卒業論文

校閲時の事実確認作業における誤り箇所の自動推定

古田 朋也

2021 年 2 月 10 日

岐阜大学 工学部 電気電子・情報工学科 情報コース
鈴木研究室

本論文は岐阜大学工学部に
学士（工学）授与の要件として提出した卒業論文である.

古田 朋也

指導教員：
鈴木 優 特任准教授

校閲時の事実確認作業における誤り箇所の自動推定*

古田 朋也

内容梗概

本研究では校閲作業のうち，事実確認の支援を目的とする．事実確認の必要性が高い箇所を提示することによって，作業者の労力削減を図る．事実確認が必要となる文章は内容に誤りがある文章であり，文法上は正しい文章である．そのため，誤り箇所の特徴を人手にて定義することは難しい．そこで，内容に誤りを含む箇所の特徴抽出を機械学習によって行う．その際に，内容訂正が行われた事例を用いる．訂正事例を用いることで，内容誤りを含む文章を得ることができる．また，複数人で記述・編集が行われている文書を利用することで，誰もが間違えやすい箇所の特徴が抽出できる．獲得した内容誤りを含む文章の特徴を用いて，事実確認が必要となる箇所の推定を行う．この手法による校閲作業における実用性について，評価実験を行った．

キーワード

校閲作業，支援システム，文章分類，テキスト処理，機械学習

*岐阜大学 工学部 電気電子・情報工学科 情報コース 卒業論文, 学籍番号: 1173033129, 2021 年 2 月 10 日.

目次

図目次

表目次

第 1 章　はじめに

現在，インターネット上には校閲が行われていない記事が存在しているが，そのような記事は閲覧者に誤った情報を与えてしまう．ところが，校閲作業には時間と労力がかかる．文書の校閲は，誤字脱字の訂正と事実確認の二種類に大きく分けられる．誤字脱字訂正の支援はいくつか行われているが，事実確認作業の支援はほとんど行われていない．そこで，本研究では事実確認の作業に着目し，確認が必要となる箇所の推定結果を作業者に提示することによって，作業者の労力削減を図る．

本稿では，内容訂正に関する過去の事例を用いて，誤りの発生が見込まれる箇所の推定を自動で行い，作業者に提示するという手法を提案する．事実確認作業において確認が必要となる箇所は，誤りの発生が見込まれる箇所であり，訂正の事例を利用することにより自動推定することが可能であると考えた．

本研究では，内容に誤りがありそうな箇所の推定をすることによって，事実確認作業の支援を行う．作業を行う文書中には，年代などの数字を含む文章など，誰もが間違えやすい文章が存在していると考える．このような文章について事実確認するだけでも，確認作業としては効果的である．また，事実確認作業にて作業者が訂正すべき箇所は，文章の内容について誤りが発生している箇所であるため，確認作業を行う上で誤りの発生が見込まれる箇所を重点的に確認することも有効な方法である．そして，そのような文章を自動で判定することができれば，作業者は提示された文章に対して確認作業を行うだけでよくなる．しかし，ここで提示したい内容誤りの文章は，文法上は正しい文章である．そのため実際に事実関係を確認しないと，どの文章に誤りが発生しているのか判断することが難しい．また，誤りの特徴や訂正内容は多様であるため，誤字脱字を含む文章と異なり，人手にて誤りの特徴を定義することが困難である．そのため，BERT[1] を用いた分類器を構築することによって内容誤りを含む文章の特徴を抽出する．その際，実際に訂正された文書を用いることによって，どのような文章に訂正すべき箇所が存在していたか，どんな訂正内容であったかの特徴を訓練させる．訓練に大量の訂正事例を用いることによって，多様な内容誤りを含む文章が獲得できる．その結果，人手では判断することが難しい，内容誤りの特徴が抽出できるのではないかと考えた．そして，このような内容誤りの特徴抽出によって，事実確認箇所の自動推定が可能になると考えた．

内容誤りの発生しやすい文章や，その訂正内容についての特徴を得るためには，大量の訂正事例が必要となる．しかし，著者が一人だけの文書の訂正事例を大量に用いても，その著者の癖や間違えやすい箇所にしか対応できなくなってしまう．それでは，支援システムとしての汎用性に欠けてしまう．そのため，著者が複数である文書を用いることによって，誰もが間違えやすい箇所の特徴を獲得したい．そこで，複数人で編集が行われており，編集履歴を利用することによって大量の訂正事例の確保が可能となる Wikipedia の記事をデータとして使用することにした．データとして大量に用意したいのは，内容誤りを含む文章である．そのため，文書としては訂正前のものを使用するのが適当であると考えた．しかし，訂正前の文書だけでは，どこに誤りが含まれているのか判断することは困難である．そのため，一文ずつ確認作業を行い，ラベルを付与する必要が出てくる．これでは，内容誤りを含む文章を用意することに時間がかかる．大量の文章を用意する必要があるため，これでは効率が悪い．そこで，訂正後の文書も利用することで内容誤りを含む文章を収集する．訂正前と訂正後の文章対を用いて比較することによって，訂正された箇所とその訂正内容がわかる．これらを比較すると，主に 4 種類の文章が存在していた．一字一句そのままの文章，文章全体の意味は変わっていないが言い回しが訂正された文章，記述されている情報が追加・更新された文章，文章自体が削除または事実に誤りがあった文章である．この事実から，文章は訂正内容によって，4 種類の分類ができそうであると考えた．

データセット作成時のラベル付けを行う際，すべてを人手による比較で行うと，一記事あたり約 2〜3 時間程とかなり時間がかかった．記事内の文章の約 5 割が一字一句そのまま残っていた文章であったため，機械による比較を取り入れることで，作業時間の短縮が可能になると考えた．そのため，できる限り自動で付与することを考えた．まず，文書中から機械による比較によって，一字一句そのままの文章を抽出した．これにより，残りの文章すべてが訂正された文章となる．残った文章は何かの訂正が行われており，訂正前と訂正後の文章が異なる．そして，訂正内容もそれぞれ異なるため，これ以上機械で判断することは困難である．そのため，残りの文章に対しては，人手によって訂正前と訂正後の文章を比較することが適切であると考えた．人手による比較によって訂正内容を確認し，ラベルを振り分けた．機械による比較を取り入れたことにより，すべて人手にて行った場合のおよそ半分の

所要時間にてラベル付けを行うことができた．

作業者には，文章ごとの分類結果だけでなく，文章中の事実確認すべき箇所の提示を行う．その際，確認箇所として Attention[2] を利用することが適当であると考えた．Attention は，分類時に注意を向けた箇所を示しており，その分類項目に分類された原因ともいえる．そのため，その箇所が重要視すべき確認箇所ではないかと考えた．また，訂正内容の中にも，訂正の優先度があると考える．削除された文章は相応の理由があり削除されているため，そのような文章と類似している文章の確認作業は優先して行うべきである．情報の追加・更新された文章と類似している文章は，内容誤りほどではないが確認作業の優先度は高い．言い回しのみ訂正された文章や一字一句そのままの文章と類似している文章は，確認作業の優先度はそれほど高くない．そのため，確認作業の優先度が高い文章に対して，確認箇所を提示することが良いと考えた．

