2023年度（令和5年度）　修士論文

補助損失を利用したBERTによる段落分割

令和5年2月

鳥取大学大学院　持続性会創生科学研究科
工学専攻　情報エレクトロニクスコース

自然言語処理研究室

M21J4023Y　ZHUO　BINGGANG

概要

段落分割は, テキストセグメンテーションタスクの一種である. 飯倉らは, BERT の基本的な使い方の上で Focal Loss を導入することで, 段落分割の性能を改善した. 本研究は, 毎日新聞および小説データセットにおける段落分割を調査した. 飯倉らの手法の上で, 我々は補助損失を導入してモデルをトレーニングし, 性能をさらに改善した. その結果, 飯倉らの手法によって得られた毎日新聞における平均 $F1$ スコアは 0.6704, 我々の手法では 0.6801 であり, 約 1% の性能改善を達成した. 小説データセットでも性能の改善が確認された. さらに, 統計的な検定より, 2 つの手法によって得られた結果の差が統計的に有意であることが分かった.

目 次

表 目 次

図 目 次

第1章　はじめに

文章の形式段落を推定する問題は, 対象とする 2 文が同一の形式段落に所属するかどうかという二項分類問題として捉えることができる. 本研究の主要な目的は, 自動段落分割の性能向上である.

自動段落分割は, テキストの読みやすさを向上させることができるだけではなく, 言語モデルの性能指標にもなる. 近年, 自然言語の分野では数多くの言語モデルが出てきた. ある言語モデルが過去の言語モデルより優れていると言う場合, そのモデルがあらゆるタスクで過去のモデルよりも優れるべきであり, 段落分割タスクは無視できない指標である.

自動段落分割の分野にある最近の研究では, 飯倉ら [1] は, BERT に Focal Loss を導入することにより, 優れた性能を達成した. BERT の一般的な使用法に基づいて, 性能を向上させるために, 飯倉らはバイナリクロスエントロピー (BCE) 損失を Focal Loss に置き換えた. 飯倉らの手法の詳細を 2 章で説明する.

Focal Loss [2] は, モデルの自信過剰な予測にペナルティを課すことにより, データ不均衡の問題を軽減できる損失関数である. データ不均衡とは, 分類問題におけるラベルの各クラス数の比率が大きく偏っている状況である. BERT [3] とは, 複数のエンコーダー層と自己注意ヘッドを備えた Transformers に基づく [4] 言語モデルである. BERT は, 近年, さまざまな自然言語処理タスクで優れた性能を達成することが証明された. ニューラルネットワークの性質上, BERT が優れる性能を達成する理由を特定することは困難だが, Clark ら [5] は, BERT が自己注意メカニズムによって大量な構文情報を学習したことを指摘した.

本研究では, 飯倉らの研究手法 [1] に基づいて, 段落分割問題を研究した. 優れた性能を達成したが, 飯倉らの研究では 2 つの問題点がある.

まず, 彼らの手法が依存している Focal Loss は, データ不均衡問題が大きい小説データセットには効果的だが, 他のデータセットにも同様に効果的かどうかは不明である. そのため, 本研究は 2 つのデータセットで段落分割問題を調査した. 1 つ目は毎日新聞データセットで, 2 つ目は飯倉らの研究と同様, 夏目漱石の小説データセットである. 小

説データセットは, 毎日新聞データセットよりも高いデータの不均衡を示している.

次に, ニューラルネットワークモデルの共通問題点として, モデルが分割点の周囲の文脈に注目する傾向があり, 距離の遠い情報を有効に利用できない. そこで本研究は補助損失を提案し, モデルが距離の遠い情報に注目させることを試みた. 補助損失については, 3.3 節で詳しく説明する.

本研究の主な主張点を以下に整理する.

- 本研究の新規性は, 飯倉らの研究の上で補助損失を導入し, 手法を改良することである. この改良は, モデルのパフォーマンスの改善につながる. その結果, 飯倉らの手法によって得られた毎日新聞における平均 $F1$ スコアは 0.6704, 我々の手法では 0.6801 であり, 約 1% の性能改善を達成した. 小説データセットでも性能の改善が確認され, 飯倉らの手法の平均 $F1$ スコアは 0.8261, 我々の手法は 0.8339 であった.

- 統計的検定で提案手法の優位性が統計的に有意であることを確認した.

- 特殊トークン配置の違いによるモデルの性能変動を調査した. 提案手法のように, 両側に補助的な [SEP] トークンで文の分割点をマークする場合, モデルの性能が基本配置の 0.6684 から 0.6797 に向上した. この結果は, BERT を利用する今後の研究の参考になる.

本論文の構成は以下の通りである. 第 2 章では, 本研究に関連する研究としてどのような研究が行われてきたかを記述し, 幅広く先行研究を紹介する. 第 3 章では, 提案手法を構成部分に分けて詳細な説明を行う. 第 4 章では, 性能指標とデータセットなど, 実験設定を記述する. 第 5 章では, 実験と有意差検定の結果を記述する. 第 6 章では, 幅広い考察について記述する. 第 7 章では, まとめを行い, 今後の課題を記述する.

第2章　先行研究

この章では先行研究を紹介する.

図 2.1 は, 自然言語処理の分野におけるタスクを整理したものあり, 図の示すように, 段落分割とトピック分割は, テキスト分割の分野に属し, 非常に類似している手法である. 二者の分割の粒度も類似しているため, 教師あり学習であれば基本的にデータセットを変更するだけで手法をそのままで使える. その他, 段落分割と比べれば, トピック分割に関する研究はより豊富で成熟している. それゆえ, トピック分割にある先行研究を無視できない.

2.1　トピック分割

テキストセグメンテーションの主要な分野の 1 つとして [6], トピックセグメンテーションは段落分割よりもはるかに注目されている. さまざまな研究で提案されているトピックセグメンターは, 内因性手法と外因性手法の 2 つの大きなカテゴリーに分類できる [7]. 内因性手法はテキストの表面 (text surface infomation) の情報に依存し, 外因性手法は深層的な意味 (semantic) 情報 (通常外部モデルによって導入される) に依存する. テキスト表面の情報から素性を抽出する機械学習手法は内因性手法と見なす.

内因性手法は, Halliday と Hasan の研究にまでさかのぼることができる [8]. 彼らは, 同じトピックに属するテキストフラグメントが似ている語彙を持つと指摘する. この指摘に基づく手法の中には TextTiling[9] と C99[10] があり, どちらも語彙の重複 (lexical repetition) に基づく古典的な内因性手法である.TextTiling は, 2 つのテキストブロック間のコサイン類似度によってセグメンテーション境界を決定し, C99 は, 文の間のコサイン類似度マトリックスの上でクラスタリングすることによってセグメンテーション境界を決定する.

他には, LCseg [11], F06 [12], TopicTiling[13] などの内因性手法がある. 語彙チェーン (lexical chain) と機械学習手法に基づく LCseg は, 内因性手法の中で最も性能の高いメソッドと報告される [7].F06 と TopicTiling はどちらも TextTiling に基づいている.F06

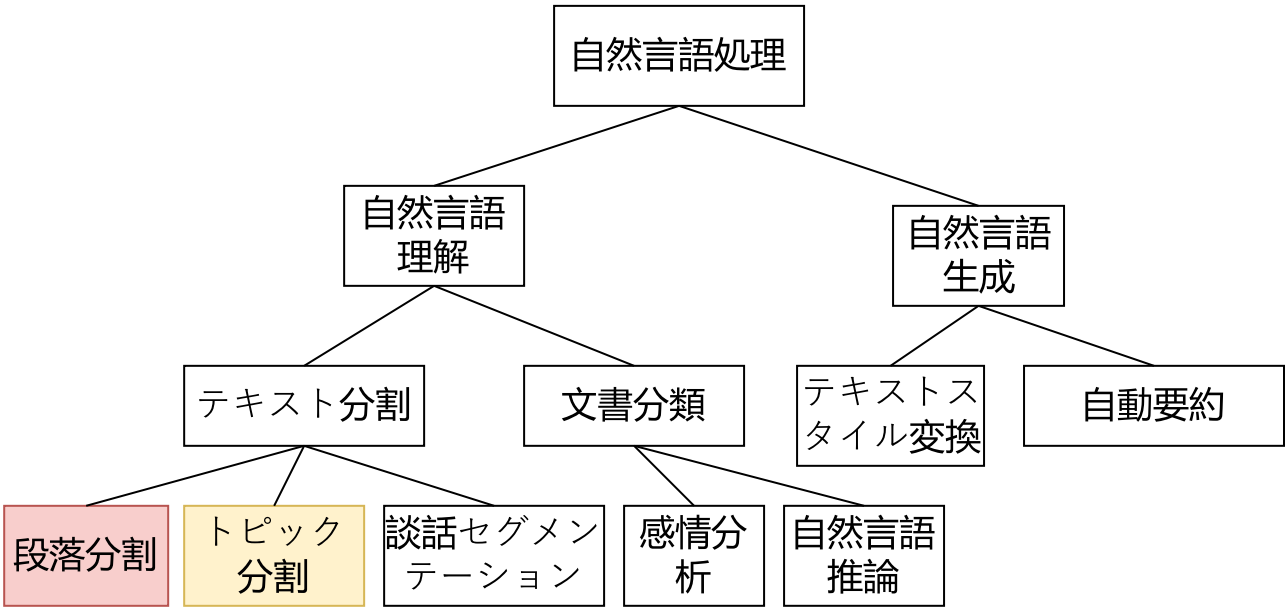

図 2.1: 自然言語処理タスク

は, 文間の類似度を計算するときに, cosine metric の代わりに dice metric を使用している.TopicTiling は, 文間の類似度を計算するときに, LDA モデルに基づくトピックベクトル表現を使用する.

外因性手法について, 高密度で低次元の静的な単語表現, LSA[14], word2vec[15], Glove[16] などの導入は, 初期の典型的な外因性手法である.Naili たちの研究 [7] は, 高密度で低次元の静的単語表現に基づく外因性トピックセグメンターの性能が, 内因性トピックセグメンターの性能よりもはるかに高いことを示した. ドメインに依存しない (domain-independent) 大規模なデータセットでトレーニングすることにより, word2vec などのモデルは単語の深い意味を学習すると考えられており, これが性能向上の理由と見なされている.

数年前にトピック分割の分野で優れた結果を出した研究のほとんど [17, 18, 19, 20] は, word2vec や GloVe などの外因性手法に基づいた.

近年, 静的な単語表現では扱いにくい多義性の問題に対処するために, 文脈によって動的な単語表現を学習するいくつかの手法が登場した. これらの手法 [3, 21, 22, 23] から取り出した動的な単語表現は, 多数の NLP タスクで静的な単語表現よりもはるかに優れていた [24]. トピック分割の分野でも, 動的な単語表現の導入は, 先行手法 [25, 26, 27] よりも優れた結果を生み出した.

2.2 段落分割

トピックセグメンテーションと同様に, 段落分割の手法を内因性と外因性の 2 つのカテゴリに分けられる. トピックのセグメンテーションと比較して, 段落分割はあまり研究されていない. 私たちが知っている研究のほとんどは内因性手法に属している [28, 29, 30, 31].

Bolshakov らの手法 [28] は texttiling に似ているが, 提案された text cohesion measure に基づいて段落分割を行う.Genzel ら [29] は, lexical と syntactic 素性を使用した sparse voted perceptron モデルであり, 初期のニューラルネットワークモデルと見なすことができる. Sporleder ら [30] と Filippova ら [31] の手法は, どちらも機械学習手法であるが, Sporleder らの手法は言語学素性に基づき, Filippova らの手法より良い性能を達成した.

Naili ら [7] の結論に基づいて, 本研究は外因性手法に焦点を当てる. 外因性手法を使用した段落分割研究の中では, 飯倉ら [1] の研究は, 動的な単語表現モデルを段落分割の分野に導入した最初のものである. 彼らの研究対象は小説データセットであり, データ不均衡の問題の影響を軽減するために, BERT に加えて Focal Loss[2] を導入し, 優れた結果を達成した.

