

疑似訓練データを用いたニュース記事間の続報判定

松本 直彰[†] 湯本 高行^{††} 山本 岳洋^{††} 大島 裕明^{†,††}

[†] 兵庫県立大学 大学院応用情報科学研究科 〒650-0047 兵庫県神戸市中央区港島南町 7-1-28

^{††} 兵庫県立大学 大学院情報科学研究科 〒651-2197 兵庫県神戸市西区学園西町 8-2-1

E-mail: [†]{aa20y510,ohshima}@ai.u-hyogo.ac.jp, ^{††}{yumoto,t.yamamoto}@sis.u-hyogo.ac.jp

あらまし あるトピックについてニュース記事から調べる時に、出来事が起こった順に読み進めるとスムーズな理解につながる。また、トピックのまとめを作成する時に、ユーザが必要だと考えて手元に残した記事の続きがどの記事となるのかが分かればスムーズにまとめが作成できる。本研究では、BERT を用いて 2 つの記事の続報関係の有無を判定するモデルを作成する。モデルのファインチューニングを行う際、通常であれば人手で作成した大量の続報記事ペアを訓練データとして必要とし、多大な労力が必要となる。そこで、大量の人手で作成する続報記事ペアに代わり、同一記事内の段落間の関係に着目することで、リード段落とそれに続く段落を疑似訓練データとして扱うことを提案する。これで、人手で作成した少量の続報記事ペアを検証データとして用意するだけで続報判定モデルが作成できる。また、入力に用いる 2 つの記事の段落ペアにおいて、共通する名詞を MASK トークンに置き換えて学習させ、続報判定において文脈に特化したモデルの作成を行う。

キーワード 情報検索, デジタル図書館, 自然言語処理, 可視化, 情報要約

1 はじめに

過去に起こった出来事について知りたい時、私たちはニュース記事を収集することで調べることがある。調べる出来事が時間的に古くなるほどニュース記事は入手しにくくなっていく。新聞会社などの報道にかかわる企業の一部は、過去数十年間にも及ぶニュース記事を保存した記事アーカイブを提供している。記事アーカイブを検索することで過去のニュース記事を容易に入手することができる。そして、大きな災害からの復興などの長期間にわたるトピックについて調べる際には、記事アーカイブから検索した内容をまとめることで私たちは必要な情報を入手する。

しかし、記事アーカイブ検索には様々な問題が存在する。例えば、調べた情報をまとめる際に記事同士の関係が分からなくなることや、記事間の移動がスムーズに行えないことが挙げられる。

既存の記事アーカイブの検索インタフェースは Web 検索インタフェースのような形態である。ユーザがあるトピックについて検索を行う場合、調べた内容をまとめる時に必要となる記事をブックマークとして保存する。従来のインタフェースではブックマークした記事の個々の情報のみしか保存されない。しかし、まとめを作成する時にユーザはある程度小さなまとまりに記事内容を集約するため記事間の関係を必要とする。検索を行っていくうちに検索初期にブックマークした記事同士の関係は分からなくなっていく。そのため、まとめを行う時に記事同士の関係を思い出すため再度記事を読み返すことがあり、まとめをスムーズに行えない。

また、記事検索の最中にブックマークした記事に関係する内容の記事に出会った場合、内容を確認するために対象記事へと

移動することがある。この時、ユーザは検索画面と記事閲覧画面を行き来しながらトピックを調べている。既存のインタフェースで該当記事に到達するには、ブラウザバックを複数回押すか、ブックマークした記事を表示する画面に移動して該当記事を見つける作業が必要となる。どちらもユーザに複数回の画面操作を要求するため、スムーズな検索の妨げになる。

これらの問題に対して我々は、ユーザが気に入った記事の関係性を確認しながら検索を行え、記事アーカイブでの効率的な検索とまとめ作成が行えるように、ブックマーク記事の続報関係をグラフ化した探索者ビューを提案している [12]。また、探索者ビューを既存の記事アーカイブ検索インタフェースに組み込んだ新しいインタフェースを作成し、ユーザに既存のインタフェースとの使用感の違いを調査した [13]。

しかし、提案した記事間の言及関係を続報関係とみなす方法では、記事の続報関係をうまく発見できていないことが分かった。そこで本稿では、自然言語処理で使われる機械学習モデルである BERT [3] を用いて、入力した 2 つの記事ペアが続報かどうかを判定する記事間続報判定モデルを作成する。2 つの記事ペアにおいて、時間的に古い記事を先行記事、時間的に新しい記事を続報候補記事と呼ぶ。図 1 は記事間続報判定モデルの概要を示したものである。

