの仕方など良くわかっていないので、勉強したい。恥ずかしながらテキスト生成の自動評価に使われていたBERTScoreという指標を知らなかった。 調べたところBERTScoreを改良したMoveScoreというものもあるようなので、今後評価指標として使っていきたい。

**(08)** **ChatGPTを用いたニュース記事のESGトピック分析**

ロイターニュース記事から、抽象度の異なる３種類のESGトピックを抽出するというもの。ChatGPTを用いて、プロンプトエンジニアリングでトピックを自動抽出し、概念抽象度が異なるトピックの抽出が可能であることを示した。プロンプトには出力例を１つだけ示すOne-shot promptingを用いた。

（感想）

プロンプトをどうするかという話ではなく、どうやって抽象度が異なるトピックを抽出できたと定量的に評価するかという評価手法に主眼が置かれた内容のようであった。何をもって精度を評価すべきかいうのは常に頭を悩ます問題なので、もう少し細かく読んでみたい。

**(11)** **BERTおよびChatGPTを用いたサステイナビリティレポートからのSDGs関連文抽出**

SDGｓに関連する文が、17のゴールのどれに相当するかを分類するというもの。データとして国連のSDGｓの定義の日本語約や経団連の事例集などを用いた。データ数が少ないため、固有名詞を類義語でランダムに置換するというやり方で、データオーギュメンテーションを行った。また一部のゴールに関する文は非常に少なかったため、ChatGPTでデータを生成して補完した。それらのデータを用いた、日本語事前学習済みのＢＥＲＴで分類モデルを作成した。ChatGPTでデータ補完を行った場合とそうでないものを比較し、サンプル数が少ない場合はデータ補完により精度が向上することが示された。

（感想）

　ChatGPTを学習データの水増しに使うという発想がこれまでなかったので、このような活用方法があるのだと勉強になった。確かに同じ意味で表現を変えるということは得意そうなので、単に固有名詞を置換するというよりは良さそうに感じる。

**(12)** **ChatGPTを活用した運用報告書の市況コメントの自動生成**

ChatGPTを活用し、運用報告書における市況コメントと見通しを自動生成するというもの。まず期間中のリターンを計算し、期間中上昇した場合はリターン最大、下落した場合はリターン最小の日付付近に関するロイターニュースを抽出。それらのニュースをChatGPTのプロンプトに一緒に入れて、変動要因を抽出させた。次に、リターンと変動要因の要約をChatGPTのプロンプトに一緒に入力し、市況コメントを生成させた。最後に、これまでの会話履歴も含め、今後の見通しコメントを生成するよう指示した。評価は市況コメントについてのみ行い、人手による定性的な評価とSentenceBERTのコサイン類似度を用いた定量的な評価を行った。ある程度、運用者が書く市況コメントに近いものを生成できたものの、一部ニュース記事の少ないアセットではHallucinationとみられる記述が多くみられるなど、課題はあるようであった。

（感想）

　定量評価でSentenceBERTと並んでSentenceLUKEというモデルが用いられていた。とあるデータセットでは、SentenceBERTよりも精度が高いようなので、今後比較のために利用したい。

（その他）

海外から参加していた人がYouTubeで感想戦を配信していたので、ご参考まで。（みずほの発表も取り上げてくれて）

<https://www.youtube.com/watch?v=UeLkLCjqcSI>

なぜかこのタイミングで前回の発表取り上げてポストしてくれている人もいました。

https://twitter.com/Yoshihiko\_ICKW/status/1712108939943547308

以上。