過去の事例を用いてデータセットを作成し，BERT による分類器を構築した．誤りの有無についての分類を行い，確認箇所として Attention を可視化して提示する．そして，上記手法を Web アプリケーションとして実装した．提案手法の評価実験として，評価値に基づくモデルの性能評価と，実例に基づく実用性の評価を行った．Accuracy だけに注目した場合，低いもので 0.47，高いもので 0.73 となった．また，本手法によって確認すべき文章と判定された文章は，校閲作業の観点で見た場合，確認作業が必要な文章となっていた．確認箇所についても，Attention をそのまま使用して問題ない結果がいくつか見られた．内容に関する誤りを含む文章は，誤字脱字を含む文章と比較して，誤りの特徴が掴みづらいため，事実確認作業の支援を行うことは困難であった．しかし，評価実験の結果から，過去の内容誤りに関する訂正事例を用いて，内容誤り箇所の推定を行うことにより，事実確認作業の支援が可能であることが明らかとなった．一方で，誤りの可能性が高い文章の取りこぼし，不適当と思われる箇所が確認箇所の中にいくつか見られる，など問題点もいくつか見つかった．校閲作業における実用性を向上させるために，上記の問題点について改善策を講じる必要がある．

本論文における貢献は以下の通りである．

- 内容誤りに関する訂正事例を用いて，内容誤り箇所の推 定を行うことにより，事実確認作業の支援が可能であることが明らかとなった．

- 事実確認すべき箇所の提示に，分類時の Attention が利用可能であることが明らかとなった．

本論文の構成は以下の通りである．2 章では本論文にて用いた技術や手法についての基本的事項を述べる．3 章では関連研究と提案手法の似ている点，異なる点について述べる．4 章では本論文の提案手法について述べる．5 章では評価実験の目的と内容，結果・考察について述べる．最後に 6 章では本論文のまとめと今後の課題について述べる．

第 2 章　基本的事項

本論文にて用いた技術や手法について，基本事項を述べる．

2.1　BERT

Transformer[2] の Encoder を使用した 12 層のニューラルネットワークモデル．モデルの構造を修正せずとも，転移学習することで，様々な自然言語処理タスクに応用できる汎用性の高いモデルとなっている．転移学習前の事前学習として MLM(Masked Language Modeling) と NSP(Next Sentence Prediction) を長い文章を含むデータセットを用いて行っている．MLM は複数箇所が穴になっている文章に対して穴埋め単語を予測するタスク，NSP は入力された二文が連続した文かどうかを判定するタスクとなっている．事前学習にて獲得したネットワークの重みを別のタスク用にファインチューニングすることで，高い精度を発揮することが期待できる．

2.2　Attention

文章などの系列データを扱う際に，要素ごとの関係性や注意を向ける箇所を学習する機構．Transformer はこの Attention のみを使用 した Encoder-Decoder モデルであるため，BERT にも存在する．BERT の Self Attention は 12 個の Multi head Attention で構成されており，一単語につき 12 個の数値が得られる．複数の Attention を用意することで、異なる特徴空間から，判定根拠を学習できる．それにより，判定根拠について異なる特徴空間からの情報を得ることができ，分類時などの判定根拠を取得する際に有益なものとなる．

2.3　交差検証

作成したモデルの評価をする際に行う検証方法．まず，データセットをいくつかに分割する．分割したうちの一つを評価用データ，残りを訓練用データとして訓練

を行う．評価用データを入れ替えることで，評価用データと訓練用データのすべての組み合わせで訓練が終了するまで，これを繰り返す．それぞれで得られた評価値の平均をとることで，モデルの汎化性能の評価を行う．また，データ数が少ないときにも有効な検証方法となっている．

2.4 評価指標

作成したモデルの評価のために用いる．表 2.1 のような評価用データ予測時の混同行列を基に導出され，主に Accuracy，Recall，Precision，F 値が使われる．

Accracy は式 (2.4.1) のように導出される．この数値は，単純な正解率を示しており，どれだけ正確に予測できているかを測ることができる．しかし，Accuracy だけでは不均衡データを扱う場合，不適切な評価になりかねない．例えば，正例のデータ数が極端に少ない場合，すべて負例であると予測すれば自ずと Accuracy は高い数値になる．しかし，正例と予測されることが一切ないため，良い予測とは言えない．そのため，Recall，Precision，F 値という評価指標を使用する．

Recall は式 (2.4.2) のように導出される．全正例データのうち，予測結果が正例とされたデータの割合を示しており，正例データをどれだけ取りこぼしなく予測することができたかを測ることができる．Recall は，予測するデータをすべて正例とすれば高い数値になる．

表 2.1　評価値導出のための混同行列

	正例予測	負例予測
正例ラベル	TP	FN
負例ラベル	FP	TN

TP　正例と予測されたうち，ラベルも正例であるデータの数.
FP　正例と予測されたうち，ラベルが負例であるデータの数.
FN　負例と予測されたうち，ラベルが正例であるデータの数.
TN　負例と予測されたうち，ラベルも負例であるデータの数.

Precision は式 (2.4.3) のように導出される．正例と予測されたデータのうち，ラベルも正例である割合を示しており，正例と予測した結果がどれだけ正しかったかを測ることができる．Precision は，確実なものだけ正例と予測して，判断が難しいものはすべて負例とすれば高い数値になる．

F 値は式 (2.4.4) のように Recall と Precision の調和平均をとることで導出される．これは，Recall と Precision の両方の値を考慮した評価値である．Recall と Precision は片方の数値が高くなるともう片方は低くなるように，トレードオフの関係となっている．そのため，調和平均をとることで二つの値を考慮した総合的な評価を行う．

どの数値を評価指標として重視するかはモデルのタスクによって異なる．例えば，癌検診であれば、癌である人を癌でないと誤って予測することは避けなければならない．そのため，Recall を重視することで，取りこぼしを少なくすることが適当である．一方，通販サイトのおすすめ機能であれば，何でもお勧めするのではなく，確実に興味を持ちそうなものを勧めるべきである．このような場合は，Precision を重視することで，確実性を高めることが適当である．このように，タスクによって適した評価指標を用いることが大切である．

$$Accuracy = \frac{TP + FP}{TP + FP + TN + FN} \tag{2.4.1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2.4.2}$$

$$Precision = \frac{TP}{TP + FP} \tag{2.4.3}$$

$$F\text{ 値} = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}} = \frac{2Recall \cdot Precision}{Recall + Precision} \tag{2.4.4}$$

第 3 章　関連研究

校閲作業に関連する研究は，いくつか行われている．高橋ら [3] は，誤字脱字箇所推定と訂正候補文字の提示についての手法を提案している．Bidirectional LSTM を用いて，文字毎に誤字かどうかの判定を行う確率モデルと周辺単語から単語間に入る文字を予測する言語モデルを構築し，入力文に対して，それぞれのモデルで判定を行う．これら二種のモデルの判定結果から得られた情報をランダムフォレストの入力とすることによって誤字脱字の有無を判定している．そして，誤字脱字が存在していた場合，確率モデルと言語モデルの出力から得られた情報によって訂正を行っている．

今村ら [4] は，日本語助詞誤りの訂正についての手法を提案している．条件付き確率場 (CRF) を用いることによって，誤り箇所に対して訂正候補文字を決定している．その際に，マッピング素性とリンク素性の二種類の素性を用いている．さらに，リンク素性については，n-gram 素性と言語モデル確率を併用することにより，訂正結果が日本語文として正しいかを測っている．