2 節では本稿に関連する研究について述べる。3 節では記事間続報判定モデルについて述べる。中でも 3.1 節では続報判定における文脈情報の必要性について、3.2 節では BERT モデルのファインチューニングについて、3.3 節ではファインチューニングに用いる訓練、検証、テストデータの作成について、3.4 節では BERT モデルの学習結果について、3.5 節では記事間続報判定モデルの入出力とその結果について述べる。4 節ではユーザによる記事間続報判定モデルの評価について述べる。

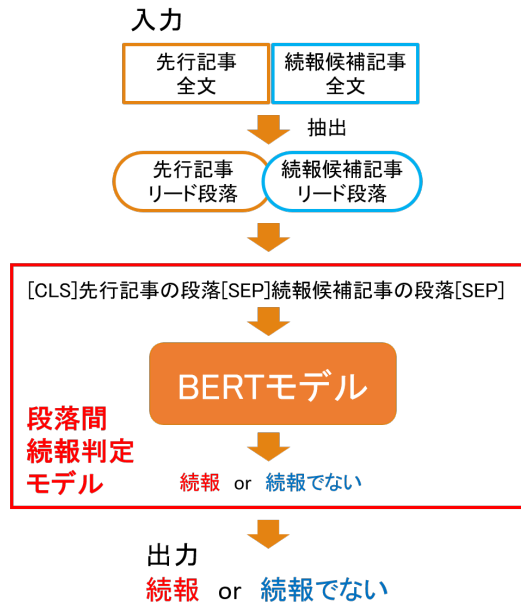


図 1 記事間続報判定モデル

2 関連研究

新聞記事などのデジタルアーカイブを積極的に利用しようという取り組みには次のようなものがある。Diego らは、新聞記事の中でも特に切り抜きに特化したデジタルライブラリについて検索を安易にするためのネットワークの構築や、デジタル化、索引付けについて提案した [1]。Ramesh らは、図書館で新聞記事の切り抜きを製本して保存していたものをデジタル化することで、図書館の運営や研究者に貢献できたことを実施に行った取り組みを例に挙げて示した [4]。ニュース記事をデジタル化しアーカイブとして運用することは図書館の省スペース化や検索効率の向上などにおいて注目されている分野である。

BERT を用いた新聞記事の解析には次のようなものがある。Liu らは、中国語の記事に対してトピック予測モデルの向上を図るべく、BERT と LSTM を組み合わせた新しいトピック分析モデルについて提案した [7]。Sanchez らは、法律の専門的なニュース記事検索に対して、BERT を用いて文脈情報を考慮した推薦を行うことで、類似した判例の記事を推薦することを提案した [9]。Deshmukh らは、ニュース記事検索において、任意のクエリによる BM25 でランキングされた記事集合に対して、ユーザが直近に閲覧した記事との類似を Sentence-BERT を用いて判断し類似記事を推薦することを提案した [2]。Glazkova らは、フェイクニュースの発見のために BERT を分類器として用いた。また、Wikipedia による事前学習モデル以外にも過去のフェイクニュースを事前学習に用いることを提案した [5]。Zhang らは、ニュース記事検索において個別のユーザの特徴をユーザが閲覧した記事本文で直接表現することを提案し、クエリによる単語レベルとユーザレベルのそれぞれで BERT の学習を行い、2つのモデルを組み合わせることで記事の推薦を行った [10]。アーカイブ内の記事を検索する時、既存の単語レ

ベルでの記事検索のみでは不十分な場合がある。記事アーカイブを用いて長期間報道されたトピックについて検索する場合、トピックの話の流れをつかむためにユーザは語の共通性以外に記事の文脈や記事同士の関係に着目している。関連研究では記事間の類似度などに着目しているが、本研究では記事間の続報関係に着目する。ユーザが閲覧した記事やブックマークしている記事に対して、話の続きとなる続報記事を推薦することでスムーズな検索とトピックの容易な理解ができると考えた。

