修士学位論文　発表原稿
**それでは BERT を用いた原文と要約文の**
**分散表現の統合手法と題しまして**
創発ソフトウェア研究室 M2 の 高山が発表いたします.

発表の流れはこのようになっております.

はじめに

近年、**大規模言語モデル, LLM** の進化に伴い、
自然言語処理分野においては、
BERT や GPT といった
LLM を活用したシステムに注目が集まっています。
そのため、より高性能なモデルを構築するためには、
文章全体の**適切な分散表現**を獲得し
それを**最適に処理する手法の選択**が重要です

そこで本研究では、分散表現の処理方法の 1 つである**プーリング**に注目します.
まず、一般的に自然言語処理分野で用いられているプーリング手法について説明します。
1 つめは、入力データの**文頭**を表す CLS トークンの埋め込み表現,
ベクトル E_cls を文の分散表現として用いる手法、
そして､2 つめは各単語トークンの埋め込み表現の平均ベクトル E_avg を文の分散表現として用いる手法です。
**適切なプーリング手法の選択はモデルの性能に直接影響を与える**ため、
非常に重要な要素であるのに対して
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という背景があります。

このような背景を踏まえて
先行研究として、大和は **CLS-Average Pooling（CAP）を提案しました**
これは学習可能な和が 1 となる非負のパラメータ $p,\ q\ (\geq 0)$ を用いて
$E_{[CLS]}$ 及び, $E_{Avg}$ の重み付き和 C= $pE_{[CLS]} + qE_{Avg}$ を
文の分散表現とするプーリング手法であり
テキスト分類タスクにおいて
$E_{[CLS]}$ や $E_{Avg}$のみを用いた場合よりも高い性能を示しました

また、更新後のパラメータを $p', q'$とすると更新式は

以下の式で表されます。

修士学位論文　発表原稿
ここで、本研究の目的と概要について説明します。
本研究では、
**より適切な文章の分散表現を獲得し、**
**原文に含まれる重要な文脈情報をより効果的に学習する手法の確立　を目標とし、**
この目標を達成するために今回は
まず、LLM の API を用いることで原文から要約文を生成し、
次に、大和による手法をベースとした**要約文情報を組み込んだプーリング手法を提案**します。
そして、テキスト分類タスクを解くことで提案手法の有効性を検証しました。

**【2:20】**
次に要素技術について

本研究ではテキストの埋め込み手法として BERT を用いました。
詳細については時間の都合上省略します

PLaMo は Preferred Elements 社が開発している日本語に特化した大規模言語モデルで
既存モデルをベースに用いず合計 2 兆トークン分のテキストデータで事前学習 が行われています。
本研究では,要約文の生成において plamo-beta モデル
分類精度の検証には‘plamo-ワン-prime’モデルを使用しました.

**【2:50】**
次に提案手法について

まず、本研究で用いるデータセットについて説明します。
本研究では livedoor ニュースコーパスデータセットを使用しました.
これは 9 つの ニュースカテゴリ，計 7,367 個の記事からなるデータセットで
各記事データ D はカテゴリ名 $D_{label}$，記事タイトル $D_{title}$，記事本文 $D_{body}$ などのメタデータを持ちます。
また本研究では以下のようにラベルとカテゴリ名を対応させています

次に LLM　API を用いた要約文の生成手法について説明します
まず、各記事データ D について,
D_{title}と D_{body}の情報から図のようなプロンプト D_{prompt}を作成します。
そして, PLAmo API から得られた回答を記事データ D に対する要約文 D_{summary}とします。
**このとき、温度パラメータは 0 としました.**

次に**大和の手法をベースとした要約文情報を組み込んだプーリング手法**について説明します
本研究では 2 種類のモデルを提案します
提案手法 1 は **BERT の並列化による要約文情報の統合手法モデル**です
これは、大和の手法に対して、赤枠に当たる部分を追加したものです。
この赤枠の中では、左側と同様に、要約文に対する分散表現を並行して学習します
**各 BERT モデル**への入力形式は図の通りです.

修士学位論文　発表原稿

そして学習可能な和が 1 となる非負の*パラメータ*$p,\ q,\ r, s$を用いて

これら 4 つのベクトルの重み付き和 C を入力文の分散表現とします

また、学習可能なパラメータ p, q, r ,s の更新式は

大和に倣って更新後*のパラメータを* $p', q', r'$, s'とすると

以下の式で表されます。

提案手法 2 は**スケーリング層を導入した要約文情報の統合手法モデル** です
これは、大和の手法に対して, BERT モデルの出力先に スケーリング層を組み込んだモデルです。
このスケーリング層では bert から得た単語ベクトル E_i に対して、
tok_i に対応するスコアを掛け合わせることで
スケーリング後の単語ベクトル E ハット i を得ます
ここで、score は学習前に事前に計算可能な固定の重みであり、カスタムストップワード辞書を用いて算出されます。
**これにより、提案手法 1 と比較してモデルのパラメータ数を増やすことなく、要約文情報の統合が期待できます**

*Score の計算方法は*
*Tok が[CLS], [SEP]の場合は* score は 1 となりスケーリングは行いません.
*Tok がストップワードであればペナルティとして正の定数パラメータ* $\alpha$ *を用いて* $-\alpha$ *を返します.*
*また、Tok がストップワードではない場合、score は 以下の式で表されます。ここで*

$tfidf$*項*は、モデルへの入力系列と対応する要約文の 2 つのサブワード列から算出した

要約文におけるトークン i の tfidf スコアです.
ストップワード辞書には spaCy, slothLib で定義されているものに加えて句読点や記号, **[PAD]を設定しました。**
**本研究では $\alpha$ が 1 または 5 の場合について実験を行いました。**