鈴井ら [5] は，日本語文章の不自然箇所検知についての手法を提案している．Yahoo!知恵袋と Wikipedia から，自然な日本語文章のみをデータとして使用することによって，ニューラルネットワークによる言語モデルを構築している．このモデルに，周辺単語のみを入力として，判定したい単語についての出現確率の推定を行う．この自然な日本語文章における出現確率の推定を行うことによって，不自然箇所を決定している．

校閲作業は大きく二種類の作業に分けられる．誤字脱字訂正を行う作業と事実関係の確認を行う作業である．上記の先行研究は二種類の作業のうち，誤字脱字訂正に関連する研究となっている．先行研究では，誤字脱字訂正についての研究が多く，事実確認作業についての研究は多くない．そこで本研究では，二種類の作業のうち，事実確認作業に焦点を当てた．なお，事実確認作業に関連する研究として，ファクトチェックについての研究が行われている．

内山ら [6] はファクトチェックにおける要検証記事の探索支援についての手法を提案している．Twitter におけるニュース記事に対する言及を利用することによって，検証の必要性を示唆する端緒情報である確率の推定をツイートごとに行う．そ

して，端緒情報である確率によってスコア付けを行い，記事ごとにまとめる．まとめたスコアを用いて，それぞれの記事について検証必要度のランク付けを行い，作業者へ提示している．これにより，作業者はランクの高い記事から検証していけば良いことになる．

上記研究では，ファクトチェック作業を行うかどうかの判断支援をしている．この支援では，ファクトチェックの作業自体には干渉していない．一方で，本研究では，校閲作業自体の支援を行う．過去の事例のみを用いて，事実確認が必要な箇所の推定を試みる．提案手法により，先行研究とは異なる作業支援を行うことにより，労力の更なる削減が可能であると考えた．

第 4 章　提案手法

本手法では，誤りの発生が見込まれる箇所の推定を分類問題として扱う．提案手法の概要を図 4.1 に示す．提案手法によるモデル構築は Step.A，モデル構築後の誤り箇所推定は Step.B の流れで行う．それぞれの詳細については，括弧内に示す節にて述べる．

Step A-1 過去の訂正事例から，訂正前と訂正後の二種類の文書を用意する．(4.1.1 節)

A-2 二種類の文書を比較することによって，四種類のラベルを付与する．(4.1.2 節)

A-3 過去の訂正事例を利用したデータセットにて，分類器のファインチューニングを行う．(4.2 節)

Step B-1 事実確認作業を行う文書を入力する．その際，文章単位に分割して，各文章ごとに確認箇所の推定を行う．(4.2.1 節)

B-2 構築した分類器を用いて，四種類に分類する．(4.2.1 節)

B-3 レベル 1, 2, 3 と判定された文章については，確認箇所として，Attention の可視化を行う．(4.3.1 節)

B-4 分類器による分類結果と，Attention による確認箇所を作業者へ提示する．

4.1　データセット作成

本手法では，与えられた文章に対して誤りが発生しているかどうかを判定する．そのため，訓練データとして誤りが発生している文章を用意して，分類器にその文章の特徴を訓練させる必要がある．そこで，過去の訂正事例を利用する．データとして大量に用意したいのは，内容誤りを含む文章である．そのため，文書としては訂正前のものを使用することが適当であると考えた．しかし，訂正前の文書だけでは，どこに誤りが含まれているのか判断することは困難である．そこで，訂正後の文書も利用する．訂正前と訂正後の文書を用意し，比較することによって誤りが発

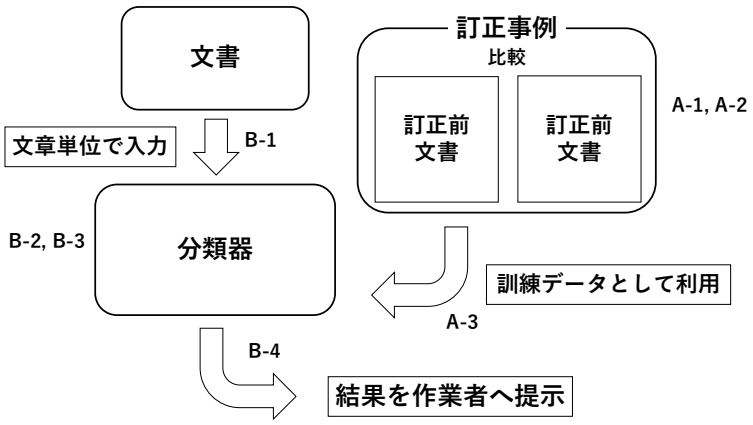


図 4.1　提案手法の概要

生していた文章を抽出する．訂正前の文章に対して，表 4.1 にて示すラベルを比較時に付与することによりデータセットを作成する．

4.1.1　使用データ

Wikipedia の編集履歴を利用して，訂正前の文書として過去バージョンの記事を，訂正後の文書として最新バージョンの記事を使用する．内容誤りの発生しやすい文章や，その訂正内容についての特徴を得るためには，大量の訂正事例が必要となる．また，誰もが間違えやすい箇所の特徴を獲得したいため，著者が複数である文書を使用したい．そのため，複数人で編集が行われており，編集履歴を利用することによって大量の訂正事例の確保ができる Wikipedia の記事をデータとして使用することにした．

ここで，過去の記事を選ぶ際に，二つの条件を考える．一つは，修正箇所ができるだけ多い古いバージョンの記事であること，もう一つは，最新バージョンと同程度の文章数を持つ記事であることである．古いバージョンであればあるほど，訂正された箇所は多くなる．しかし，あまりに古すぎると記事として不十分，不適切なものとなってしまう．そのため，上記の条件を設けることとした．条件を満たすた

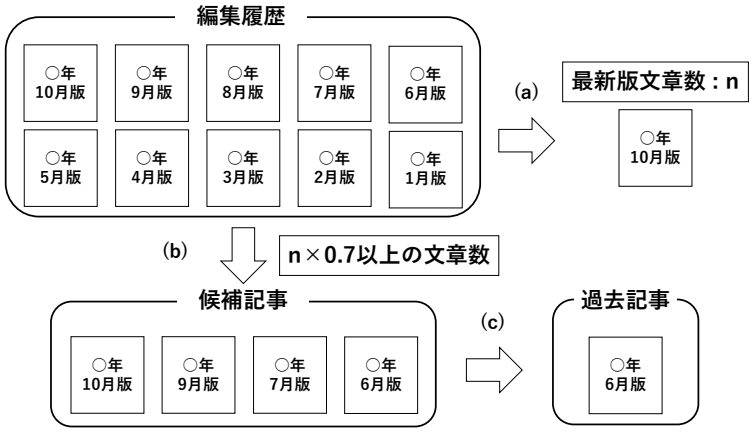


図 4.2　過去記事の選出方法

めに，図 4.2 の流れで過去の記事の選出を行う．

(a) Wikipedia の編集履歴から最新記事の文章数を求めることによって，基準となる文章数 n を決定する．
(b) 基準となる文章数 n の 7 割以上の文章数を持つ記事を編集履歴から抽出して，候補記事とする．
(c) 候補記事のうち，最も古いバージョンの記事を過去の記事として採用する．