また、BERT のファインチューニングのためには大量の学習データセットが必要となる。本研究では、続報記事のペア、非続報記事のペアが大量に必要である。訓練データを人手で作ることが高いコストを要する場合、学習させたい要素に類似した要素を自動的に抽出し訓練データとして扱うことがある。このような訓練データを関連研究では疑似訓練データと呼んでいる。ニューラルネットワークの学習における疑似訓練データについては、次のような研究がされている。Liu らは、ゼロ代名詞の係り受け解析のためにニューラルネットワークでのモデルを提案し、モデルを訓練するためにニュース記事などから疑似訓練データを生成することを提案した [8]。Li らは、英語のスペルミス訂正のための nested RNN モデルを提案し、単語の音素の類似に着目してモデルを学習するための疑似訓練データを作成することを提案した [6]。水野らは、系列データの収集や前処理などは作成コストが高く、大規模なデータセットの収集、利用は困難であるため、深層学習で十分な学習データを確保するために LSTM AutoEncoder を用いて系列データから固定長のベクトルを抽出して疑似訓練データを作成することを提案した [14]。

3 記事間続報判定モデル

本節では、2つの記事ペアを入力した時に、記事ペアが続報関係にあるかどうかを判断する記事間続報判定モデルを提案する。続報判定は BERT を用いたモデルで行うが、記事ペア全文は入力長としては長すぎる。そこで、入力を記事ペアからそれぞれ 1つずつ選んだ段落ペアとし、続報判定を行う段落間続報判定モデルを提案する。また、BERT のファインチューニングを行うには大量の続報記事ペアが必要となる。人手でデータセットを作成することはコストが高いため、記事要素間の関係に着目して続報関係を学習できる疑似訓練データの作成を提案する。

3.1 続報判定における文脈情報

ユーザが2つの記事同士を比較して続報を発見する時、2つの記事間の語の類似性だけではなく文脈にも注目している。図 2 は使われている語は共通しているが続報関係を持たない記事ペアの記事内容の一部を示したものであり、図 3 は続報関係を持つ記事ペアの記事内容の一部を示したものである。図 2 の先行記事は、阪神・淡路大震災発生時に京都府の救援隊が派遣されるという報道である。続報候補記事は新潟県中越地震の時に兵庫県を含めた近畿の警察が救援隊を派遣するという報道であ

発行日：1995/01/18

タイトル：丹後大震災の恩返し救援隊/弥栄町が派遣/阪神・淡路大震災
本文：...十七日明け方に発生した兵庫県南部地震被害の救援のため、京都府竹野郡弥栄町は救援隊(隊長・上田俊六町消防団長、十五人)の派遣を決め、同日夕、一行が被害の最もひどい神戸市に向かって出発した。...



続報ではない

発行日：2004/10/29

タイトル：新潟県中越地震 近畿5府県警計200人派遣へ

本文：近畿5府県警計200人派遣へ 新潟県中越地震で、兵庫、京都、大阪、奈良、和歌山各府県警は二十八日、近畿管区広域緊急援助隊として、計約二百人を二十九日から被災地に派遣することを決めた。...九五年の震災では、新潟県警から延べ約五千八百人が兵庫県内の被災地に派遣されたといひ、兵庫県警は「震災で受けた支援を、今回の救援活動で少しでもお返ししたい」としている。

図2 使われている語は共通しているが続報関係を持たない記事ペア

発行日：2009/08/14

タイトル：県西・北部豪雨 JR姫新線・播磨新宮-佐用間「復旧には1週間程度」 佐用-上月はめど立たず

本文：JR西日本は13日、豪雨被害で不通になっている姫新線の播磨新宮-美作江見間のうち、播磨新宮-佐用間が1週間程度で復旧する見通しになったと発表した。運転再開日は未定。残る佐用-美作江見間は被害状況を調査中だが、特に佐用-上月間の被害がひどく、復旧のめどは立たない。...



続報

発行日：2009/08/20

タイトル：県西・北部豪雨 姫新線あす一部再開 播磨新宮-佐用間 全面復旧は10月上旬

本文：兵庫県西、北部の豪雨で、JR西日本は19日、不通になっている姫新線の播磨新宮-美作江見間のうち、播磨新宮-佐用間の運行を21日の始発から再開すると発表した。...「復旧には半年かかる」との厳しい見方もあったが、その後の調査で、今後1カ月半程度で全線復旧できることが分かった。...