これらの先行研究から結論をまとめると, 動的単語表現モデルが静的な単語表現モデルより性能が良く, 外因性手法が内因性手法よりも優れていると言える. 提案手法も飯倉らの手法も動的単語表現モデルに基づく外因性手法であるが, 提案手法は飯倉らの手法に基づく補助損失を導入し, 新聞データセットでの段落分割の性能をさらに向上させた.

第3章　手法

本研究では, 以下の 4 つの手法で比較実験を行った. 最初の 2 つはベースラインで, 後ろの 2 つは提案された手法である.

- BERT + BCE Loss (BERT の基本的な使い方, 略して Vanilla, ベースライン手法)
- BERT + Focal Loss (略して FL, 飯倉らの手法)
- BERT + BCE Loss + Auxiliary Loss（略して AUX, 本研究の提案手法）
- BERT + Focal Loss + Auxiliary Loss (略して FL+AUX, 本研究の提案手法)

図 3.1 は, 4 つの手法の違いを示している.

BERT の基本的な使い方と飯倉らの手法には共通問題点が存在する. それはニューラルネットワークモデルは分割点の周囲の文脈に注目する傾向があり, 距離の遠い情報を有効に利用できないことである. そこで本研究は補助損失を使って, モデルが距離の遠い情報に注目させることを試みた.

上記４つの手法は, BERT, 損失関数, および補助損失で構成され, 次に, これらの構成部分を順次説明する.

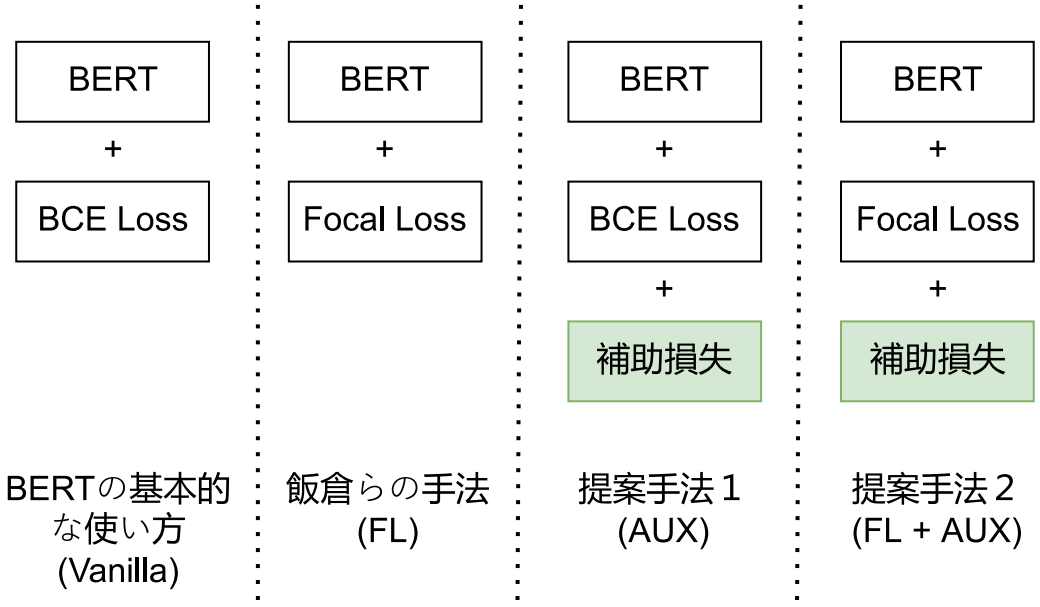

図 3.1: ４つの手法

3.1 BERT

近年, 大規模なデータセットで事前トレーニングされたモデルの使用が, 自然言語処理の主流になっている. BERT (Bidirectional Encoder Representations from Transformer の略) [3] とは, 最も研究されている大規模予備訓練自然言語処理モデルの一つである. BERT の有用性が, 近年多数の論文により確認されている. したがって, 本研究は BERT に基づく自動段落分割を検討する.

文章の形式段落を推定する問題は, 対象とする 2 文が同一の形式段落に所属するかどうかという二項分類問題として捉えることができる. BERT を使用してテキスト分類問題を解く手順は次のとおりである.

1. BERT を使用して入力を文ベクトル $vector$ に変換する.
2. 文ベクトル $vector$ を分類器に入力し, 確率値 p を取得する.
3. 確率値 p と正解ラベル $label$ を使用して, 損失値 $loss$ を計算する.
4. 損失値 $loss$ を減らすようにパラメータ調整を行う.

図 3.2 で対応する手順を示す. 図の中の [CLS] と [SEP] は, BERT の特殊トークンである. 特殊トークンとは, 入力テキストの特殊な位置をマークする記号である. 一般的

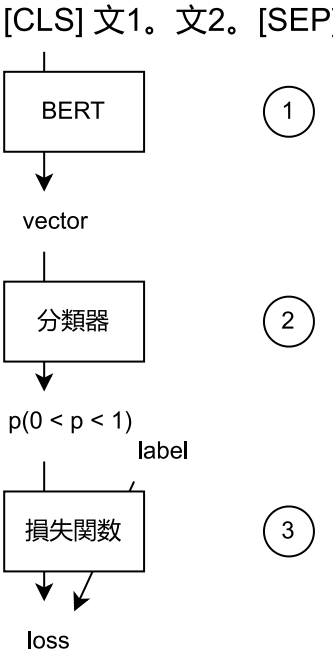

図 3.2: BERT でテキスト分類

には, [CLS] はテキストの始まりをマークし, [SEP] は入力の違う部分の接続点をマークする.

3.2 基本的な損失関数

本章で, 本研究で使われた 2 つの損失, BCE Loss と Focal Loss を紹介する. 基本的な損失関数と違って, 本研究が提案する補助損失 (Auxiliary Loss) とは, 損失を組み合わせる方法と考えられ, あらゆるの損失関数の上で使用できる. 補助損失について 3.3 節で紹介する.

3.2.1 BCE Loss

文章の形式段落を推定する問題は, 対象とする 2 文が同一の形式段落に所属するかどうかという二項分類問題として捉えることができ, そして二項分類問題で最も一般的に使用される損失関数は BCE Loss である.

BCE Loss は式 3.1 で表すことができる.

$$BCE(p, y) = -ylog(p) - (1 - y)log(1 - p), \tag{3.1}$$

ここで, y とは正解ラベルで, 0 は非分割, 1 は分割である. p は分類器が出した段落分割の確率である.p と y の差が大きいほど, 損失値が大きくなり, p と y が等しい場合, 損失値は 0 になる. したがって, この損失関数で得られる損失値の範囲は 0 から無限大である.

式 3.2 に表示されている p_t を導入することで, 式 3.1 を式 3.3 に単純化できる.

$$p_t = \begin{cases} p & if \quad y = 1 \\ 1 - p & if \quad y = 0 \end{cases} \tag{3.2}$$

$$BCE(p_t) = -log(p_t). \tag{3.3}$$

データ不均衡問題に対処するために, 実際に BCE Loss を使用する場合, 式 3.3 に重み $\alpha_t (0 < \alpha_t < 1)$ 導入することがよくある. ただし, 最適なパラメーターを探索する時間を短縮するために, 本研究では $\alpha_t = 1$ で BCE Loss を使用する.

3.2.2 Focal Loss

飯倉らの手法 (FL) は、BERT の基本的な使い方に基づき, BCE Loss を Focal Loss に置き換えた単純な改良と考えられる.

Focal Loss とは, コンピューター ビジョンの分野でよく使用される損失関数である. Focal Loss は, 自信過剰 (overconfident) な予測にペナルティを課すことで, データの不均衡問題を適切に対応できる. 対照的に, 自然言語処理の分野で Focal Loss を使用した研究はまれであり, 飯倉らの研究の新規性がそこから見える.

Focal Loss は式 3.4 で表される.

$$FL(p_t) = (-\alpha_t)(1 - p_t)^{\gamma} log(p_t). \tag{3.4}$$

γ は調整できるパラメータで, γ が 0 の場合, Focal Loss は BCE Loss と同じになる. p_t は式 3.2 と同じである. 3.2.1 節で説明したのと同じ理由で, α_t を 1 に設定する.

前述のように, Focal Loss は自信過剰 (overconfident) な予測にペナルティを課す. 具体的には, 予測結果が正解ラベルに近いほど, モデルが結果に自信があることを意味し, このとき, Focal Loss で計算された損失値は, BCE Loss よりも小さくなる. γ は Focal Loss の重要なハイパーパラメータである. γ の最適値を決定するために, グリッドサーチによって [0.5, 1.0, 2.0, 5.0] の 4 つの値を調べた.

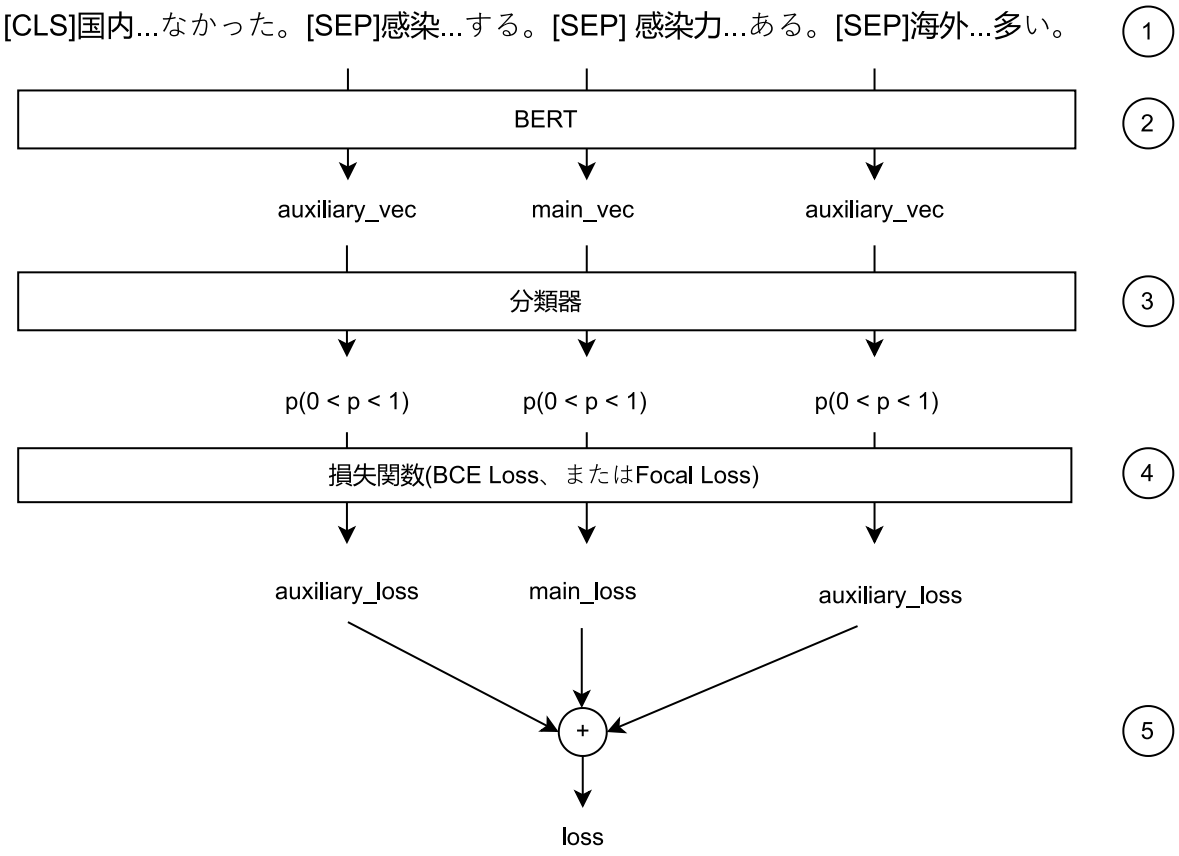

図 3.3: 補助損失で最終損失を計算する手順

3.3 Auxiliary Loss

本研究が提案する補助損失とは, 損失を組み合わせる方法と考えられ, あらゆるの基本損失関数の上で使用できる. 補助損失を使用して最終損失を計算する手順を図 3.3 に示す. 以下は, ステップごとの説明である.