4.1.2　作成方法

訂正前と訂正後の文章対を比較すると，主に 4 種類の文章が存在していた．そのため，本手法では，入力された文章を表 4.1 に示す 4 種類の項目に分類する．図 4.2 のように選出した過去の記事と最新の記事を比較することによって，過去の記事の文章に対応したラベルを付与する．二種類の記事を比較する際に，表 4.1 に示す基準に従い，ラベル付け作業を行う． まず，過去の記事を基準として，一字一句完全一致している文章が最新の記事に存在しているか否かの判定を機械によって自動で行う．それにより，レベル 0 のラベルを付与する文章を抽出する．残った文章

は何らかの訂正が行われており，訂正内容もそれぞれ異なる．そのため，残りの文章に対して，これ以上機械で判断することは困難であり，人手によって訂正前と訂正後の文章を比較することが適切であると考えた．残った文章に対して，人手による目視で訂正前と訂正後の比較を行うことによってレベル 1，2，3 のラベルを付与する．文章比較時の具体例を表 4.2 に示す．ここでは，過去の記事には誤った内容が含まれており，最新の記事には誤りが一切含まれていないという前提の下で比較作業を行う．また，この分類は文章の内容に関する誤り箇所の推定を目的としている．そのため，文章の比較時に誤字脱字については考慮しないこととしている．

表 4.1　分類ラベルと付与基準

ラベル	分類項目	付与基準
レベル 0	誤りの発生ほとんどなし	一字一句そのまま残っている
レベル 1	言い回しや表記についての誤り発生の可能性あり	言い回しや表記のみの変更，分割・統合された
レベル 2	情報の追加が見込める可能性あり	情報の追加・更新がされた
レベル 3	内容について誤り発生の可能性あり	削除または内容について訂正された

表 4.2　比較時の具体例

ラベル	訂正前	訂正後
レベル 0	2014 年 6 月に初ライブを行った（ただし、2014 年 5 月にシークレットでライブを行っている）	2014 年 6 月に初ライブを行った（ただし、2014 年 5 月にシークレットでライブを行っている）
レベル 1	米津が創造した架空のかいじゅうの**イラスト**を描き、その特徴、習性を紹介するという内容だった	米津が創造した架空のかいじゅうの**イラストレーション**を描き、その特徴と習性を紹介するという内容だった
レベル 2	徳島県立徳島商業高等学校を卒業後、大阪の美術専門学校に通いながらバンド活動を行う	徳島県立徳島商業高等学校を卒業後、大阪の美術専門学校に通いながら**「Ernst Eckmann」で**バンド活動を行う
レベル 3	ドットハック セカイの向こうに（**2012 年 1 月 21 日公開**） - 岡野智彦 役（声の出演）[37]	ドットハック セカイの向こうに（**2012 年 1 月 26 日公開**） - 岡野智彦 役 [66]

4.2　モデル構築

BERT は，あらゆる自然言語処理タスクにおいて，汎用性の高いモデルである．モデルの構造を修正せずとも，転移学習することによって，様々なタスクに応用でき，高い精度を発揮している．

内容に関する誤りを含む文章は，誤字脱字を含む文章と比べて，誤りの特徴が掴みづらい．それは，事実確認作業にて訂正される文章は，内容には誤りがあるが，文法上は正しい文章となっているためである．そこで，分類器として BERT を利用することにより，内容誤りの特徴を学習させる．4.1 節のように，内容訂正に関する過去の事例を用いて，内容誤りがある文章を収集する．BERT による分類器を構築する際に，収集した誤り文章を用いてファインチューニングを行うことによって，掴みづらかった内容誤りの特徴が抽出でき，分類を行うことが可能であると考えた．

本稿では，訓練済み BERT モデルとして，東北大学の乾・鈴木研究室の訓練済み日本語 BERT モデル*を使用した．訓練済みモデルは日本語版 Wikipedia にて，事前学習が行われており，語彙数は 32,000 となっている．訓練済み BERT モデル 12 層，これに出力層を 1 層加えた 13 層のモデルを転移学習にて構築する．4.1 節に従って作成したデータセットを用いて，ファインチューニングを行う．ファインチューニング時には，BERT モデルのパラメータは，最終層のみ更新するように設定する．

4.2.1　モデルによる分類

構築したモデルによる分類の流れを示す．まず，与えられた文章を形態素解析して，単語ごとに BERT モデルの語彙に対応した ID 化を行い，数値に変換する．形態素解析には，辞書として mecab-ipadic-NEologd を用いた MeCab を使用している．単語の ID 化は BERT の訓練済みモデルのトークナイザにて行う．ID 化した単語を w_i として，入力文章の単語列 $S = w_1, w_2, \cdots, w_n$ を構築したモデルに入力する．入力後，Embedding レイヤーによって，w_i を対応した単語ベクトルに変換

*https://github.com/cl-tohoku/bert-japanese

する．ここでは単語ベクトルとして，BERT の事前学習時に得られた 768 次元の分散表現を使用している．その後，BERT を使用した 12+1 層のモデルにて，計算を行う．その結果，4 次元のベクトルである出力 $O = (o_0, o_1, o_2, o_3)$ が得られる．この出力 O に対して，式 (4.2.1) のように定義される Softmax 関数を適用することにより．$O' = Softmax(O) = (o'_0, o'_1, o'_2, o'_3)$ を得る．O' の要素 o'_i は，とりうる値の範囲が $0 < o'_i < 1$ となっており，O' のすべての要素を足し合わせると 1 になる．そのため，O' の要素 o'_i は，入力文章 S がラベル (i) に属する確率を表している．よって，入力文章に対する判定は，O' の要素のうち，最大値となる要素に対応したラベルを選択することによって決定する．

$$Softmax(x_i) = \frac{\exp(x_i)}{\exp(x_1) + \exp(x_2) + \cdots + \exp(x_n)} \tag{4.2.1}$$

4.3 可視化

分類時の Attention を可視化することによって，確認箇所として利用する．Attention は，文章などの系列データを扱う際に，要素ごとの関係性や注意を向ける箇所を学習する機構であり，BERT にも存在している．この Attention を可視化することによって，入力データのどの部分に注目して分類を行ったのか，という分類時の判定根拠を確認することができる．文章中に誤りを含むとされた場合の判定根拠には，誤りの原因や特徴が含まれていると考える．誤りを含むと判定した根拠や原因を示す箇所の確認作業は重点的に行うべきである．そこで，分類時の Attention の可視化を行い，確認箇所として作業者へ提示することにより，確認作業の実施を促す．

本手法では，四種類への分類後，レベル 1，2，3 と判定された文章については確認箇所として Attention の可視化をする．分類項目の中にも，確認作業の優先度があると考える．内容誤りのを含む可能性が高いとされるレベル 3 は優先度が最も高い．情報の追加・更新が見込まれるレベル 2 も次点で優先度が高い．一方，レベル 0 は誤りがほとんど存在しないと判定された文章であり，作業の優先度は低い．そのため，レベル 0 とされた文章に対しては確認箇所としての Attention の可視化を