図3 続報関係を持つ記事ペア

る。続報候補記事では派遣理由について、阪神・淡路大震災時に新潟県警の支援を受けたと述べられている。この2つの記事は、地域の違い以外は使われている語は共通している記事である。しかし、先行記事は京都府の救援隊の救援を報じた記事であり、続報候補記事にある新潟県の支援については報じていない。したがって、この続報候補記事は先行記事の話の続きである続報としてみなすことは適切ではない。また、図3の先行記事は、豪雨被害によって電車の全線復旧にめどが立たないと報じた記事である。続報候補記事では、その後の調査で電車の全線復旧は1か月以上必要だと報じた記事である。この2つの記事は、語が表層的に共通しており内容的にも続報関係にあると判断できる。

従って、記事間の続報の判定には、語の共通性や類似性だけではなく文脈的な要素も考慮したモデルが必要である。本稿では、文章中の語の類似や文脈情報を学習できるBERTを用いて図1に示すような記事間続報判定モデルを作成する。記事間続報判定モデルの入力は先行記事と続報候補記事の全文であり、出力は続報か否かとなる。

モデル作成時の課題として、BERTの入力長に対して記事全文のペアは長すぎる。また、通常の2値ラベル付けタスクによるファインチューニングを行うと文脈よりも語の類似が大きく考慮されるモデルとなる。そこで3.2節では、入力を記事ペアのそれぞれから1つずつ抽出した段落ペアとし出力を続報か否かとする、続報判定において文脈を重視する段落間続報判定モ

デルの構築を提案する。加えて、BERTのファインチューニングにおいて大量の続報記事ペアが必要となるが、人手でのデータセット作成には多大な労力を必要とする。そこで3.3.1節では、同一記事間のリード段落と後述段落を疑似的な続報関係とみなして、疑似訓練データを自動生成することを提案する。また、検証データやテストデータは続報記事ペアを人手で少量作成する。3.3.2節では偏りのない検証データ、テストデータ作成のための方法を示す。3.4節では、段落間続報判定モデルの学習結果とテスト結果を示す。3.5節では、記事間続報判定モデルの入力である記事ペアそれぞれの全文から段落間続報判定モデルの入力である段落ペアを選ぶ方法を示し、記事間続報判定モデルのテスト結果を示す。

3.2 段落間続報判定モデルの構築

記事間続報判定モデルの作成のために、2つの記事からそれぞれ選んだ段落ペアを入力とし、続報かどうかを判断するモデルを構築する。図1に示した通り、記事間続報判定モデルの入力は記事ペアの全文となる。しかし、記事全文が続報判定に必要なわけではない。ニュース記事などの報道文書において、1つの段落に対して1つの出来事について述べられていることが多い。また、BERTの入力長に対し、記事ペアそれぞれの全文は入力として長すぎる。そこで、記事ペアからそれぞれ選んだ段落ペアに対して、続報か続報でないかを判定するモデルが必要である。

また、入力段落ペアにおいて語の類似よりも文脈について重視した、段落間続報判定モデルを構築する。続報の有無を判定するモデルを構築するので、事前学習されたBERTに対して2値分類タスクを課してファインチューニングを行う。しかし、続報記事ペアが続報と判断できる段落ペアを正例、ランダムに記事を組み合わせて非続報記事ペアを作成したリード段落ペアを負例とすると、語の類似性が強く考慮されたモデルになってしまう。そのため、次に示す方法でファインチューニングを行うことで、語の類似性よりも文脈的な要素から続報を判断する段落間続報判定モデルを構築する。

BERTのファインチューニングの方法として2種類を検討する。1つ目は、先行記事と続報記事の段落を入れ替えた類似負例を含むデータセットを使ってファインチューニングを行った通常段落間続報判定モデルである。2つ目は、2つの段落ペアの共通語をMASKトークンに置き換えたデータセットでファインチューニングを行った共通語MASK段落間続報判定モデルである。

ファインチューニングに用いるデータセットの負例は、語が表層的に共通しない負例の他に語が表層的に共通する負例を用意する。続報記事ペアの本文を比較すると、語彙的に類似している。また、語が表層的に共通しているものも多く、続報を判断するときに語の類似性を強く考慮する要因の1つになっていると考えられる。したがって、図4のように、続報候補記事から抜き出した段落、先行記事から抜き出した段落の順にペアとし、語が表層的に共通するが続報ではない負例ペアである類似負例を作成する。この方法で作成したデータセットを用いてファイン

[CLS] ...被害がひどく、復旧のめどは立たない...[SEP]...残る佐用—美作江見間は10月上旬ごろに復旧する見込みという。...[SEP]



[CLS]...残る佐用—美作江見間は10月上旬ごろに復旧する見込みという。...[SEP]...被害がひどく、復旧のめどは立たない...[SEP]

図 4 段落ペアの順番を入れ替えた類似負例

[CLS]...**JR姫新線**佐用駅近くの踏切が流され...[SEP]...**JR姫新線**は線路などの被害が大きく...[SEP]