1. ステップ 1: 入力内のすべての文の接続点を特殊トークン [SEP] でマークする. ここが従来手法と異なるところである. 特殊トークンについて 3.3.1 節で詳しく説明する.

2. ステップ 2: 入力を BERT で処理した後, [SEP] ごとに対応するベクトルを取り出す. 両側の [SEP] が補助タスクに導入されているため, 対応するベクトルと損失の前に auxiliary というプレフィックスが付けられる.

3. ステップ 3: 分類器を使用してベクトルを分類し, 分類器の出力 p を段落分割の確率値とする.

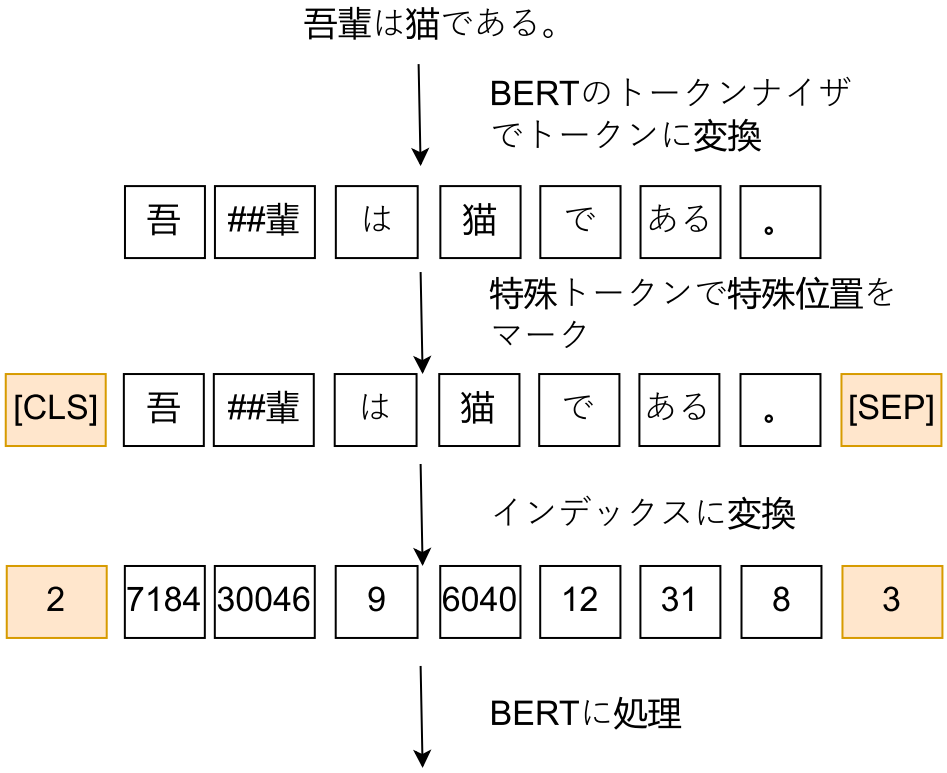

図 3.4: テキストをトークン ID に変換

4. ステップ 4: 分類器の出力 p と対応する位置の正解ラベルを使用して, 損失値を計算する.

5. ステップ 5: すべての損失値を組み合わせ, 最終損失を取得する. この損失値を使用して, モデルのパラメーター調整を行う. 損失値を組み合わせ方法について, 3.3.2 節で詳しく説明する.

補助損失を使用する利点をまとめる. まず, モデルがより広い範囲の文脈情報に注意を当てることを促進することができ, これは段落分割タスクを解くには有益である. そして, 連続した段落分割はほとんど発生しないため, 周囲に段落分割点があるかどうかを知ることが現在の判断に役立つ.

3.3.1 特殊トークンの使い方

BERT はテキストを直接処理できないため, BERT に入力する前には, BERT のトークナイザーを使ってテキストをトークン ID に変換する必要がある. テキストを BERT へ入力するための準備は図 3.4 のとおりである.

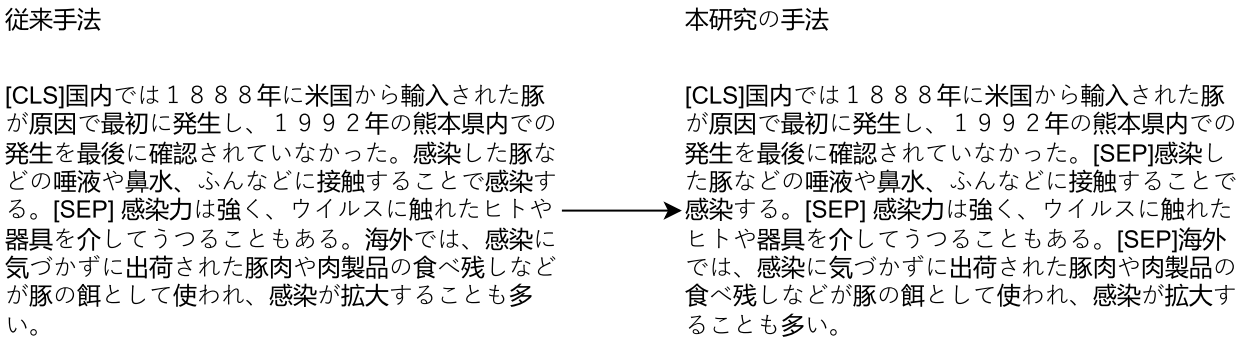

図 3.5: [SEP] の使い方の違い

「吾輩」が「吾」と「##輩」に分かれたことを図 3.4 から見える．これは, 東北大学のトークナイザーが WordPiece を使用して単語を処理しているためである．WordPiece は, 希な単語をより小さな断片に分割し, ## のプレフィックスは, この断片が前のステムに接続されていることを示す.

テキストから変換されたトークンに加えて, 特殊トークンもある．特殊トークンとは, 入力テキストの特殊な位置をマークする記号である．一般的には, [CLS] はテキストの始まりをマークし, [SEP] は入力の違う部分の接続点をマークする.

BERT の基本的な使い方と飯倉らの方法では, [SEP] は段落分割を判断する箇所をマークする．例えば, ウィンドウサイズが 2（入力が 2 文）の場合, [SEP] は最初の文の後に配置され, ウィンドウサイズが 4（入力が 4 文）の場合, [SEP] は 2 番目の文の後に配置される．それに対して, 本研究の手法では, [SEP] を使用してすべての文の接続点をマークする．ウィンドウサイズが 4 の場合, 従来手法と本手法の違いを図 3.5 に示す.

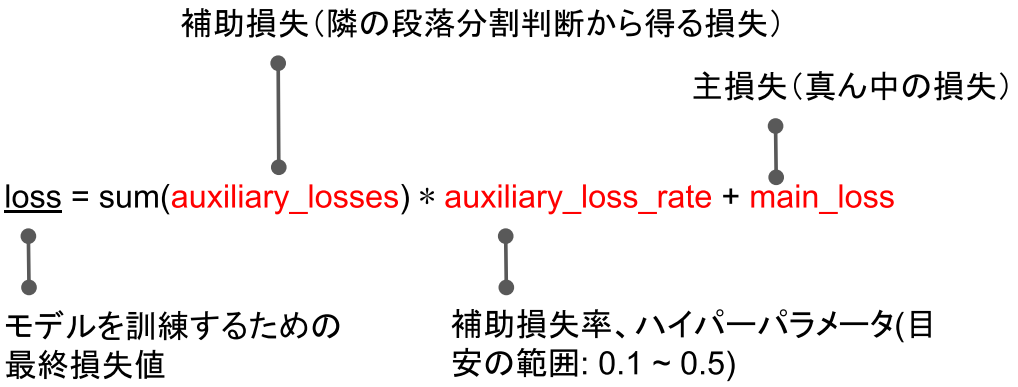

図 3.6: 損失値の計算式

3.3.2 損失値の組み合わせ方法

すべての文接続点に対応する損失値を取得した後, それらを 1 つの損失に結合する必要がある. 最も単純なやり方として, すべての損失値を合計することができる.

ただし, 損失の重要性によって異なる重みを付けることが一般的である. メインタスクは補助タスクよりも重要であるため, 対応する損失の重みはより高く設定したい. ウィンドウサイズ 4 の場合, メインタスクは真ん中の段落分割であり, 両側の段落分割は補助タスクになる. メインタスクからの損失値を main_loss と呼び, 補助タスクからの損失値を auxiliary_loss と呼ぶ. 最後に, 次の式で最終的な損失値を計算する.

$$\begin{aligned} loss \quad = \quad & sum(auxiliary\_losses) \\ & *auxiliary\_loss\_rate + main\_loss, \end{aligned} \tag{3.5}$$

理解しやすくするために, 図 3.6 で式を説明する. $auxiliary\_loss\_rate$ (補助損失の重み) はハイパーパラメーターで, データセットやウィンドウサイズなどの実験設定が変更する時, グリッドサーチで再選択する必要がある.

正式な実験を開始する前に, $auxiliary\_loss\_weight$ が 0.1 または 0.2 のときに, 本手法の性能が一番良いため, 正式な実験のとき, グリッドサーチで [0.0, 0.1, 0.2, 0.3] か

ら最適なパラメーターを選択した.

第4章　実験方法

この章では, Vanilla, FL, AUX, および FL + AUX の性能を比較する. 次に, 実験で使用された性能指標, データセット, およびパラメータ設定について説明する.

4.1　性能指標

$F1$ スコアは, テキストセグメンテーションの分野で一般的に使用される性能指標であり, 本実験では Google と飯倉らの研究と同様に, 性能指標として $F1$ スコアを使用する.

特殊なケースの処理については, 記事の最初の文が段落の先頭にあるかどうかを判断するのは簡単すぎるため, トレーニングと評価の際に記事の先頭にある特殊なケースをスキップする.

4.2　データセット

提案手法の一般的な有効性を確認するために, 2 つのデータセットで実験を行った. １番目のデータセットは夏目漱石の小説であり, 飯倉らが研究で使用したデータセットと一致する.2 番目のデータセットは 2019 年の毎日新聞である. 表 1 と表 2 は, 2 つのデータセットの詳細な情報を示している. 非分割点/分割点の数の比率からわかるように, 小説データセットのデータの不均衡問題はより深刻である. 統計的な検定を行うために, 毎日新聞データセットを人為的に 10 のグループに分割し, 各データセットで手法の性能を観察できるようになる. 小説のテストデータセット自体には 188 章があるため, 章に基づいて手法の $F1$ スコアを求める.

毎日新聞データセットの処理について. 元の毎日新聞データセットでは, 記事は時系列順に並べられていて, 同じカテゴリの記事もグループ化されなかったため, ランダム化処理しなくてもデータの均等性を保つことができる. それゆえ, 本実験は記事の順序

表 4.1: データセット情報 (毎日新聞)

データセット	記事の数	文の数	分割点の数	非分割点の数	非分割点/分割点の数
Train	2000	27436	8604	18832	2.19
Dev	1000	13508	2073	4566	2.20
Test1	500	6644	2106	4538	2.15
Test2	500	6349	2081	4268	2.05
Test3	500	7193	2180	5013	2.30
Test4	500	6364	2033	4331	2.13
Test5	500	6430	2031	4399	2.17
Test6	500	7130	2222	4908	2.21
Test7	500	7026	2142	4884	2.28
Test8	500	6942	2203	4739	2.15
Test9	500	7806	2327	5479	2.35
Test10	500	7410	2312	5098	2.21
テストセットの平均	500	6929	2164	4766	2.20

表 4.2: データセット情報 (小説)

データセット	章の数	分割点の数	非分割点の数	非分割点/分割点の数
Train	340	3625	24553	6.77
Dev	110	630	4327	6.87
Test	188	1874	9225	4.923

を変更しなかった. 「特別版」などの特別なカテゴリの記事を削除した後, トレーニング, テスト, 開発セットを順番に, 記事をスキップせずに作成した.

参考のために, 毎日新聞のテストセットから 1 つの記事 (テストセットごとに 500 記事がある) と, 小説のテストセットから 1 つの章 (テストセットには合計 188 章がある) を例に挙げる.