行わないこととする．

4.3.1 可視化方法

Attention の可視化方法について示す．可視化する Attention は，ファインチューニング時に，パラメータ更新を行った BERT の最終層の数値を使用する．BERT の Self Attention は 12 個の Multi head Attention で構成されており，一単語につき 12 個の数値が得られる．これら 12 個の数値は，それぞれが異なる特徴空間にて得られた判定根拠となっている．そこで，得られた 12 個の数値に対して，単語ごとに加算平均を計算することによって，確認箇所として利用する．判定結果を表示する際に，Attention の数値に応じて，背景色を単語ごとに変化させる．ある単語の Attention の加算平均結果を A として，背景色を構成する RGB の値をそれぞれ式 (4.3.1) のように決定する．

$$\begin{cases} R = 255 \times (1 - A) \\ G = 255 \\ B = 255 \times (1 - A) \end{cases} \tag{4.3.1}$$

決定した RGB の値を各単語の背景色に設定して表示する．Attention が強く掛かっている単語については背景色が濃い緑色に，Attention があまり掛かっていない単語については背景色が白色に近くなる．この可視化方法に従って，Attention を確認箇所として提示することによって，作業者が確認すべき箇所を直感的に把握できるようにする．

第 5 章　評価実験

提案手法について，以下の評価実験を行った．

実験 1 10 分割交差検証にて提案手法によるモデルの性能評価
　目的 1 提案手法による分類によって，どの程度の精度が出るのか確認する．
　目的 2 不均衡データの扱い方によって精度が変動するのか確認する．
実験 2 提案手法による推定の校閲作業における実用性評価
　目的 1 提案手法によるモデルにて，適当な判定が行われているか校閲作業の観点で確認する．
　目的 2 事実確認作業における確認箇所として，適当な箇所が提示されているのか確認する．
実験 3 追加実験
　目的 1 データ数の増加に伴う性能の変化を確認する．

データセット作成時に，データ源として 2020 年 10 月 16 日時点を最新記事とした日本語版 Wikipedia より，カテゴリ「日本の芸能人」に属する閲覧数が多い記事 16 件を使用した．Wikipedia 内の記事において，人物記事は閲覧数が多い記事となっており，そのような記事は優先して確認作業を行うべきであると考える．また，使用する記事はできるだけ訂正された箇所が多いものを採用したい．以上の理由で，更新が頻繁に行われている「日本の芸能人」に属する記事を対象として使用した．使用した記事内の文章 3,638 文に対し，4.1 節にて示した手順に従って著者がラベルを付与し，データセットを作成した．ラベルを付与した結果，データセット内のラベルの割合は，レベル 0 から順に，55%，27%，7%，11% となっており，ラベルの内訳はレベル 0 から順に 2,027 件，984 件，240 件，387 件となった．このデータセットを使用して，実験 1，2 を実施した．

5.1　実験 1

本節では，提案手法による分類モデルの性能評価を 10 分割交差検証による評価値に従って行う．また，データ使用方法を変えて実行し，それぞれの評価値を比較

することによって，不均衡データの扱い方による精度の変動具合の確認をする．

5.1.1 実験内容

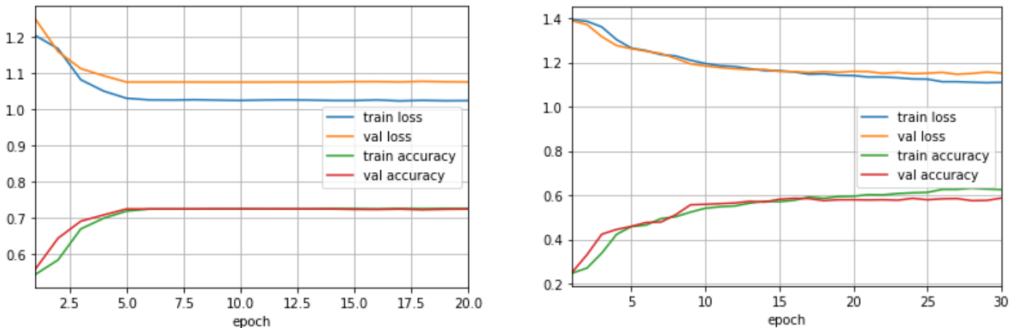
まず，前述のデータセットをラベルの割合を保った状態で 10 分割する．その後，訓練用データに対して，ラベル間の偏りを解消するための処理を行う．基準となるデータ数より多いラベルに対しては余剰データの削除を，少ないラベルに対しては同じデータを追加することによってラベル間のデータ数の偏りを解消した．以下四種類のデータ使用方法でそれぞれ 10 分割交差検証を行った．

(1) ラベル間の偏りは考えずにデータをそのまま使用
(2) データ数を最も少ないラベルに揃えて使用
(3) データ数を最も多いラベルの半数に揃えて使用
(4) データ数を最も多いラベルに揃えて使用

(1) から (4) までの訓練データの総数は，(1) から順に，3,274 件，864 件，3,648 件，7,296 件となり，ラベルの内訳は表 5.1 のようになった．BERT のファインチューニングは，まず 20 エポックを基準として行い，Validation Loss が下がりきっていない場合はエポック数を増やして再度実行するという方針をとった．その結果，(1) は 20 エポック，(2)，(3)，(4) は 30 エポックでの実行となった．それぞれの学習曲線は図 5.1 のようになった．

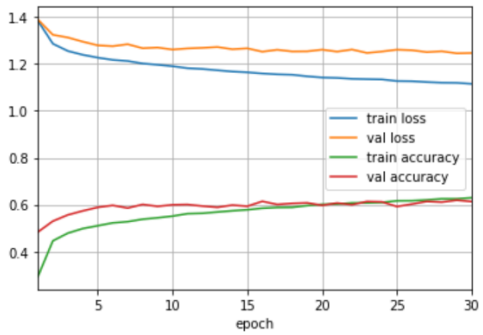
表 5.1　訓練データのラベル内訳

使用方法	レベル 0	レベル 1	レベル 2	レベル 3	総数
(1)	1,824	886	216	348	3,274
(2)	216	216	216	216	864
(3)	912	912	912	912	3,648
(4)	1,824	1,824	1,824	1,824	7,296

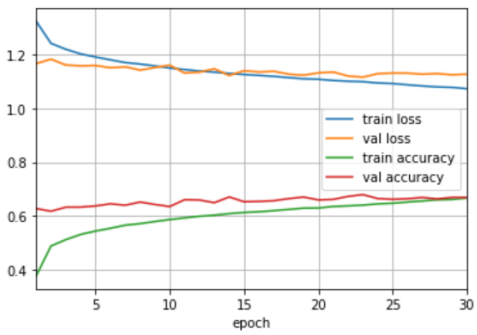


(1)

(2)



(3)



(4)

図 5.1　交差検証時の学習曲線

5.1.2　結果・考察

訓練後のそれぞれの評価値を表 5.2 に示す．Recall，Precision，F 値については，それぞれレベル 0，レベル 1，レベル 2，レベル 3 を正例とした場合の数値を示している．

Accuracy だけに注目した場合，低いものは 0.58，高いものは 0.72 という結果となった．(2)，(3)，(4) における数値のばらつきは，データ数の違いによる影響であると考えられる．一方，この分類タスクにおいて重要視するべきは Recall であると考える．確認作業において，誤りの可能性が低いものが高いと判定されることは問題ないが，高いものが低いと判定されることは問題となるためである．このような，誤りの可能性が高い文章の取りこぼしを少なくしたい．従って，今回はレベル 3 の Recall を重視するべきだと考える．レベル 3 についての Recall は，(2)，(3)，(4) で 0.4 程度の数値であった．この結果から，誤りの可能性が高い文章の取りこ

