Q. livedoorデータセットを使った他の研究と比較して精度はどうか

A. 近年の言語処理学会などでこのデータセットを使った論文ではデータ拡張やデータ分布の可視化に関するものが多く、分類性能を記載してないものが多かった。また、研究によって扱うラベルや、試行回数、入力形式が異なるため単純な比較は難しいのですが、提案手法１における96.5%の精度は十分な値だと思います。ただ今回は同じ条件下で大和の手法と比較して精度が向上したという結果だけ得られたのでこれについて報告しました。

Q. 他に使えそうなデータセットは無いか.

A. 要約文の生成コストを気にしないなら一般的な分類タスクで用いられるデータセットは扱えると思います. 正しい要約文がセットでなっているものは少ないのが現状です.  
  
Q. 要約文の妥当性について、検討はしてるか

A. 現状, 主観的に見て原文と大きくずれてはいないことと bertscoreによる統計でしか妥当性を確認していません。ハルシネーションなどの問題もあるので, LLM API を用いてより良い要約文を生成する手法、プロンプトに関してはまだ検討段階です。  
  
Q. アテンション、ストップワードにも付いているみたいだけど？

A. bert の出力先でのスケーリングが行われているので, 順伝播時にbertはそのscoreを知りえないことに起因すると考えられます。スケーリング層の先にもう一つTransformerを追加してその中でまたアテンションを見るという解決策が考えられます。  
  
Q. 要約文生成と分類精度で違うplamo使っているのはなぜ？  
  
Q. ストップワードどんなのがあるの？  
  
Q. TFIDFの定義からこの式意味あるの？  
  
Q. plamoってgpt-4o-miniと比較してどうなの？  
  
Q. タイトルだけで分類したら？clsとavgだけでやった場合の結果は？  
  
Q. 要約文の生成コスト、トークン数について

A. 1データにつき入力トークンだけで平均2000ほどあるので, データ数掛け合わせると入力トークンだけで1500万トークン、出力合わせたら2000万トークンは行くと思います.  
  
Q. どういうデータ、クラスにおいて間違いが多かったか

Q. もう一度、目的、結果、何が新しいのか、説明して

Q. 固有名詞だけの抽出だけで事足りるのでは？

Q. 提案手法の要約文の品質が低い場合、逆に精度が悪化する可能性は？

Q. 提案手法は CAP の拡張であるが、CAP の問題点をどこまで解決できているのか？何が新しい？

Q. 自然言語処理において CAP みたいなプーリング手法は他にあるか？

A. Window\_sizeを設定して複数のインデックスから抽出していく手法や、各Transformerの出力を並べることで画像のように扱い、それにCNNを適応する手法など

Q. 提案手法の新規性は本当にあるのか？単に CAP に要約文を足しただけでは？

Q. LLM API の進化が進めば、結局本手法の有用性は失われないか？

Q. なぜT5などのローカルLLMで要約文を生成しなかった？

A. データセットに、十分な長さの正しい要約文としてのデータが無かったため、ファインチューニングができなかった。教師なしで精錬された要約文を生成するためにはLLM の API が現状理にかなっていると判断したから。  
  
Q. 自然言語処理における独自のプーリング手法は画像処理分野と比べて数少なく,

その効果に関する理解は不十分であるのはなぜ？

1. その理由としては、単語は数値的に連続したデータに直接変換することが出来ず、言語モデルを介して数値表現に変換しても、単語の埋め込み表現の各次元の間には数値的な大小が存在しないことが大きな理由。また、単語ベクトルを操作することでかえってノイズになってしまう。

Q. BERT を使い続ける理由

A. 本研究の目的は大和の手法に対して, 埋め込み表現のプーリング手法の改善を第一目標としているので, ベースとなる埋め込み部分に関しては大和の手法と変えずに比較した. もちろん最新のモデルで埋め込み表現を得ることで精度が上がることが推測されるが, モデルの性能と計算効率の観点から本研究の有意性を吟味したい。できるだけ小さいモデルの中で原文の情報を濃くしていきたい.