次は毎日新聞データセットの例である.

木造住宅の建築販売会社の営業マンが次のような話をしてくれた。「預金が１００万円、年収４００万円程度の若いサラリーマンが、３０００万円前後の家の購入を申し込んでくる。その人たちに聞くと、残りの資金は全て銀行が超低金利で融資してくれ、住宅取得促進税制を活用すると１０年間所得税がほとんどかからなくなるから、借家で生活するより得になると銀行に勧められたという。

今や給与が毎年上昇していく時代ではないし、金利も３０年間固定とは限らず、いずれ元利返済が増えていく。子供が生まれ、教育費も必要になった時、その人たちは生活していけるのかと心配になる。

昔は銀行が厳格な返済計画を作らせて、返済可能な人にだけ融資が行われたものだ。ところが日銀の異次元金融緩和で貸し付けを急ぐ銀行側が、個人向け住宅融資を盛んに行うようになった。その上、住宅不足時代に作られた住宅取得促進税制は今回の税制改正で、減税の適用期間が１０年から１３年に延びる。家余りの時代に、無理な住宅取得にますます拍車がかかる可能性が強い。

人口減少が進む日本では、いずれ3軒に1軒が空き家になると言われている。例えば、千葉県の松戸駅周辺や、東京都目黒区の高級住宅街などでも空き家ばかりが目に付く。本来は既存住宅のリフォームが求められている時代だ。しかし、相変わらず住宅取得に一層の減税を行い、将来売れるとは思えない遠郊の地で住宅建設が行われている。

今後も人口減が加速していく中で、来年の東京五輪後のそう遠くない時期に、金融危機の引き金となったサブプライムローンの日本版とも言える問題が浮上してくる恐れがある。

次は小説データセットの例である.

その日の帰りがけに津田は途中で電車を下りて、停留所から賑にぎやかな通りを少し行った所で横へ曲った。質屋の暖簾のれんだの碁会所ごかいしょの看板だの鳶とびの頭かしらのいそうな格子戸作こうしどづくりだのを左右に見ながら、彼は彎曲わんきょくした小路こうじの中ほどにある擦硝子張すりガラスばりの扉を外から押して内へ入った。扉の上部に取り付けられた電鈴ベルが鋭どい音を立てた時、彼は玄関の突き当りの狭い部屋から出る四五人の眼の光を一度に浴びた。窓のないその室へやは狭いばかりでなく実際暗かった。外部そとから急に入って来た彼にはまるで穴蔵のような感じを与えた。彼は寒そうに長椅子の片隅かたすみへ腰をおろして、たった今暗い中から眼を光らして自分の方を見た人達を見返した。彼らの多くは室の真中に出してある大きな瀬戸物火鉢ひばちの周囲まわりを取り巻くようにして坐っていた。そのうちの二人は腕組のまま、二人は火鉢の縁ふちに片手を翳かざしたまま、ずっと離れた一人はそこに取り散らした新聞紙の上へ甜なめるように顔を押し付けたまま、また最後の一人は彼の今腰をおろした長椅子の反対の隅に、心持身体からだを横にして洋袴ズボンの膝頭ひざがしらを重ねたまま。

電鈴ベルの鳴った時申し合せたように戸口をふり向いた彼らは、一瞥いちべつの後のちまた申し合せたように静かになってしまった。みんな黙って何事をか考え込んでいるらしい態度で坐っていた。その様子が津田の存在に注意を払わないというよりも、かえって津田から注意されるのを回避するのだとも取れた。単に津田ばかりでなく、お互に注意され合う苦痛を憚はばかって、わざとそっぽへ眼を落しているらしくも見えた。

この陰気な一群いちぐんの人々は、ほとんど例外なしに似たり寄ったりの過去をもっているものばかりであった。彼らはこうして暗い控室の中で、静かに自分の順番の来るのを待っている間に、むしろ華はなやかに彩いろどられたその過去の断片のために、急に黒い影を投げかけられるのである。そうして明るい所へ眼を向ける勇気がないので、じっとその黒い影の中に立ち竦すくむようにして閉とじ籠こもっているのである。

津田は長椅子の肱掛ひじかけに腕を載のせて手を額にあてた。彼は黙祷もくとうを神に捧げるようなこの姿勢のもとに、彼が去年の暮以来この医

者の家で思いがけなく会った二人の男の事を考えた。

その一人は事実彼の妹婿いもとむこにほかならなかった。この暗い室の中で突然彼の姿を認めた時、津田は吃驚びっくりした。そんな事に対して比較的無頓着むとんじゃくな相手も、津田の驚ろき方が反響したために、ちょっと挨拶あいさつに窮したらしかった。

他の一人は友達であった。これは津田が自分と同性質の病気に罹かかっているものと思い込んで、向うから平気に声をかけた。彼らはその時二人いっしょに医者の門を出て、晩飯を食いながら、性セックスと愛ラヴという問題についてむずかしい議論をした。

妹婿の事は一時の驚ろきだけで、大した影響もなく済んだが、それぎりで後あとのなさそうに思えた友達と彼との間には、その後ご異常な結果が生れた。

その時の友達の言葉と今の友達の境遇とを連結して考えなければならなかった津田は、突然衝撃ショックを受けた人のように、眼を開いて額から手を放した。

すると診察所から紺こんセルの洋服を着た三十恰好がっこうの男が出て来て、すぐ薬局の窓の所へ行った。彼が隠袋かくしから紙入を出して金を払おうとする途端とたんに、看護婦が敷居の上に立った。彼女と見知り越ごしの津田は、次の患者の名を呼んで再び診察所の方へ引き返そうとする彼女を呼び留めた。「順番を待っているのが面倒だからちょっと先生に訊きいて下さい。明日あしたか明後日あさって手術を受けに来て好いかって」

奥へ入った看護婦はすぐまた白い姿を暗い室へやの戸口に現わした。「今ちょうど二階が空あいておりますから、いつでも御都合の宜よろしい時にどうぞ」

津田は逃のがれるように暗い室を出た。彼が急いで靴を穿はいて、擦硝子張すりガラスばりの大きな扉を内側へ引いた時、今まで真暗に見えた控室にぱっと電灯が点ついた。

表 4.3: 各手法の最適パラメータ

パラメータ	Vanilla	FL	AUX	FL+AUX
新聞データセット				
γ (Focal Loss)	-	2.0	-	5.0
補助損失率	-	-	0.2	0.1
エポック数	2	3	2	2
小説データセット				
γ (Focal Loss)	-	5.0	-	5.0
補助損失率	-	-	0.3	0.1
エポック数	2	2	2	2

4.3 パラメータ設定

本研究は PyTorch と Hugging Face というディープラーニングフレームワークを使っている. 学習済みの BERT は東北大の bert-base-japanese-whole-word-masking である. BERT のパラメータは, 12 層, 隠れ層 768, 12 ヘッドである. オプティマイザーは AdamW で, 学習率は 2e-5 である.

補助損失を有効にするには, 入力が 2 文以上である必要がある. モデルの期待する文の数を ウィンドウサイズ で表す. 本研究では, ウィンドウサイズを 4 に固定し, より大きい ウィンドウサイズ の実験は今後の課題にする.

他のパラメータについて, dev データセット上のグリッドサーチで, 各手法の最適なパラメータを調べた. グリッドサーチの時, パラメータ設定ごとに 10 個モデルを訓練し, dev データセットにおける出力の平均 F 値で性能を比較する. グリッドサーチの結果による各手法の最尤パラメータは表 4.3 のとおりである. パラメータ設定が変更されると, モデルを訓練する最適なエポックが変わる可能性があることを考慮し, 1 から 3 の間の最適なエポックを調査した. 実験時間を節約するために, エポックが独立しているパラメータではないことに注意してください. つまり, 1, 2, 3 のエポックで同じ 10 個のモデルが使用される.

第5章　実験結果

5.1　性能

毎日新聞データセットの実験結果は, 表 5.1 に示されている. Vanilla とは BERT の基本的な使い方, AUX とは BERT の基本的な使い方の上で補助損失を導入する手法, FL とは飯倉らの手法, FL+AUX とは飯倉らの手法の上で補助損失を導入する手法を意味する. 最後の列 all one とは, すべての出力が 1 のときの $F1$ スコアである. 性能の良さが偶然に発生しないことを確認するためにテストセットを 10 グループに分けたので, 表にグループごとの結果を記載する. Mean($F1$) は 10 グループの平均 F1 で, Std.($F1$) は 10 モデルの F 値の標準偏差である. Mean($prec$) と Mean(rec) はそれぞれ平均精度と平均適合率である.

小説データセットの場合, テストデータセット (夏目漱石の小説「明暗」) には 188 章があるので, 章ごとに $F1$ スコアを計算したが, 表 5.2 に平均値だけを載せた. 統計的検定で結果間に有意差があることが確認された.

2 つのデータセットの実験結果から結論をまとめる. まず, FL+AUX は両方のデータセットで一番良い性能を達成し, 提案された手法の有効性を示した. 次に, FL は, データ不均衡の問題が緩和である毎日新聞データセットでの性能が良くないが, 補助損失はいつも BERT の基本的な使い方 (Vanilla) より優れている. そこで補助損失の一般的な有効性が見える.

5.2　有意性検定

結果が統計的に有意であることを確認するために, 結果に対して有意差検定を行った.

ペアのテスト スコアを取得するために, Daily News データセットで 10 個のテストセットを用意した. Novel データセットの場合, テストセットには 188 の章が含まれているため, 章に基づいてスコアを計算した.

有意差検定の方法について. 結果が正規分布している場合は両側 t 検定を使用し, そ

表 5.1: 毎日新聞データセットでの性能

	FL+AUX	FL	Vanilla	AUX	All one
Test1	**0.6784**	0.6715	0.6708	0.6759	0.4814
Test2	**0.7004**	0.6937	0.6923	0.6964	0.4937
Test3	**0.6773**	0.6629	0.6671	0.6713	0.4652
Test4	**0.6846**	0.6745	0.6783	0.6772	0.4842
Test5	**0.6860**	0.6743	0.6787	0.6805	0.4801
Test6	0.6658	0.6565	0.6620	**0.6696**	0.4752
Test7	**0.6841**	0.6750	0.6793	0.6811	0.4673
Test8	**0.6831**	0.6737	0.6760	0.6752	0.4818
Test9	**0.6698**	0.6596	0.6594	0.6644	0.4593
Test10	**0.6718**	0.6626	0.6640	0.6699	0.4756
Mean($F1$)	**0.6801**	0.6704	0.6728	0.6761	0.4764
Std.($F1$)	0.0055	0.0116	0.0144	0.0056	0.0
Mean($prec$)	0.7059	0.6994	**0.7103**	0.7082	0.3127
Mean(rec)	0.6413	0.6464	0.6431	**0.6479**	**1.0**

表 5.2: 小説データセットでの性能

	FL+AUX	FL	Vanilla	AUX	All one
Mean($prec$)	0.8802	0.8587	0.8985	**0.9291**	0.1688
Mean(rec)	0.8113	**0.8166**	0.7711	0.7697	1.0
Mean($F1$)	**0.8339**	0.8261	0.8159	0.8244	0.2889
Std.($F1$)	0.0043	0.0072	0.0301	0.0130	0.0

うでない場合は両側 Wilcoxon の符号順位検定を使用する. 表 5.3 は, 正規分布検定の結果を示している. 0.05 未満の p 値は, 結果が正規分布ではないことを意味する. 正規分布に従う結果を太字でマークした.

表 5.3 より， 小説データセットでの結果は正規分布に従わないことがわかる. その理由は, 小説データセットが章ごとに F 値を計算し, 各章が比較的短く (平均 60 文しか含まれていない), 計算された F 値に大きな偏差が生じるためである. 一方, 新聞テストセットにはそれぞれ 6000 文以上が含まれているため, 計算された F 値は比較的安定している.

正規分布検定の結果に基づいて, 新聞データセットに対して、両側 t 検定と両側 Wilcoxon の符号順位検定両方の結果を示し, 小説データセットの結果には両側 Wilcoxon の符号順位検定のみを使用する.