ぼしが，まだまだ多いことがわかる．取りこぼしが多いということは，まだまだ掴みきれていない内容誤りの特徴があることが考えられる．

四種類のデータ使用方法を比較することにより，不均衡データの扱い方による精度の変動具合の確認をする．(1) はデータの偏りが結果に顕著に表れており，特にレベル 2 とレベル 3 を正例とした場合の評価値が極端に低い値になっている．そのため，良い結果とは言い難い．一方，(2)，(3)，(4) におけるレベル 3 の Recall は，(2) は相対的に低く，(3) と (4) は同程度の値となった．また，Accuracy においては訓練データ数が最も多い (4) が最も高くなる結果となった．以上の結果より，(4) の最も多いラベルにデータ数を揃えて使用する方法が，この四種類の中では総合的に良い精度が得られることが確認できた．

表 5.2　データ使用方法別の評価値

使用方法	評価値	レベル 0	レベル 1	レベル 2	レベル 3
(1)	Accuracy		0.724		
	Recall	0.965	0.685	0.000	0.002
	Precision	0.728	0.716	0.000	0.014
	F 値	0.830	0.700	0.000	0.004
(2)	Accuracy		0.586		
	Recall	0.791	0.637	0.491	0.425
	Precision	0.609	0.578	0.629	0.573
	F 値	0.683	0.597	0.549	0.480
(3)	Accuracy		0.613		
	Recall	0.769	0.583	0.437	0.387
	Precision	0.688	0.754	0.357	0.355
	F 値	0.723	0.656	0.387	0.365
(4)	Accuracy		0.668		
	Recall	0.796	0.577	0.425	0.382
	Precision	0.809	0.736	0.278	0.310
	F 値	0.801	0.646	0.328	0.336

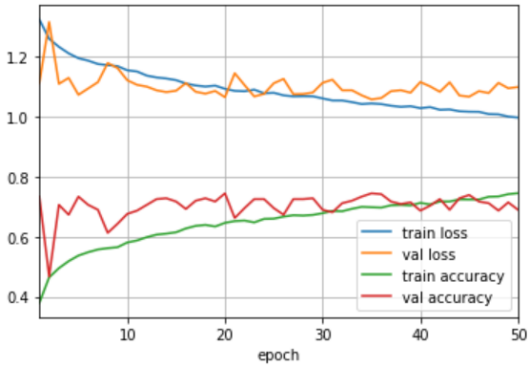


図 5.2　採用モデルの学習曲線

5.2　実験 2

本節では，校閲作業における実用性を確かめる．提案手法による確認箇所の推定を行い，結果を確認・分析することによって事実確認作業に適用可能かどうかの評価をする．

5.2.1　実験内容

5.1 節の結果から，(4) のデータ数を最も多いラベルに揃えて使用する方法を採用して，モデルを作成した．訓練データの総数は 7,296 件，ファインチューニングの最大エポック数は 50 エポックで訓練を行った．また，訓練時に Validation Loss が 15 エポック改善しなければ，Early Stopping を実行して訓練を停止するように設定した．そのため，訓練後のネットワークの重みは Validation Loss が最も低かった時の重みとなる．モデルはいくつか作成した中で，重視するべきレベル 3 の Recall ができる限り高い値となったものを採用した．採用されたモデルは Early Stopping の実行により，50 エポック目で終了したため，訓練後のネットワークの重みは 35 エポック目のものとなった．採用したモデルの学習曲線は図 5.2 のようになった．採用したモデルによる評価用データ分類時の混同行列を表 5.3 に，それに基づく評価値を表 5.4 に示す．表 5.4 における，Recall，Precision，F 値は，それぞれレベル 0，レベル 1，レベル 2，レベル 3 を正例とした場合の数値を示している．

このモデルを Ruby on Rails にて作成した Web アプリケーションに導入して，文書の判定を行った．レベル 1，2，3 と判定された文章に対しては分類結果だけでなく，確認箇所として Attention も可視化して提示している．ここでは，データセットに含まれていない文書の判定を行い，結果を確認することにより評価を行う．判定を行う文書には、データセットと同様に，日本語版 Wikipedia のカテゴリ「日本の芸能人」に属する記事を使用した．

5.2.2 結果・考察

判定結果の一部を図 5.3 に示す．レベル 2 と判定された文章の多くが，情報の追加・更新を行うことができそうな文章となっていた．また，レベル 3 と判定された文章の多くが，確認が必要となる文章となっており，概ね期待通りの結果が得られている．例えば，図 5.3 のレベル 2 については，日付や役名の追加，「予定」と記載してある箇所の更新などを行うことができそうである．レベル 3 については，3

表 5.3　採用モデルの評価用データ分類時の混同行列

		分類結果			
		レベル 0	レベル 1	レベル 2	レベル 3
ラベル	レベル 0	174	8	11	10
	レベル 1	17	68	4	10
	レベル 2	8	4	10	2
	レベル 3	9	10	0	19

表 5.4　採用モデルの評価値

評価値	レベル 0	レベル 1	レベル 2	レベル 3
Accuracy	0.744			
Recall	0.857	0.719	0.416	0.500
Precision	0.836	0.755	0.400	0.463
F 値	0.846	0.719	0.408	0.481

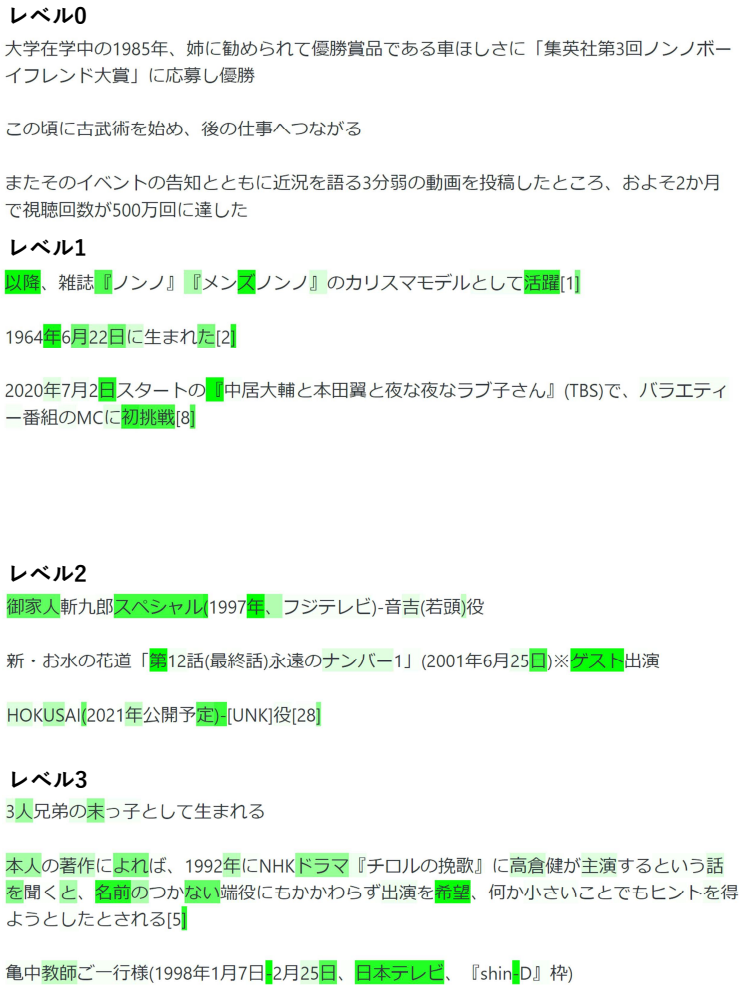


図 5.3　判定結果の一部

人兄弟の末っ子かどうか，本当に 1992 年なのか，というような事実確認が必要となる．

提示した確認箇所について，レベル 2 においては，多くが日付や役名などの追加や更新ができそうな箇所の前後に存在する結果となっていた．レベル 3 においては，そのままでも確認箇所と見なせる箇所が提示されていたが，中には不適当と感じる箇所もいくつか提示されていた．Attention を可視化することによって確