[CLS]... **[MASK][MASK]**佐用駅近くの踏切が流され...[SEP]... **[MASK][MASK]**は線路などの被害が大きく...[SEP]

図 5 共通する名詞を MASK トークンで置き換え

チューニングを行った BERT モデルを通常段落間続報判定モデルとする。

また、語が表層的に共通することによるモデルの続報判定の影響をより抑えるため、段落ペア中に共通して存在する語を MASK トークンに置き換える。図 5 のように、入力における 2 つの段落において共通する名詞を MASK トークンで置き換えたものをデータセットとする。負例に類似負例を混ぜたデータセットに共通する名詞を MASK トークンで置き換える処理をしたデータセットを用いてファインチューニングを行った BERT モデルを共通語 MASK 段落間続報判定モデルとする。

3.3 データセットの作成

BERT の事前学習モデルをファインチューニングするために、続報となる記事ペアの続報だと判断した段落ペアが大量に必要となる。しかし、人手で続報記事ペアを作成することは多大な労力を要する。

そこで、本研究では同一記事内の段落同士の関係性に着目して、疑似訓練データを自動で作成することを提案する。また、検証データとテストデータについてはトピックに偏りのないようなデータを手動で少量作成する。

3.3.1 疑似訓練データ

続報記事ペアを大量に人手で作成することは多大な労力を必要とするため、訓練データを自動で作成する。続報となる 2 つの記事ペアは、続報記事が先行記事の話の続きとなっている。図 3 では、先行記事で全線復旧にめどが立たないと報道して、続報記事で詳しい調査の結果復旧に 1 か月半かかると報道された。同じ電車の路線の被害の程度の話をしており、先行記事の話の続きが続報記事になっていることが分かる。つまり、話の続きとなる段落ペアを自動的に収集できれば、疑似的な続報関係とみなし疑似訓練データが作成できる。

そこで、同一記事内に存在する段落の関係に着目して疑似訓練データを作成する。報道記事において、第 1 段落は記事で報道する内容を簡潔に要約した段落となることが多く、リード段落と呼ばれる。また、リード段落以降の段落は報道内容を詳しく

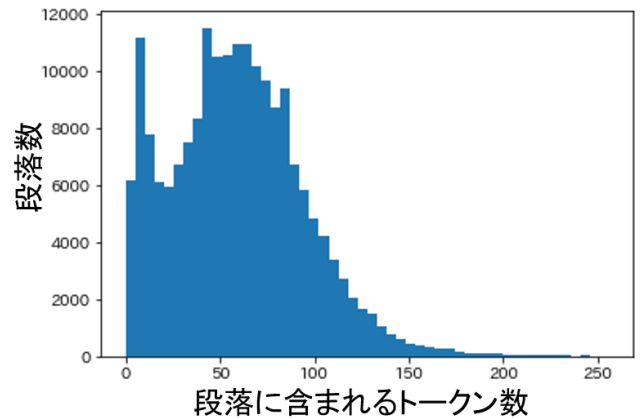


図 6 段落ごとのトークン数のヒストグラム

く説明した段落である。本研究ではリード段落以降の段落を後述段落と呼ぶ。ここで、報道内容を詳しく説明した後述段落は、報道概要を示すリード段落について詳しく説明したものであるととらえることができる。つまり、リード段落の話の続きが後述段落であるといえる。そこで、同一記事内においてリード-後述段落を疑似的な続報関係とみなし、疑似訓練データを作成する。

図 6 は 2019 年に神戸新聞から発行された記事のそれぞれの段落が、BERT の事前学習で用いたトークナイザでトークン化した場合、1 段落につきどれくらいのトークン数があるかを示すものである。横軸がトークン数、縦軸が段落数となっている。図 6 を見るとほとんどの段落が 256 トークン以下で構成されている。また、40 トークン以下の段落は文章として成り立っていないものが多くみられた。よって、40 トークン以上 256 トークン未満の段落に対し疑似訓練データの生成を行った。

後述段落はリード段落以降の段落を指すため複数存在することがある。この時、後述段落は時間的にリード段落以後のものを正例ペアとして扱った。まず、すべての段落の時間情報 [12] を抽出する。リード段落直後の後述段落から順に、リード段落と後述段落の時間情報を比較し、時間情報がリード段落以後の後述段落をリード-後述段落ペアとした。本研究では 1 つの記事につき、最初にリード-後述段落ペアとなった 1 ペアのみを抽出した。