表 5.4 は, 有意差検定の p 値を示す. p 値が 0.05 以下 (太字で示されている) の場合, 2 つの手法を使用して得られた結果の差は統計的に有意であると見なされる.

表 5.3: 正規分布検定の p 値

	Vanilla	FL	AUX	FL+AUX
新聞データセット	**0.7492**	**0.2107**	0.0437	**0.4947**
小説データセット	4.85e-12	3.34e-10	9.33e-13	1.38e-12

表 5.4: 有意差検定からの p 値

	Vanilla	FL	AUX	FL+AUX
毎日新聞データセット (両側 t 検定)				
Vanilla	-	-	-	-
FL	**0.0135**	-	-	-
AUX	**0.0047**	**0.0005**	-	-
FL+AUX	**0.000001**	**0.0000002**	**0.0047**	-
毎日新聞データセット (両側 Wilcoxon の符号順位検定)				
Vanilla	-	-	-	-
FL	**0.0273**	-	-	-
AUX	**0.0097**	**0.0019**	-	-
FL+AUX	**0.0019**	**0.0019**	**0.0136**	-
小説データセット (両側 Wilcoxon の符号順位検定)				
Vanilla	-	-	-	-
FL	**0.0003**	-	-	-
AUX	**1.32e-5**	0.055	-	-
FL+AUX	**1.79e-11**	**6.33e-7**	**0.0418**	-

第6章　考察

6.1　ケーススタディー

実験では提案手法が従来手法よりも優れていることが示されているが, この優位性は統計レベルであり, 個々の例に反映することは困難である. 一方, ニューラルネットワークモデルのパラメータがたくさんあるため, モデルの判断の根拠を理解するのが難しい.

近年, 人々は深層学習の判断根拠を理解するために多くの努力をしてきた. 例えば, 注意力 (Attention) を可視化することはできる. しかし, この方法の解釈可能性も疑問視されている [32].

参考のために, この章では以下のカテゴリからランダムに各１例とその注意力を示す. 注意力の可視化について, トークンが赤くなるほど, 集めた注意力が高いことを意味する.

- 例の 1: 新聞データセット, 正解が分割, FL + AUX が正しく, FL が間違う.
- 例の 2: 新聞データセット, 正解が分割, FL + AUX が間違っており, FL が正しい.
- 例の 3: 小説データセット, 正解が分割, FL + AUX が正しく, FL が間違う.
- 例の 4: 小説データセット, 正解が分割, FL + AUX が間違っており, FL が正しい.

例の１ (新聞データセット, 正解ラベルは 1, FL + AUX の出力は 0.69, FL の出力は 0.23, モデルは訓練済みなモデルからランダムに選出された) は次のとおりであり, ２つのモデルの注意力を可視化した結果は図 6.1 である.

> 自民党の三原じゅん子女性局長は１５日、議員活動と 育児の両立を目指す「超党派ママパパ議員連盟」会長・野田聖子衆院予算委員長の会合で、ネット投票に関する論点整理を提示 した。衆院法制局が複数の憲法学者に

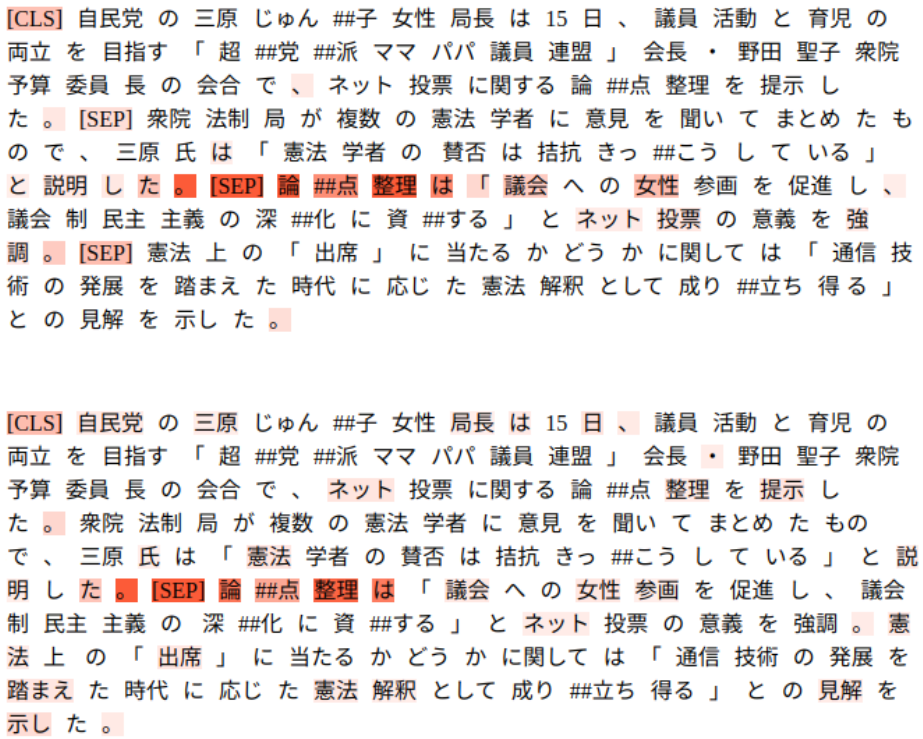

図 6.1: 例の１に対する注意力の可視化 (上は AUX+FL, 下は FL)

意見を聞いてまとめたもので、三原氏は「憲法学者の賛否は拮抗きっこうしている」 と説明した。

論点整理は「議会への女性参画を促進し、議会制民主主義の深化に資する」とネット投票の意義を強調。憲法上の「出席」に当たるかどうかに関しては「通信技術の発展を踏まえた時代に応じた憲法解釈として成り立ち得る」との見 解を示した。

例の２ (新聞データセット, 正解ラベルは 1, FL + AUX の出力は 0.46, FL の出力は 0.86) は次であり, ２つのモデルの注意力を可視化した結果は図 6.2 である.

政情不安が続く南米ベネズエラに日本人旅行客の女性が隣国ブラジルから入国後、国境が封鎖されたため、ブラジル側に戻れず足止めされていることが２６日、分かった。在ベネ ズエラ日本大使館が現地に職員を派遣し、首都カラカスから出国するなどの方策を検討している。

関係者によると、女性は ４０歳で、ブラジル国境に近いベネズエラ南部サンタエレナデウアイレンで足止めされている。今月１７日にベネズエラ入 りし、ギアナ高地に位置するロライマ山に登った。

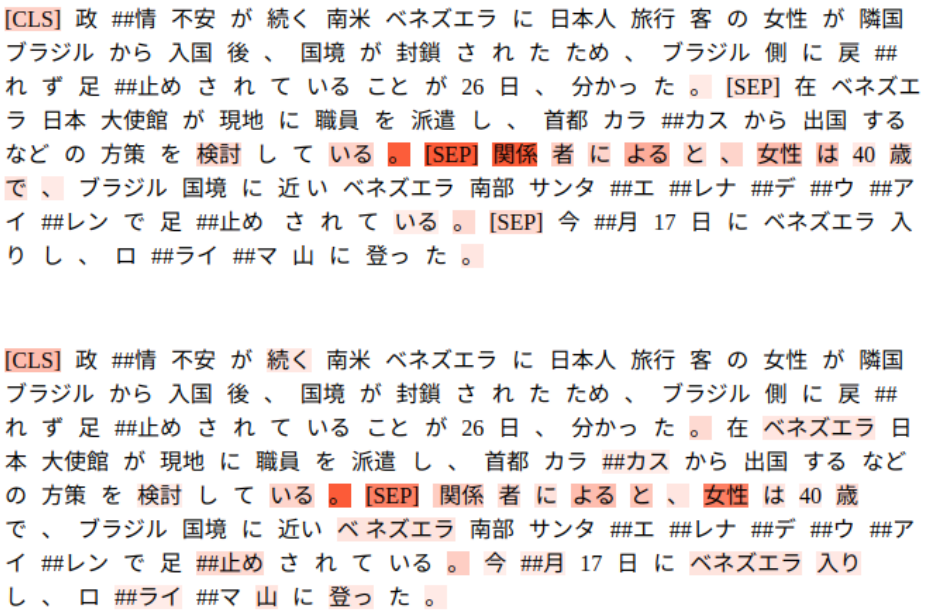

図 6.2: 例の 2 に対する注意力の可視化 (上は AUX+FL, 下は FL)

例の 3(小説データセット, 正解ラベルは 1, FL + AUX の出力は 0.84, FL の出力は 0.32) は次のとおりであり, 2 つのモデルの注意力を可視化した結果は図 6.3 である.

「御父さんからまだ手紙は来なかったかね」

「いいえ来ればいつもの通り御机の上に載せておきますわ」

津田はその予期した手紙が机の上に載っていなかったから、わざわざ下りて来たのであった。

「郵便函ゆうびんばこの中を探させましょうか」

例の 4(小説データセット, 正解ラベルは 1, FL + AUX の出力は 0.36, FL の出力は 0.55) は次であり, 2 つのモデルの注意力を可視化した結果は図 6.4 である。

「大丈夫」

俥は再び走かけ出した。

彼らの医者に着いたのは予定の時刻より少し後おくれていた。しかし午ひるまでの診察時間に間に合わないほどでもなかった。

これらの例は参考のためにリストした. 何かの知見を見つけることは難しい. 注意力の可視化に関しては, 2 つの手法の注意力が分割点の近くに集中していることがわかる.

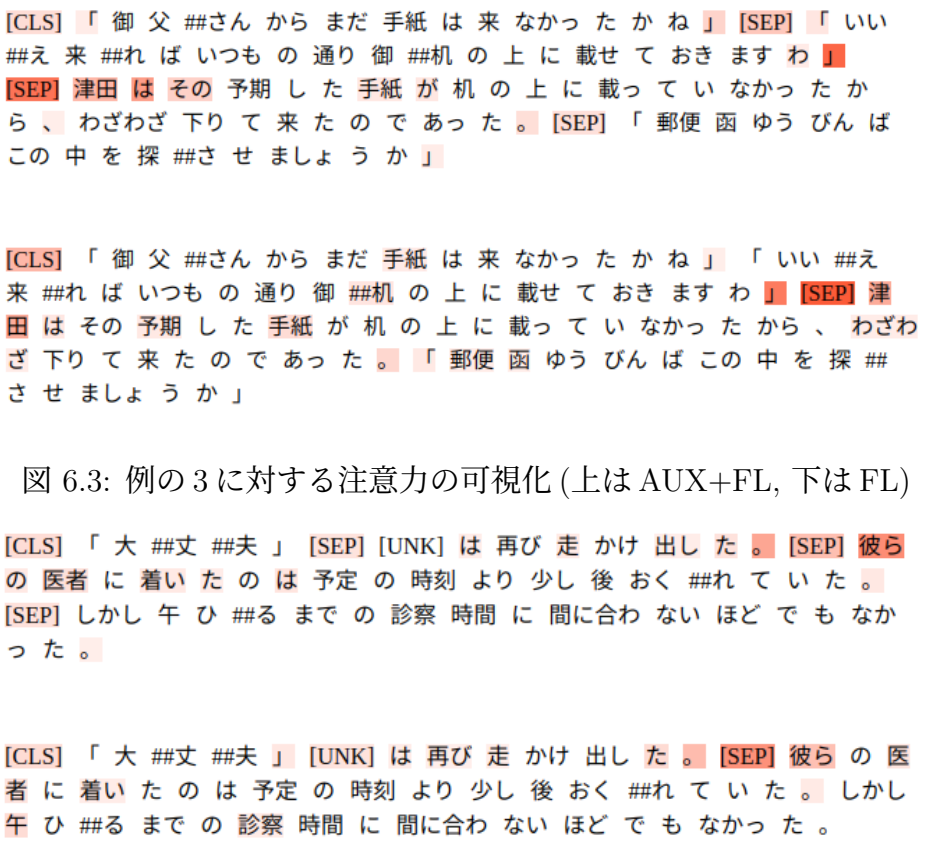

図 6.3: 例の 3 に対する注意力の可視化 (上は AUX+FL, 下は FL)

図 6.4: 例の 4 に対する注意力の可視化 (上は AUX+FL, 下は FL)

6.2 一記事に対する出力例

次に, 新聞の一記事と小説の一章に対する提案手法と従来手法の出力を示す. 【1】とはモデルは「ここが分割点」と判断し, 【0】とは非分割点と判断する意味である. 改行は, 生データに存在する段落分割点を意味する.