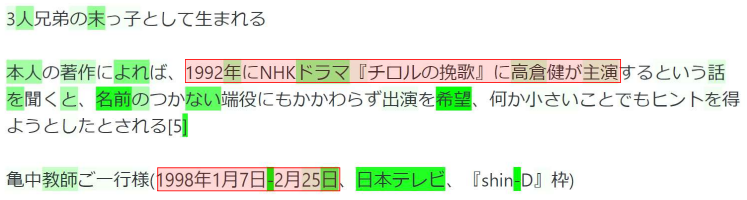


図 5.4　確認箇所の改善案

認箇所とする方法については，適当な箇所が提示されることが多かった．しかし，Attention をそのまま提示すると，レベル 3 にて見られたように，確認箇所として不適当な箇所も中には存在している．不適当な確認箇所となる箇所は，Attention が独立した部分のみに掛かっており，確認内容について断片的な情報しか持たない．そのため，この箇所のみ提示されても，何について事実確認をするべきなのか把握することは困難である．改善案の一つとして，Attention とその周辺単語を含めた一定のまとまりを作り，確認箇所として提示することが考えられる．例えば，図 5.4 にて，赤枠で示された部分は，Attention のみを提示した場合，「年」「ドラマ」「日」などの単語しか提示されない．このような箇所に対して，Attention とその周辺単語を含めた赤枠のようなまとまりを作り提示する．周辺単語を含めることにより，情報が補足され，何について事実確認をすればよいかが直感的に把握できるようになる．このように，提示方法に一工夫加えることによって，確認箇所として，より良いものになるのではないかと考える．

一方で，レベル 0 や 1 には年代などの数字を含む文章がいくつか分類されていた．年代などの数字を含む文章は，事実確認作業において重要視したい文章の一つである．このような結果となった原因の一つとして，過去の事例にて，数字に関する誤りが少なかったことが考えられる．訓練データに過去の事例のみを用いているので，訓練データに存在しなかった誤りは推定できない．校閲作業における実用性

を高めるためには，この問題点の解決は必要不可欠である．

5.3 実験 3

実験 1，2 において，レベル 3 の Recall が低く，内容誤り文章の取りこぼしが発生していた．その原因として，訓練データ数が少ないことにより，誤りの特徴抽出が不十分であったり，誤りパターンが不足していることが考えられる．訓練データに存在しない誤りのパターンには対応できないため，訓練データ数を増加させることによって性能の改善が期待できると考えた．そこで，データ数をおよそ 2 倍に増加させ，実験 1，2 と同じ条件で 10 分割交差検証と実用性評価を行う評価実験を再実施することによって，データ数の増加に伴う性能の変化を確認した．

5.3.1 実験内容

日本語版 Wikipedia より，カテゴリ「日本の芸能人」に属する 11 件の記事の文章に対して，新たにラベルを付与した．これを実験 1，2 で使用したデータセットに追加して，追加実験を実施した．データを追加した結果，データ数が 6,328 件，ラベルの割合はレベル 0 から順に，56%，28%，6%，9% となっており，ラベルの内訳はレベル 0 から順に，3,541 件，1,794 件，402 件，591 件となった．

10 分割交差検証において，5.1 節と同様にして，以下のようにラベル間の偏りを解消する．

(1) ラベル間の偏りは考えずにデータをそのまま使用
(2) データ数を最も少ないラベルに揃えて使用
(3) データ数を最も多いラベルの半数に揃えて使用
(4) データ数を最も多いラベルに揃えて使用

その結果，訓練データの総数は，(1) から順に，5,696 件，1,448 件，6,376 件，12,748 件となり，ラベルの内訳は表 5.5 のようになった．BERT のファインチューニングは，まず 30 エポックを基準として行い，Validation Loss が下がりきっていない場合はエポック数を増やして再度実行するという方針をとった．その結果，(1) から

(4) まで，すべて 30 エポックでの実行となった．10 分割交差検証による評価値を 5.1 節の結果と比較することによって，データ数を増加させたことによる分類精度の変動具合を確認する．

実用性の評価においても，5.2 節と同様に，(4) のデータ数を最も多いラベルに揃えて使用する方法を採用して，モデルを作成した．訓練データの総数は 12,748 件，ファインチューニングの最大エポック数は 50 エポックで訓練を行った．モデルはいくつか作成した中で，重視するべきレベル 3 の Recall ができる限り高い値となったものを採用した．採用されたモデルは，最大エポックの 50 エポックまで訓練が実行され，訓練後のネットワークの重みは 46 エポック目のものとなった．モデルの評価用データ分類時の Accuracy は 0.691，レベル 3 の Recall は 0.440 であった．作成したモデルに，5.2 節にて用いた文書と同じものを判定させた．その結果を 5.2 節と比較することによって，訓練データ数の増加による判定結果の変化について確認する．

5.3.2 結果・考察

10 分割交差検証の結果は，表 5.6 のようになった．Recall，Precision，F 値については，それぞれレベル 0，レベル 1，レベル 2，レベル 3 を正例とした場合の数値を示している．表 5.7 に実験 1 と実験 3 の評価値の比較結果を示す．なお，Recall についてはレベル 3 を正例とした場合の数値を示している．表 5.7 の比較結果から，データ数を増加させたことによる分類精度の大きな変動は見られなかった．

訓練データ数の増加によって，判定結果が変化した文章の一例を表 5.8 に示す．データの増加前では，年代などの数字を含む文章の取りこぼしが発生していた．増

表 5.5　訓練データのラベル内訳

使用方法	レベル 0	レベル 1	レベル 2	レベル 3	総数
(1)	3,187	1,616	362	531	5,696
(2)	362	362	362	362	1,448
(3)	1,594	1,594	1,594	1,594	6,376
(4)	3,187	3,187	3,187	3,187	12,748

加後の判定では，依然としてレベル 0 や 1 と判定される文章も見られたが，増加前と比較すると数字を含む文章の取りこぼしが少なくなっていた．一方，確認箇所として可視化した Attention については，データの増加前後で大きな変化は見られなかった．

これらの結果から，訓練データ数を増加させることが，データ数が少ないことによる誤りパターンの不足などに対して，一定の効果を発揮することが確認できた．一方で，レベル 3 の Recall が低いことによる内容誤り文章の取りこぼしについては，改善が見られなかった．モデル性能の向上のために，訓練データ数を増加させることは，多様な誤りのパターンに対応できるようになる点では無意味ではない．しかし，データ数の増加前後で分類性能に大きな変化が見られなかったことから，