**【6:00】**
次に数値実験について説明します

改めて、
本研究では livedoor ニュースコーパスデータセットを用いた 9 クラスの記事ラベル推定を行い、提案手法の有効性を確認することを目標とします。
訓練データ：検証データ：テストデータ の比率が 8 : 1 : 1 となるように分割し、
検証データにおける F1 値 が最も高い epoch におけるモデルを用いてテストデータで評価を行いました。
分類器は 1 層の全結合層とし、 各 bert モデルは最終層のみをファインチューニングしました
ラベルごとの細かいデータ数は表のようになっています。

修士学位論文　発表原稿
また、**分類精度の比較対象として、大和による従来手法に加えて**
比較手法 1 として要約文のみを入力とした場合,

そして、比較手法 2 としては‘plamo-ワン-prime’ および ‘gpt-4o-mini’ のみを用いた
Zero-Shot, Few-Shot による分類です。プロンプトや手法の詳細については時間の都合上省略します。
LLM が選定した**上位 3 つのラベル**を JSON 形式で回答してもらい、その中に正しいラベルがあれば正解としました。
**このとき、温度パラメータは 0 としました.**
LLM API のみで十分な精度が得られる場合，学習コストの観点から,
**提案手法のような複雑なモデルを構築し、運用する必要性が失われてしまう可能性があるため，これについて検討しました.**

実験で用いたパラメータは表の通りになります。

3 回試行の平均結果を示します。カッコの中は標準偏差を表しています。
比較手法 2 には最も性能が高かった gpt-4o-mini で Few-Shot を行った結果を示しています。

表より、各評価指標について
提案手法 1 および提案手法 2 は従来手法を上回り、テキスト分類タスクにおける有効性を示すことが出来ました。

次に比較手法 1 の結果より要約文のみで分類を行った場合、訓練用データに対しては他の手法と同等に学習が進みましたが、
汎化性能が上がらず精度が下がる結果となりました。この理由に関しては、
要約文を生成するためのプロンプト内で回答の出力形式を自由にしていたため,
要約文データの一貫性を担保できていなかったことが原因だと考えられます。

**また、比較手法 2 の結果より提案手法は現時点での LLM API のみを用いる手法と比較して，有効性のあるアプローチであり，実用的な意義はあると考えられます.**

**提案手法 2 の結果では、アルファ=5 の場合の方が高い分類性能を示しました。**
**これは score のノルムが大きくなったことによって**
**ストップワードとそれ以外の境界がより明確になったことに起因すると考えられます。**

また、提案手法 2 の結果は提案手法 1 の結果には劣りますが、
モデルのパラメータ数は約 50%であることを踏まえると、**分類性能と計算効率のバランス**を考えた際に
有効な選択肢であると言えます。

修士学位論文　発表原稿
次に提案手法 1 に対する考察を述べます。
左が提案手法 1 における重みパラメータ p,q,r ,s の推移で右が従来手法における p,q の学習推移を表しています。
**それぞれ、横軸がエポック数、縦軸が重みパラメータの値を示しています。**
**青が p, 黄色が q, 緑が r,赤が s の推移を表しています。**
図より、どちらの場合も原文の cls ベクトルに対する重み p が単調に増加し、
原文と要約文の平均プーリングベクトルの重み q、 s が減少傾向にあることから、
**原文情報を含むテキスト分類タスク**において
[CLS]トークンに対応する分散表現 Ecls の方をより重要視する傾向が強いことが分かりました。
ここで, もしこのタスクが Ecls のみに強く依存しているのであれば,
提案手法 1 において E'cls の重み r に関しても単調に減少していくと考えられますが、
左図のように, r の値は一定の値まで減少した後, わずかに増加する傾向が見られました。
**このことから学習の後半において要約文の cls ベクトルが追加の情報源として再評価され,**
**重みがわずかに増加したことで汎化性能が上がり分類精度の向上に寄与したと推測できます.**

次に提案手法 2 における結果の考察を述べます。
提案手法 2 における重みパラメータの学習推移は従来手法と比較して
p,q の学習傾向が逆転しており、スケーリング層の中で各単語の重要度が事前に調整されることで,
$E$_Avg に含まれる情報の精度が向上し, モデルがより重要視するようになったと考えられます。

実際にスケーリング層における重み付けによって、
推論時にどのような変化が生じたかについて例と共に示します。
これは、bert の最終層におけるアテンションを可視化したもので、
上が従来手法、下が提案手法 2 でアルファ=5 の場合を示しています。
特に、提案手法 2 において強くアテンションがかかるようになったものには
Android やカメラアプリ、LINE といった固有名詞が挙げられ、**定性的に見て**
提案手法が**原文に含まれる重要度の高い単語**を選択的に強調しながら
最適な重みパラメータを学習できることが分かりました

**[11:10]**
最後に、まとめと今後の展望です

本研究では LLM API を用いて原文データから要約文を事前に生成し,
要約文情報を先行研究である CLS-Average Pooling に
組み込んだプーリング手法について提案しました
そしてテキスト分類タスクにおいて Accuracy, F1 値ともに
提案手法が従来手法を上回り, その有効性を確認することができました。

今後の展望としては,要約文に対する妥当性の検討および生成手法の改善,
他のタスクやデータセットにおける提案手法の有効性の検証,
提案手法 2 におけるパラメータ $\alpha$ のチューニングなどが挙げられます。
以上で発表を終わります。ご清聴ありがとうございました。**【11:50】**