新聞データセットからの一記事、FL + AUX(提案手法) の出力:

木造住宅の建築販売会社の営業マンが次のような話をしてくれた。【1】「預金が１００万円、年収４００万円程度の若いサラリーマンが、３０００万円前後の家の購入を申し込んでくる。【0 】その人たちに聞くと、残りの資金は全て銀行が超低金利で融資してくれ、住宅取得促進税制を活用すると１０年間所得税がほ とんどかからなくなるから、借家で生活するより得になると銀行に勧められたという。

【1】今や給与が毎年上昇していく時代ではないし、金利も３０年間固定とは限らず、いずれ元利返済が増えていく。【1】子供が生まれ、教育費も必要に なった時、その人たちは生活していけるのかと心配になる。

【0】昔は銀行が厳格な返済計画を作らせて、返済可能な人にだけ融資が行われたものだ。【0】ところが日銀の異次元金融緩和で貸し付けを急ぐ銀行側が、個人向け住宅融資を盛ん に行うようになった。【1】その上、住宅不足時代に作られた住宅取得促進税制は今回の税制改正で、減税の適用期間が１０年 から１３年に延びる。【0】家余りの時代に、無理な住宅取得にますます拍車がかかる可能性が強い。

【1】人口減少が進む日本では、いずれ３軒に１軒が空き家になると言われている。【0】例えば、千葉県の松戸駅周辺や、東京都目黒区の 高級住宅街などでも空き家ばかりが目に付く。【0】本来は既存住宅のリフォームが求められている時代だ。【0】しかし、相 変わらず住宅取得に一層の減税を行い、将来売れるとは思えない遠郊の地で住宅建設が行われている。

【1】今後も人口減が加速していく中で、来年の東京五輪後のそう遠くない時期に、金融危機の引き金となったサブプライムローンの日本版 とも言える問題が浮上してくる恐れがある。

新聞データセットからの一記事、FL(従来手法) の出力:

木造住宅の建築販売会社の営業マンが次のような話をしてくれた。【0】「預金が１００万円、年収４００万円程度の若いサラリーマンが、３０００万円前後の家の購入を申し込んでくる。【0 】その人たちに聞くと、残りの資金は全て銀行が超低金利で融資してくれ、住宅取得促進税制を活用す

ると１０年間所得税がほ とんどかからなくなるから、借家で生活するより得になると銀行に勧められたという。

【0】今や給与が毎年上昇していく時代ではないし、金利も３０年間固定とは限らず、いずれ元利返済が増えていく。 【0】子供が生まれ、教育費も必要に なった時、その人たちは生活していけるのかと心配になる。

【0】昔は銀行が厳格な返済計画を作らせて、返済可能な人にだけ融資が行われたものだ。 【0】ところが日銀の異次元金融緩和で貸し付けを急ぐ銀行側が、個人向け住宅融資を盛ん に行うようになった。 【1】その上、住宅不足時代に作られた住宅取得促進税制は今回の税制改正で、減税の適用期間が１０年 から１３年に延びる。 【0】家余りの時代に、無理な住宅取得にますます拍車がかかる可能性が強い。

【1】人口減少が進む日本では、いずれ３軒に１軒が空き家になると言われている。 【0】例えば、千葉県の松戸駅周辺や、東京都目黒区の 高級住宅街などでも空き家ばかりが目に付く。 【0】本来は既存住宅のリフォームが求められている時代だ。 【0】しかし、相 変わらず住宅取得に一層の減税を行い、将来売れるとは思えない遠郊の地で住宅建設が行われている。

【0】今後も人口減が加速していく中で、来年の東京五輪後のそう遠くない時期に、金融危機の引き金となったサブプライムローンの日本版 とも言える問題が浮上してくる恐れがある。

小説データセットからの一章、FL + AUX(提案手法) の出力:

その日の帰りがけに津田は途中で電車を下りて、停留所から賑にぎやかな通りを少し行った所で横へ曲った。【0】質屋の暖簾のれんだの碁会所ごかいしょの看板だの鳶とびの頭かしらのいそうな格子戸作こうしどづくりだのを左右に見ながら、彼は彎曲わんきょくした小路こうじの中ほどにある擦硝子張すりガラスばりの扉を外から押して内へ入った。【1】扉の上部に取り付けられた電鈴ベルが鋭どい音を立てた時、彼は玄関の突き当りの狭い部屋から出る四五人の眼の光を一度に浴びた。【0】窓のないその室へやは狭いばかりでなく実際暗かった。【0】外部そとから急に入って来た彼にはまるで穴蔵のような感じを与えた。【1】彼は寒そうに長椅子の片隅かたすみへ腰をおろして、たった今暗い中から眼を光らして自分の方を見た人達を見返した。【1】彼らの多くは室の真中に出してある大きな瀬戸物火鉢ひばちの周囲まわりを取り巻くようにして坐っていた。【0】そのうちの二人は腕組のまま、二人は火鉢の縁ふちに片手を翳かざしたまま、ずっと離れた一人はそこに取り散らした新聞紙の上へ甜なめるように顔を押し付けたまま、また最後の一人は彼の今腰をおろした長椅子の反対の隅に、心持身体からだを横にして洋袴ズボンの膝頭ひざがしらを重ねたまま。

【1】電鈴ベルの鳴った時申し合せたように戸口をふり向いた彼らは、一瞥いちべつの後のちまた申し合せたように静かになってしまった。【0】みんな黙って何事をか考え込んでいるらしい態度で坐っていた。【0】その様子が津田の存在に注意を払わないというよりも、かえって津田から注意されるのを回避するのだとも取れた。【0】単に津田ばかりでなく、お互に注意され合う苦痛を憚はばかって、わざとそっぽへ眼を落しているらしくも見えた。

【1】この陰気な一群いちぐんの人々は、ほとんど例外なしに似たり寄ったりの過去をもっているものばかりであった。【0】彼らはこうして暗い控室の中で、静かに自分の順番の来るのを待っている間に、むしろ華はなやかに彩いろどられたその過去の断片のために、急に黒い影を投げかけられるのである。【0】そうして明るい所へ眼を向ける勇気がないので、じっとその黒い影の中に立ち竦すくむようにして閉とじ籠こもっているのである。

【1】津田は長椅子の肱掛ひじかけに腕を載のせて手を額にあてた。【0】彼は黙祷もくとうを神に捧げるようなこの姿勢のもとに、彼が去年の暮以来この医者の家で思いがけなく会った二人の男の事を考えた。

【0】その一人は事実彼の妹婿いもとむこにほかならなかった。【1】この暗い室の中で突然彼の姿を認めた時、津田は吃驚びっくりした。【0】そんな事に対して比較的無頓着むとんじゃくな相手も、津田の驚ろき方が反響したために、ちょっと挨拶あいさつに窮したらしかった。

【0】他の一人は友達であった。【0】これは津田が自分と同性質の病気に罹かかっているものと思い込んで、向うから平気に声をかけた。【0】彼らはその時二人いっしょに医者の門を出て、晩飯を食いながら、性セックスと愛ラヴという問題についてむずかしい議論をした。

【0】妹婿の事は一時の驚ろきだけで、大した影響もなく済んだが、それぎりで後あとのなさそうに思えた友達と彼との間には、その後ご異常な結果が生れた。

【0】その時の友達の言葉と今の友達の境遇とを連結して考えなければならなかった津田は、突然衝撃ショックを受けた人のように、眼を開いて額から手を放した。

【1】すると診察所から紺こんセルの洋服を着た三十恰好がっこうの男が出て来て、すぐ薬局の窓の所へ行った。【0】彼が隠袋かくしから紙入を出して金を払おうとする途端とたんに、看護婦が敷居の上に立った。【0】彼女と見知り越ごしの津田は、次の患者の名を呼んで再び診察所の方へ引き返そうとする彼女を呼び留めた。【0】「順番を待っているのが面倒だからちょっと先生に訊きいて下さい。【0】明日あしたか明後日あさって手術を受けに来て好いかって」

【1】奥へ入った看護婦はすぐまた白い姿を暗い室へやの戸口に現わした。【0】「今ちょうど二階が空あいておりますから、いつでも御都合の宜よろしい時にどうぞ」

【1】津田は逃のがれるように暗い室を出た。【0】彼が急いで靴を穿はいて、擦硝子張すりガラスばりの大きな扉を内側へ引いた時、今まで真暗に見えた控室にぱっと電灯が点ついた。

小説データセットからの一章、FL(従来手法) の出力:

その日の帰りがけに津田は途中で電車を下りて、停留所から賑にぎやかな通りを少し行った所で横へ曲った。【0】質屋の暖簾のれんだの碁会所ごかいしょの看板だの鳶とびの頭かしらのいそうな格子戸作こうしどづくりだのを左右に見ながら、彼は彎曲わんきょくした小路こうじの中ほどにある擦硝子張すりガラスばりの扉を外から押して内へ入った。 【0】扉の上部に取り付けられた電鈴ベルが鋭どい音を立てた時、彼は玄関の突き当りの狭い部屋から出る四五人の眼の光を一度に浴びた。 【0】窓のないその室へやは狭いばかりでなく実際暗かった。 【0】外部そとから急に入って来た彼にはまるで穴蔵のような感じを与えた。 【0】彼は寒そうに長椅子の片隅かたすみへ腰をおろして、たった今暗い中から眼を光らして自分の方を見た人達を見返した。 【0】彼らの多くは室の真中に出してある大きな瀬戸物火鉢ひばちの周囲まわりを取り巻くようにして坐っていた。【0】そのうちの二人は腕組のまま、二人は火鉢の縁ふちに片手を翳かざしたまま、ずっと離れた一人はそこに取り散らした新聞紙の上へ甜なめるように顔を押し付けたまま、また最後の一人は彼の今腰をおろした長椅子の反対の隅に、心持身体からだを横にして洋袴ズボンの膝頭ひざがしらを重ねたまま。

【0】電鈴ベルの鳴った時申し合せたように戸口をふり向いた彼らは、一瞥いちべつの後のちまた申し合せたように静かになってしまった。【0】みんな黙って何事をか考え込んでいるらしい態度で坐っていた。 【0】その様子が津田の存在に注意を払わないというよりも、かえって津田から注意されるのを回避するのだとも取れた。 【0】単に津田ばかりでなく、お互に注意され合う苦痛を憚はばかって、わざとそっぽへ眼を落しているらしくも見えた。

【0】この陰気な一群いちぐんの人々は、ほとんど例外なしに似たり寄ったりの過去をもっているものばかりであった。【0】彼らはこうして暗い控室の中で、静かに自分の順番の来るのを待っている間に、むしろ華はなやかに彩いろどられたその過去の断片のために、急に黒い影を投げかけられるのである。【0】そうして明るい所へ眼を向ける勇気がないので、じっとその黒い影の中に立ち竦すくむようにして閉とじ籠こもっているのである。

【0】津田は長椅子の肱掛ひじかけに腕を載のせて手を額にあてた。【0】彼は黙祷もくとうを神に捧げるようなこの姿勢のもとに、彼が去年の暮以来この医者の家で思いがけなく会った二人の男の事を考えた。

【0】その一人は事実彼の妹婿いもとむこにほかならなかった。【0】この暗い室の中で突然彼の姿を認めた時、津田は吃驚びっくりした。【0】そんな事に対して比較的無頓着むとんじゃくな相手も、津田の驚ろき方が反響したために、ちょっと挨拶あいさつに窮したらしかった。

【0】他の一人は友達であった。【0】これは津田が自分と同性質の病気に罹かかっているものと思い込んで、向うから平気に声をかけた。【0】彼らはその時二人いっしょに医者の門を出て、晩飯を食いながら、性セックスと愛ラヴという問題についてむずかしい議論をした。

【0】妹婿の事は一時の驚ろきだけで、大した影響もなく済んだが、それぎりで後あとのなさそうに思えた友達と彼との間には、その後ご異常な結果が生れた。

【0】その時の友達の言葉と今の友達の境遇とを連結して考えなければならなかった津田は、突然衝撃ショックを受けた人のように、眼を開いて額から手を放した。

【1】すると診察所から紺こんセルの洋服を着た三十恰好がっこうの男が出て来て、すぐ薬局の窓の所へ行った。【0】彼が隠袋かくしから紙入を出して金を払おうとする途端とたんに、看護婦が敷居の上に立った。【0】彼女と見知り越ごしの津田は、次の患者の名を呼んで再び診察所の方へ引き返そうとする彼女を呼び留めた。【0】「順番を待っているのが面倒だからちょっと先生に訊きいて下さい。【0】明日あしたか明後日あさって手術を受けに来て好いかって」

【1】奥へ入った看護婦はすぐまた白い姿を暗い室へやの戸口に現わした。【0】「今ちょうど二階が空あいておりますから、いつでも御都合の宜よろしい時にどうぞ」

【1】津田は逃のがれるように暗い室を出た。【0】彼が急いで靴を穿はいて、擦硝子張すりガラスばりの大きな扉を内側へ引いた時、今まで真暗に見えた控室にぱっと電灯が点ついた。

6.3 注意力分析

6.3.1 最も注意されているトークン

注意力がある程度の解釈可能性を持っているかどうかを判断するために, 事例ごとから, 最も注意されている 10 個のトークンを取り出し, WordCloud を作成する. [SEP] と [CLS] の割合が大きすぎるため, この 2 つの特殊トークンを無視し, 残りのトークンのみで WordCloud を作成した.

