近年、深層学習をはじめとする機械学習手法 は目覚ましい進展を遂げ、自然言語処理（NLP）やコンピュータビジョン（CV） などの分野で高い精度を達成している。特に NLP の分野において、Transformer をベースとした手法はさまざまなタスクで顕著な成果を上げており、大量のテキストデータで事前学習された Transformer を基盤とする大規模言語モデル（LLMs） が注目されています。そのため、より高精度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。

より高精度なモデルを作るための深層学習の要素技術の 1 つにプーリングという手法がある プーリングとは，入力されたデータから得た特徴量の次元を縮小し，計算量を削減するとともに，抽出された特徴量のロバスト性を高めるための手法である。しかし，自然言語処理分野における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る。

Recent advances in machine learning, particularly deep learning, have achieved high accuracy in Natural Language Processing (NLP) and Computer Vision (CV). In NLP, Transformer-based models have shown exceptional performance across various tasks, leading to the rise of Large Language Models (LLMs), which are pre-trained on massive text corpora. Therefore, constructing high-accuracy models requires effectively capturing sentence-level distributed representations and selecting an optimal processing method.

Pooling is a fundamental deep learning technique that reduces feature dimensionality, improving computational efficiency and robustness. However, in NLP, pooling methods remain less explored than in CV, and their effectiveness is not well understood.

このような背景から，大和は LLMs の 1 つである Bidirectional Encoder Representations from Transformers (BERT) において一般的に用いられる［CLS］トークンの埋め込み表現を用いたプーリング手法と，平均プーリング手法を組み合わせた CLS-Average Pooling (CAP) を考案し，テキスト分類タスクにおける各プーリング手法のみを用いた場合と比較して，その有効性を示した。CAP では、学習可能で和が 1 で一定である 2 つの非負のパラメータを用いて、これら 2 つのプーリング手法から得られたベクトルの重み付き和を文章の分散表現としている。

そこで本研究では，大和による手法を基に，LLM を用いて原文から生成した要約文の分散表現を追加で組み込んだプーリング手法を提案し，テキスト分類タスクにおける性能向上を目的とした．

###

Based on the above background, Yamato proposed CLS-Average Pooling (CAP), a pooling method that combines 2 widely used techniques in BERT, a type of LLM: the embedding representation of the [CLS] token and average pooling. Their study demonstrated that CAP is more effective than using either pooling method alone in text classification tasks. In CAP, 2 trainable non-negative parameters, whose sum is fixed to 1, are used to compute the weighted sum of the vectors obtained from these 2 pooling methods as the sentence's distributed representation.

In this study, I propose a novel pooling method that further incorporates the distributed representation of summary texts generated from the original text using an LLM. By integrating this additional representation, the proposed method aims to improve classification accuracy in text classification tasks by enhancing semantic feature aggregation.

図 1 は提案手法の全体モデルの概要を示しています。提案手法では、まず独立した 2 つの事前学習済 BERT モデルを用いて、入力された原文とその要約文のそれぞれに対して[CLS]トークンの埋め込み表現（$E_{[CLS]}$, $E'_{[CLS]}$）と、平均プーリングによって得られた埋め込み表現（E_{Avg}, E'_{Avg}）を学習します。次に proposed pooling layer の中で、学習可能な重みパラメータを複数個用いてこれら 4 つのベクトルの重み付き和を文章の分散表現とします。そして、proposed pooling layer に対していくつかの設定を与え、それぞれの条件下で先行研究である大和に倣って同様のテキスト分類タスクの実験を行いました。

Figure 1 shows an overview of the entire model of the proposed method. It utilizes 2 independent pre-trained BERT models to extract [CLS] token embeddings—denoted as $E_{[CLS]}$ and $E'_{[CLS]}$ —and average pooling embeddings—denoted as E_{Avg}, E'_{Avg} —for both the input original texts and its summary. Within the proposed pooling layer, multiple trainable weight parameters are introduced to compute the weighted sum of these 4 vectors, producing the final distributed representation of the sentence C. Then, I applied multiple configurations to the proposed pooling layer and conducted text classification experiments under each condition, following the prior work of Yamato.

実験の結果、提案手法は従来手法と比べて高い分類精度を獲得した。

As a result of the experiments, the proposed method achieved higher classification accuracy compared to previous methods.