表 5.6　データ使用方法別の評価値 (追加実験)

使用方法	評価値	レベル 0	レベル 1	レベル 2	レベル 3
(1)	Accuracy		0.732		
	Recall	0.966	0.676	0.000	0.000
	Precision	0.731	0.735	0.000	0.000
	F 値	0.832	0.704	0.000	0.000
(2)	Accuracy		0.479		
	Recall	0.716	0.629	0.323	0.248
	Precision	0.452	0.517	0.539	0.494
	F 値	0.550	0.566	0.392	0.322
(3)	Accuracy		0.602		
	Recall	0.721	0.627	0.407	0.301
	Precision	0.655	0.782	0.288	0.305
	F 値	0.683	0.695	0.334	0.299
(4)	Accuracy		0.646		
	Recall	0.752	0.605	0.390	0.306
	Precision	0.803	0.732	0.219	0.239
	F 値	0.774	0.662	0.273	0.257

現状の分類モデルでは，表 5.6 程度の精度が限界のように考えられる．そのため，新たな特徴量を追加する，別の手法と組み合わせる，など別の改善策を考える必要がある．

表 5.7　データ数増加前後の評価値比較

	Accuracy		Recall	
使用方法	増加前	増加後	増加前	増加後
(1)	0.724	0.732	0.002	0.000
(2)	0.586	0.479	0.425	0.248
(3)	0.613	0.602	0.387	0.301
(4)	0.668	0.646	0.382	0.306

表 5.8　判定結果が変化した文章の一例

レベル 0 → 3	1994 年、『しのいだれ』（細野辰興監督作品）で、憧れだった役所広司と共演を果たし、『凶銃ルガー P08』と 2 本併せて日本映画プロフェッショナル大賞・特別賞を受賞
レベル 0 → 3	またそのイベントの告知とともに近況を語る 3 分弱の動画を投稿したところ、およそ 2 か月で視聴回数が 500 万回に達した
レベル 0 → 3	好きな日本人女優ランキングでは、2020 年 9 月現在第 4 位である
レベル 1 → 3	阿部 寛（あべ ひろし、1964 年 6 月 22 日 [1] - ）は、日本の俳優、モデルである
レベル 1 → 3	女優の仲間由紀恵とはドラマ「トリック」で共演していた [10]
レベル 3 → 1	同年以降は俳優を主として活動するも、ファッションモデル出身という肩書きと顔立ちから、ありきたりな（本人は当時の事を「フェラーリで乗り付けるような」と語っている）二枚目の役しか与えられなかった

第 6 章　おわりに

本稿では，内容訂正に関する過去の事例を用いて，誤りの発生が見込まれる箇所の推定を自動で行い，作業者に提示するという手法を提案した．事実確認作業において確認が必要となる箇所は，誤りの発生が見込まれる箇所であり，訂正の事例を利用することにより自動推定することが可能であると考えた．訓練データとして，訂正前の内容誤りを含む文章が必要となる．しかし，内容誤りの文章は文法上正しい文章であるため，訂正前の文書だけではどの文章に誤りがあるのか把握することが困難である．そのため Wikipedia の編集履歴を用いて，訂正前と訂正後の文書を使用した．二つの文章対を比較して，内容誤りを含む文章を抽出し，訂正内容ごとにラベルを付与することによって，データセットを作成した．作成したデータセットを用いて，分類器とする BERT のファインチューニングを行い，内容誤りを含む文章の特徴を抽出した．構築した分類器による文章ごとの判定によって，作業者に対して，内容誤りを含む可能性が高い文章の提示を行った．また，事実確認の優先度が高い文章においては，文章中の確認箇所を提示することにより，作業の実施を促す．その際，分類時の Attention を可視化することによって確認箇所とした．上記提案手法により，本研究の目的である校閲作業における事実確認作業の支援が可能になり，作業者の労力削減につながると考えた．そして，提案手法についての評価実験を行った．提案手法による分類において，誤り発生の可能性が高いと判定された文章は，校閲作業の観点で見た場合，確認作業が必要な文章となっていた．また，Attention をそのまま確認箇所とする手法については，適当な箇所が提示されることが多く，概ね期待通りの結果が得られた．これらの結果から，過去の内容誤りに関する訂正事例を用いて，内容誤り箇所の推定を行うことによって，事実確認作業の支援が可能であることが明らかとなった．

しかし，誤り発生の可能性が高い文章の取りこぼし，不適当と思われる箇所が確認箇所の中にいくつか見られる，年代などの数字を含む文章が確認が不要であると判定されてしまう，など問題点もいくつか見つかった．今後，これらの課題を解決していき，校閲作業における実用性を高めていく．そのために，分類時に新たに別の特徴を使用する，確認箇所の提示方法に一工夫加えるなどの改善策を考え，適用していく．

また，本稿ではカテゴリ「日本の芸能人」に属する記事を対象に評価実験を行ったが，校閲作業への適用のためには，それ以外の文書も推定することによって提案手法の汎用性を確認する必要がある．そして，本稿での実用性の評価は著者の主観によるものが大きい．そのため，第三者がこのシステムを使用した際の評価も取り入れていかなければならない．

謝辞

本研究を進めるにあたって、指導教員である鈴木優特任准教授にたくさんのご指導，ご助言をいただきました．事務補佐員の井尾さんには様々な手続きをするにあたって，お世話になりました．また，同じ鈴木研究室の皆様には本研究について参考になる様々なご意見をいただきました．本論文を書き終えることができたのは，支えてくださった皆様のおかげです．心より感謝申し上げます．

参考文献

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pp. 5998–6008, 2017.

[3] 高橋諒, 蓑田和麻, 舛田明寛, 石川信行. Bidirectional lstm を用いた誤字脱字検出システム. 人工知能学会全国大会論文集, Vol. JSAI2019, pp. 3C4J903–3C4J903, 2019.

[4] 今村賢治, 齋藤邦子, 貞光九月, 西川仁. 識別的系列変換を用いた日本語助詞誤りの訂正. 言語処理学会第 18 回年次大会, pp. 18–21, 2012.

[5] 鈴井克徳, 若林啓. ニューラルネットワークを用いた日本語学習者の文章における不自然箇所検知. 第 10 回データ工学と情報マネジメントに関するフォーラム（DEIM2018）, 2018. online(G3-4).

[6] 内山香, 鈴木海渡, 田上翼, 塙一晃, 乾健太郎, 小宮篤史, 藤村厚夫, 町野明徳, 楊井人文, 山下亮. ファクトチェックのための要検証記事探索の支援. 人工知能学会全国大会論文集 第 32 回全国大会 (2018), pp. 4Pin126–4Pin126. 一般社団法人 人工知能学会, 2018.

発表リスト

[1] 古田朋也，鈴木優『校閲作業のための LSTM による確認箇所抽出』第 19 回情報科学技術フォーラム, 2020.
[2] 古田朋也，鈴木優『校閲作業のための LSTM による確認箇所抽出』東海関西データベースワークショップ, 2020.
[3] 古田朋也，鈴木優『校閲時の事実確認作業における誤り箇所の自動推定』第 13 回データ工学と情報マネジメントに関するフォーラム, 2021.