小説データセットにおける, 最も注意されているトークンの WordCloud は図 6.5 と図 6.6 であり, 新聞データセットにおける, 最も注意されているトークンの WordCloud は図 6.7 と図 6.8 である. 上位 30 個のトークンを表 6.1 と表 6.2 に示す.

先行研究の中, 笠井ら [33] は, 小説データセットにおける分割非分割に関する有用な素性を報告した. 分割に関する有用な素性には, 「やがて」, 「しばらくは」などの素性があり, 非分割に関する有用な素性には, 「でしょ」, 「それでいて」, 「ただし」, 「けれども」などの素性がある. そこからいくつかの直感に合う知見を得ることができる. 例えば, 「やがて」, 「しばらくは」は段落の先頭に現れることが多く, 文中に「だだし」や「けれども」などの接続詞がある場合, この文で新しい段落が始まる可能性が低い.

本研究では, 図 6.5 と 6.6 より, BERT が「しかし」, 「けれども」などの接続詞に多くの注意を払っていることが分かる. 問題は, BERT が「津田」「た」「は」「に」などのトークンに最も注目する理由である.

一つの推測として, これらのトークンは, データセットで頻繁に出現し, 段落分割を判断するためではなく, 文の構造を理解するために, BERT によって注意が払われているということである.

表 6.1: 最も注意されているトークン (小説データセット)

順位	FL+AUX	FL
1	。	。
2	」	「
3	「	は
4	た	た
5	は	」
6	よ	津田
7	、	、
8	津田	が
9	の	の
10	ね	を
11	が	彼女
12	彼	ね
13	お	と
14	彼女	よ
15	を	僕
16	それ	に
17	に	から
18	しかし	小林
19	から	お
20	その	夫人
21	延	彼
22	そう	だ
23	と	て
24	##して	です
25	――	なかっ
26	小林	しかし
27	けれども	ない
28	で	延
29	も	叔父
30	だ	――

表 6.2: 最も注意されているトークン (新聞データセット)

順位	FL+AUX	FL
1	。	。
2	は	は
3	た	た
4	、	、
5	「	「
6	」	」
7	だ	で
8	で	いる
9	の	が
10	いる	だ
11	に	に
12	が	も
13	も	と
14	する	日本
15	ない	から
16	と	氏
17	ある	日
18	日	ます
19	から	この
20	し	私
21	年	の
22	いう	さん
23	さん	トランプ
24	月	です
25	を	を
26	ます	監督
27	です	その
28	氏	する
29	この	まし
30	人	一方

図 6.5: 最も注意されているトークンの WordCloud(小説、AUX+FL)

図 6.6: 最も注意されているトークンの WordCloud(小説、FL)

6.3.2 注意力より有用なトークンを識別

前述のように, トークンは, 段落分割を判断するためではなく, 文を理解する目的で注意されることがある. 段落分割に役立つトークンを識別するために, この章ではこれらの共通トークンを除外しようとする. 共通トークンとは, 図 6.9 のように, 分割点と判断する時よく注意されるトークンと、非分割点と判断する時よく注意されるトークンの重なる部分と考えられる.

小説データセットにおける FL+AUX を使用する場合, 最も注意力を集めた上位 100 個のトークンのうち, 67 個の共通トークンが存在し, そして, FL を使用する場合, 最も注意力を集めた上位 100 個のトークンのうち, 80 個の共通トークンが存在する. 共通トークンを除外した結果を表 6.3 と表 6.4 に示す.

表 6.3: 分割点と判断する時注意されるトークン (小説データセット, 共通トークンを除外)

順位	FL+AUX	FL
1	こう	.....
2	下	だろ
3	##子	##」
4	##君	下さい
5	細	ます
6	##女	##え
7	##笑	さ
8	こういう	ちゃ
9	継	かい
10	清	医者
11	来	ましょ
12	眼	質問
13	へ	のに
14	笑い	今度
15	こんな	こう
16	ここ	何
17	藤井	さん
18	質問	##女
19	看護	嘘
20	##刻	芝居
21	問	
22	挨拶	
23	やがて	
24	宅	
25	先	
26	##答	
27	医者	
28	聞い	
29	今度	
30	驚	
31	事実	
32	電車	
33	結婚	

表 6.4: 非分割点と判断する時注意されるトークン (小説データセット, 共通トークンを除外)

順位	FL+AUX	FL
1	##して	##して
2	けれども	すると
3	だ	も
4	なかっ	あっ
5	君	夫
6	僕	考え
7	だから	女
8	あなた	ところが
9	すると	笑っ
10	です	[UNK]
11	ただ	し
12	何	眼
13	それから	##笑
14	ある	話
15	##に	藤井
16	女	それから
17	さ	つまり
18	もし	病院
19	すぐ	堀
20	私	玄関
21	そんな	
22	##さん	
23	でしょ	
24	かい	
25	ば	
26	だけ	
27	もの	
28	それで	
29	彼ら	
30	同時に	
31	ない	
32	じゃ	
33	ところが	

図 6.7: 最も注意されているトークンの WordCloud(新聞、AUX+FL)

図 6.8: 最も注意されているトークンの WordCloud(新聞、FL)

新聞データセットにおける FL+AUX を使用する場合, 最も注意力を集めた上位 100 個のトークンのうち, 77 個の共通トークンが存在し, そして, FL を使用する場合, 最も注意力を集めた上位 100 個のトークンのうち, 63 個の共通トークンが存在しする. 共通トークンを除外した結果を表 6.5 と表 6.6 に示す.

これらの表から引き出せる結論は曖昧である. たとえば, 表6.3では, 小説データセットにおける分割点と判断する場合, どちらのモデルも「こう」に注意を向けるが, その理由はわからない. 一方, 私たちの直感と一致する結果もいくつかある. たとえば, 表 6.4 では, どちらのモデルも「ところが」と「それから」に注意を向ける. このような接続詞が現れると, 文が段落の先頭にある可能性は低い.

表 6.5: 分割点と判断する時注意されるトークン (新聞データセット, 共通トークンを除外)

順位	FL+AUX	FL
1	分	分
2	話す	時
3	市	官邸
4	られる	[UNK]
5	トランプ	大会
6	...	いう
7	語る	女子
8	県	事件
9	政府	男子
10	・	3
11	[UNK]	五輪
12	容疑	国会
13	首相	人
14	大会	北朝鮮
15	関係	被告
16	昨	4
17	出身	判決
18	=	会長
19	世界	容疑
20	語っ	米国
21	判決	調査
22	調査	昨
23	目指す	前
24		男性
25		出場
26		市
27		都
28		いじめ
29		県
30		2017
31		政権
32		時代
33		知事
34		離脱
35		チーム
36		委
37		会談

表 6.6: 非分割点と判断する時注意されるトークン (新聞データセット, 共通トークンを除外)

順位	FL+AUX	FL
1	それ	」
2	や	を
3	ため	ば
4	ば	ない
5	こと	て
6	たい	それ
7	そう	し
8	でも	ため
9	まし	話し
10	女性	こと
11	自分	ただ
12	発表	い
13	目	れる
14	##。	歳
15	円	……。
16	試合	指摘
17	多い	語る
18	指摘	か
19	なっ	練習
20	初めて	ほしい
21	今後	自分
22	き	例えば
23	できる	き
24		振り返る
25		述べ
26		そして
27		なる
28		全国
29		以来
30		だっ
31		たい
32		でも
33		思う
34		説明
35		発表
36		語っ
37		そう

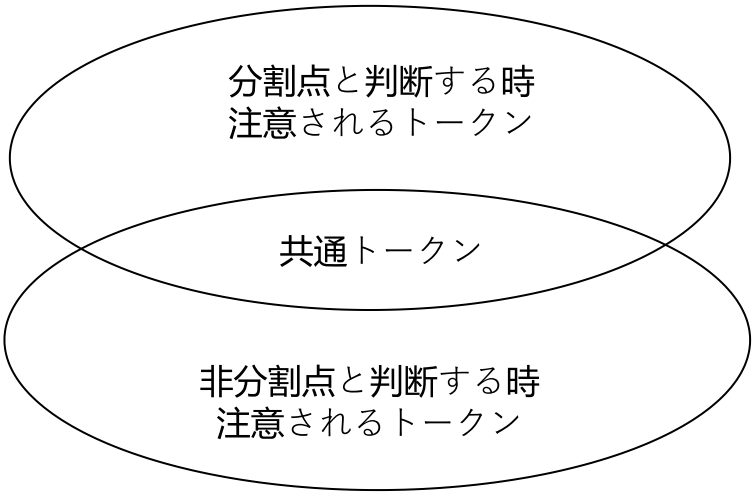

図 6.9: 共通トークン

表 6.7: 学習曲線からの平均 F 値と標準偏差

	Vanilla	FL	AUX	FL+AUX
Mean	0.6616	0.6599	0.6648	**0.6695**
Std.	0.0212	0.0270	0.0159	0.0171

6.4 学習曲線

さまざまな手法がさまざまな収束パターンにつながる可能性があることを考慮し, 図 6.10 は新聞データセットでの学習曲線を示す. y 軸は dev データセットで 3 つのモデルの平均 $F1$ スコアを表し, x 軸はイテレーションを表す. バッチサイズが 16 だったので, 最初のエポックは イテレーション 1, 715 で終了し, 2 番目のエポックは イテレーション 3, 430 で終了する.

図 6.10 から, すべてのモデルが最初のエポックの終わりに収束することがわかる. 各モデルの性能を図から区別するのは難しいため, 2 番目と 3 番目のエポックの間のすべてのポイントの平均と標準偏差を計算した. 結果は表 6.7 に示されている. この表より, 補助損失を使用したモデルが他のモデルよりも優れた性能を発揮し, より安定であることが分かる.