画像変更案 (Input の部分、詳細に書く余裕ないので)

Input: Original Texts

Input: Summary Texts

# # # # # # # # # # # # # # # # #
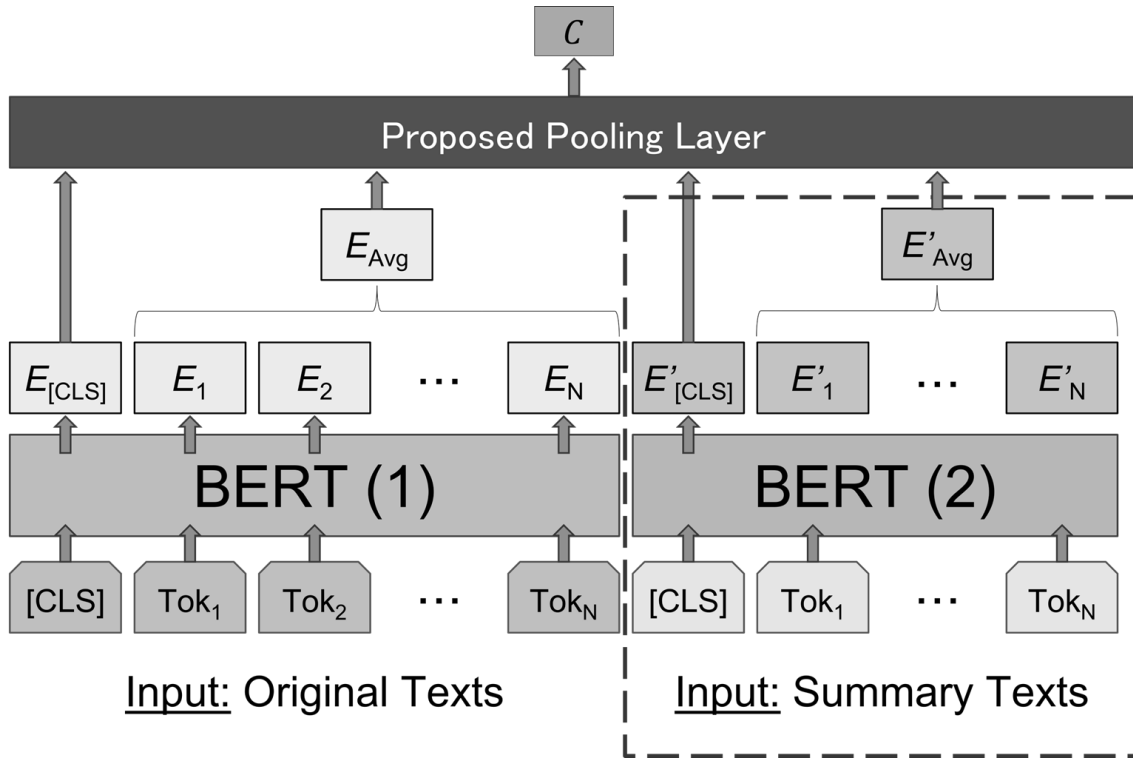
(以下添削前)

In recent years, machine learning methods such as deep learning have made remarkable advancements, achieving high accuracy in fields such as Natural Language Processing (NLP) and Computer Vision (CV). In the field of NLP, Transformer-based methods have demonstrated outstanding performance across various tasks. As a result, Large Language Models (LLMs), which are built upon the Transformer architecture and pre-trained on massive text corpora, have gained significant attention. To construct higher-accuracy models, it is crucial to effectively capture the distributed representations of entire sentences and to select an optimal processing method that maximizes their effectiveness.

One of the fundamental techniques in deep learning for building higher-accuracy models is pooling. Pooling reduces the dimensionality of extracted features from input data, lowering computational complexity while enhancing feature robustness. However, in the field of NLP, pooling techniques are far less explored compared to those in CV, and their effectiveness remains insufficiently understood.

Figure 1 shows an overview of the entire model of the proposed method. The proposed method first utilizes two independent pre-trained BERT models to obtain embedding representations of the [CLS]

token—denoted as $E_{\text{[CLS]}}$ and $E'_{\text{[CLS]}}$ —as well as embedding representations obtained via average pooling—denoted as E_{Avg}, E'_{Avg} —for both the input original texts and its summary. Next, within the proposed pooling layer, multiple trainable weight parameters are introduced to compute the weighted sum of these four vectors, forming the final distributed representation of the sentence C. Then, I applied multiple configurations to the proposed pooling layer and conducted text classification experiments under each condition, following the prior work of Yamato.