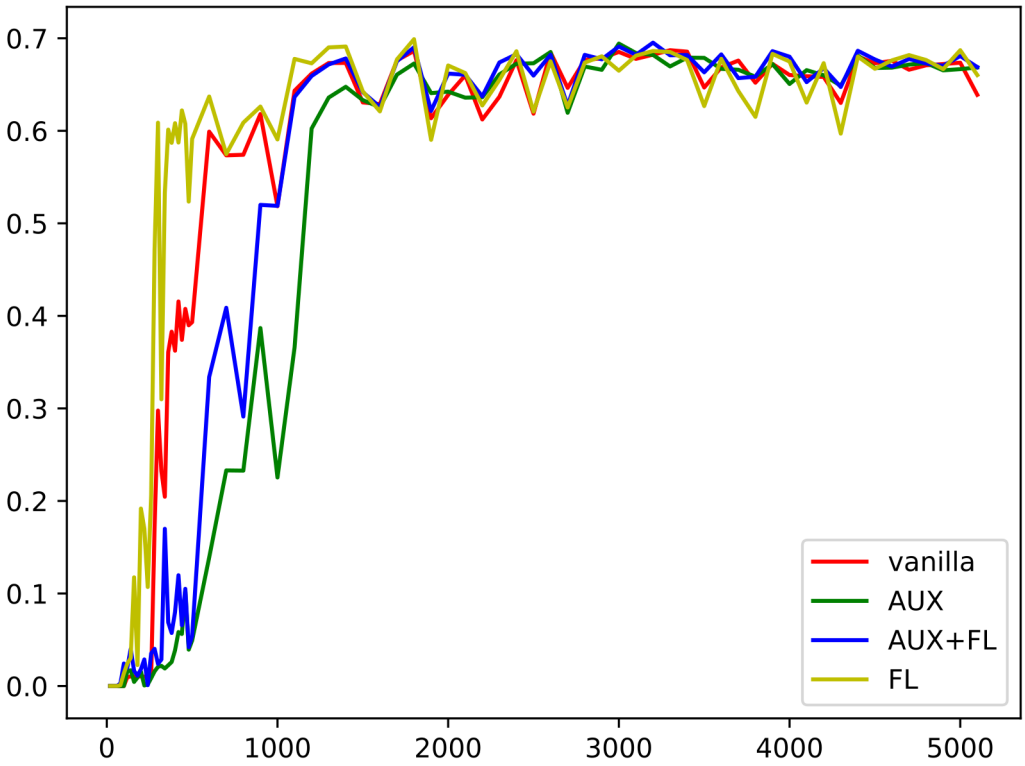

図 6.10: 毎日新聞データセットでの学習曲線

6.5 異なるプーリング戦略の影響

プーリング戦略 (pooling strategy) とは入力から文ベクトルを取り出す方法である. BERT の一般的な使い方では, [CLS] トークンに対応するベクトルを取り出す. 補助損失を使うために, 提案手法には, [CLS] の代わりに [SEP] に対応するベクトルをプーリング出力として使う.

さまざまな特殊トークン配置とプーリング戦略より性能の変化を調査するために実験を行った. 本実験で比較した特殊トークン配置とその性能を表 6.8 に示す. 太字でマークされた特殊なトークンの対応するベクトルをプーリング出力とする.

実験時間を短縮するために, 実験の設定は次のとおりである. 補助損失率を 0 に設定し（この実験の目的は特殊トークン配置とプーリング戦略の影響を調べたいだけであり, 補助損失を使わないようにする）, トレーニング エポック数を 2 に固定し, 毎日新聞データセットから未使用の記事 500 件をテストデータセットとして使う. 手法ごと

表 6.8: 異なるプーリング戦略と性能

手法名	プーリング戦略	Mean	Std.
AUX_ZERO	[CLS] s1 [SEP] s2 **[SEP]** s3 [SEP] s4	**0.6797**	0.0083
LEFT_SEP	[CLS] s1 [SEP] s2 **[SEP]** s3 s4	0.6773	0.0081
RIGHT_SEP	[CLS] s1 s2 **[SEP]** s3 [SEP] s4	0.6668	0.0076
NO_AUX_SEP	[CLS] s1 s2 **[SEP]** s3 s4	0.6604	0.0174
Vanilla	**[CLS]** s1 s2 [SEP] s3 s4	0.6684	0.0089
COUNTER_SEP	**[CLS]** s1 [SEP] s2 s3 [SEP] s4	0.6789	0.0125

に モデル 5 つをトレーニングして平均 F 値を計算する.

表 6.8 から導き出される結論は次のとおりである.

- NO_AUX_SEP と Vanilla の比較より, 補助的な [SEP] を使用しない場合, プーリング出力として [CLS] を使用する方が、 [SEP] を使用するよりも高い性能を達成できる.
- AUX_ZERO と NO_AUX_SEP の比較より, 補助的な [SEP] がモデルの性能を改善できる.
- LEFT_SEP, RIGHT_SEP と NO_AUX_SEP の比較より, 左側の補助 [SEP] を使用する場合, 性能の向上はより顕著である.

驚くべきことに, COUNTER_SEP の特殊トークン配置も優れた性能が得られた. モデルが明示的な指示トークンなしで学習目標を推測できることは注目に値する.

第7章　おわりに

本研究は, 毎日新聞および小説データセットでの段落分割を調査した. 我々は, 飯倉らの研究 [1] の上で, 補助損失を導入することでモデルの性能をさらに改善した. 実験結果より, 新聞データセットでは, BERT の基本的な使い方の平均 $F1$ スコアは 0.6728 で, 飯倉らの手法は 0.6704 である. 一方, 提案手法で得られた $F1$ スコアは 0.6801 であり, BERT の基本的な使い方よりも 0.007 高く, 飯倉らの手法よりも 0.0097 高くなっている. 小説データセットでも性能の向上が確認された. 統計的な検定より, 各手法によって得られた結果の差が統計的に有意であることが分かる.

補助損失は次の理由で有効である. まず, 連続した段落分割はほとんど発生しないため, 周囲に段落分割点があるかどうかを知ることが段落分割の判断に役立つ. 次に, ニューラルネットワークモデルは分割点の周囲の文脈に注目する傾向があり, 距離の遠い情報を有効に利用できない. そこで, 補助タスクを使って, モデルが距離の遠い情報に注目させることができる.

一方, 我々は新聞データセットに飯倉らの手法を適用した. その結果, 新聞データセットにおける BERT の基本的な使い方の平均 $F1$ スコアは 0.6728 で, 飯倉らの手法は 0.6704 であり, 性能は 0.002 低下している. その理由は, Focal Loss は, 小説データセットなど, データ不均衡問題が深刻であるデータセットに適するが, 他のデータセットには逆効果がある. したがって, データ不均衡の問題に応じて損失関数を慎重に検討する必要がある.

我々の研究成果は, 文章作成システムや Web テキストの整理に使用できる. 一方, 補助損失の使用は, 段落分割に限定されなく, 他のテキスト分割タスクにも使用できる. 今後の予定として, まず, 他のテキスト分割タスクに補助損失を適用したい. そして, 補助損失を使用し, より大きなウィンドウサイズの自動段落分割を研究したい.

謝辞

本研究を進めるにあたり，研究の進め方や本論文の書き方など，細部にわたる御指導を頂きました，鳥取大学大学院工学研究科情報エレクトロニクス専攻自然言語処理研究室の村田真樹教授に心から御礼申し上げます．また，本研究を進めるにあたり，御指導，御助言を頂きました，村上仁一准教授に心から御礼申し上げます．また，ご多忙の中，助言をいただきました櫛田大輔准教授，西山正志准教授に厚く御礼申し上げます．その他様々な場面で御助言を頂いた自然言語処理研究室の皆様に感謝の意を表します．

参考文献

[1] Riku Iikura, Makoto Okada, and Naoki Mori. Improving bert with focal loss for paragraph segmentation of novels. In *International Symposium on Distributed Computing and Artificial Intelligence*, pp. 21–30. Springer, 2020.

[2] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 6000–6010, Red Hook, NY, USA, 2017. Curran Associates Inc.

[5] Kevin Clark, Urvashi Khandelwal, Omer Levy, and Christopher D Manning. What does bert look at? an analysis of bert's attention. *arXiv preprint arXiv:1906.04341*, 2019.

[6] Irina Pak and Phoey Lee Teh. Text segmentation techniques: a critical review. *Innovative Computing, Optimization and Its Applications*, pp. 167–181, 2018.

[7] Marwa Naili, Anja Habacha Chaibi, and Henda Hajjami Ben Ghezala. Comparative study of word embedding methods in topic segmentation. *Procedia computer science*, Vol. 112, pp. 340–349, 2017.

[8] Michael Alexander Kirkwood Halliday and Ruqaiya Hasan. *Cohesion in english*. Routledge, 2014.

[9] Marti A Hearst. Text tiling: Segmenting text into multi-paragraph subtopic passages. *Computational linguistics*, Vol. 23, No. 1, pp. 33–64, 1997.

[10] Freddy YY Choi. Advances in domain independent linear text segmentation. *arXiv preprint cs/0003083*, 2000.

[11] Michel Galley, Kathleen McKeown, Eric Fosler-Lussier, and Hongyan Jing. Discourse segmentation of multi-party conversation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pp. 562–569, 2003.

[12] Olivier Ferret. Improving text segmentation by combining endogenous and exogenous methods. In *Proceedings of the International Conference RANLP-2009*, pp. 88–93, 2009.

[13] Martin Riedl and Chris Biemann. How text segmentation algorithms gain from topic models. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 553–557, 2012.

[14] Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, Vol. 41, No. 6, pp. 391–407, 1990.

[15] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[16] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543, 2014.

[17] Omri Koshorek, Adir Cohen, Noam Mor, Michael Rotman, and Jonathan Berant. Text segmentation as a supervised learning task. *arXiv preprint arXiv:1803.09337*, 2018.

[18] Pinkesh Badjatiya, Litton J Kurisinkel, Manish Gupta, and Vasudeva Varma. Attention-based neural text segmentation. In *European Conference on Information Retrieval*, pp. 180–193. Springer, 2018.

[19] Jing Li, Aixin Sun, and Shafiq R Joty. Segbot: A generic neural text segmentation model with pointer network. In *IJCAI*, pp. 4166–4172, 2018.

[20] Yizhong Wang, Sujian Li, and Jingfeng Yang. Toward fast and accurate neural discourse segmentation. *arXiv preprint arXiv:1808.09147*, 2018.

[21] Bryan McCann, James Bradbury, Caiming Xiong, and Richard Socher. Learned in translation: Contextualized word vectors. *Advances in neural information processing systems*, Vol. 30, , 2017.

[22] ME Peters, M Neumann, M Iyyer, M Gardner, C Clark, K Lee, and L Zettlemoyer. Deep contextualized word representations. arxiv 2018. *arXiv preprint arXiv:1802.05365*, Vol. 12, , 1802.

[23] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. 2018.

[24] Yuxuan Wang, Yutai Hou, Wanxiang Che, and Ting Liu. From static to dynamic word representations: a survey. *International Journal of Machine Learning and Cybernetics*, Vol. 11, No. 7, pp. 1611–1630, 2020.

[25] Michal Lukasik, Boris Dadachev, Gonçalo Simoes, and Kishore Papineni. Text segmentation by cross segment attention. *arXiv preprint arXiv:2004.14535*, 2020.

[26] Alessandro Solbiati, Kevin Heffernan, Georgios Damaskinos, Shivani Poddar, Shubham Modi, and Jacques Cali. Unsupervised topic segmentation of meetings with bert embeddings. *arXiv preprint arXiv:2106.12978*, 2021.

[27] Linzi Xing, Brad Hackinen, Giuseppe Carenini, and Francesco Trebbi. Improving context modeling in neural topic segmentation. *arXiv preprint arXiv:2010.03138*, 2020.

[28] Igor A Bolshakov and Alexander Gelbukh. Text segmentation into paragraphs based on local text cohesion. In *International Conference on Text, Speech and Dialogue*, pp. 158–166. Springer, 2001.

[29] Dmitriy Genzel. A paragraph boundary detection system. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 816–826. Springer, 2005.

[30] Caroline Sporleder and Mirella Lapata. Broad coverage paragraph segmentation across languages and domains. *ACM Transactions on Speech and Language Processing (TSLP)*, Vol. 3, No. 2, pp. 1–35, 2006.

[31] Katja Filippova and Michael Strube. Using linguistically motivated features for paragraph boundary identification. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pp. 267–274, 2006.

[32] Sarthak Jain and Byron C Wallace. Attention is not explanation. *arXiv preprint arXiv:1902.10186*, 2019.

[33] 笠井僚太. 日本語の段落分割における bert と最大エントロピー法の比較. 鳥取大学卒業研究発表会論文, 2021.