疑似訓練データの作成は、神戸新聞で 2019 年に発行された新聞記事約 2 万件を用いた。まずは、記事集合に対し、記事の第 1 段落をリード段落としてみなし、40 トークン未満 256 トークン以上となる記事を自動生成する候補から除外した。次に、記事の第 2 段落以降の段落を後述段落とし、40 トークン以上 256 トークン未満であるかを確認する。第 2 段落から順に確認していき、トークン数と時間情報の条件を満たした最初の後述段落とリード段落のペアを、リード-後述段落ペアとして正例とする。40 トークン以上 256 トークン未満となる後述段落がなかった場合、その記事は自動生成する候補から除外する。こうして作成されたリード-後述段落の正例ペアは約 1.7 万件となる。

最後に 3.2 節で示したリード段落と後述段落を入れ替えた類

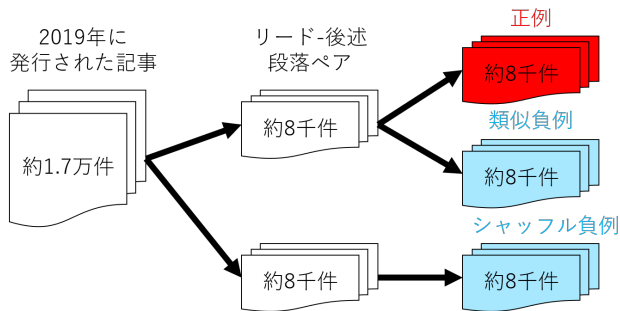


図 7 疑似訓練データの内訳

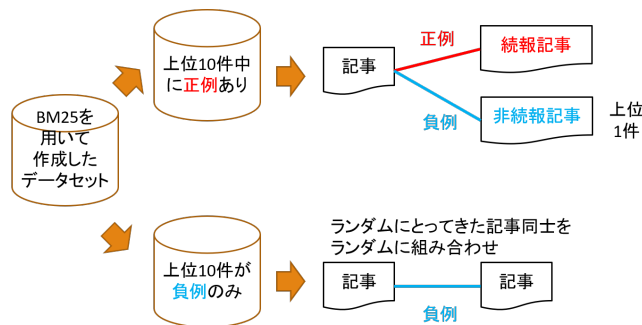


図 9 検証データの作成

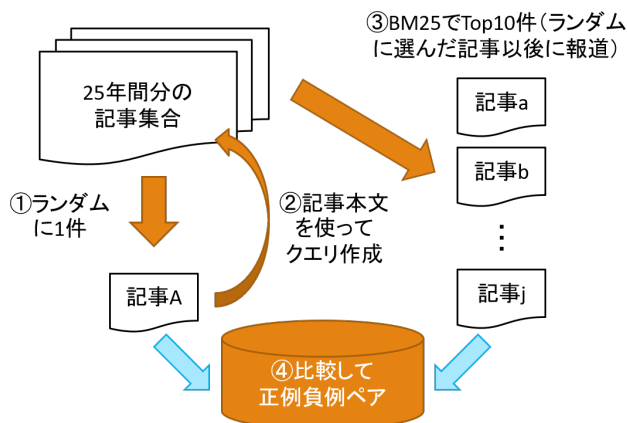


図 8 トピックに偏りのない続報記事ペアの作成

似負例と、リード-後述段落ペアの後述段落部分をランダムに入れ替えた語が表層的に共通していない負例を作成する。図 7 のように、先ほど作成した約 1.7 万件のリード-後述段落ペアを半数に分ける。図上部の半数に分けたリード-後述段落ペアはそのまま正例となる。正例ペアの 2 つの段落の順序を入れ替えて後述-リード段落とし、語が表層的に共通した負例とする。また、正例に使われなかったリード-後述段落は、後述段落をランダムに入れ替えて語が表層的に共通していない負例とし、シャッフル負例と呼ぶ。2019 年に発行された記事から生成された疑似訓練データは正例、語が表層的に共通した負例、語が表層的に共通しない負例の割合がほぼ同数となる約 2.6 万件のデータセットとなった。

3.3.2 検証データ、テストデータ

検証データとテストデータはモデルが実データで損失値が最小となる点で学習を止め、その結果が確認ができるように、人手で少量作成する必要がある。この時、記事のトピックなどに偏りが出ないように図 8 のような方法でデータセットを作成する。

まずは、記事集合からランダムに 1 件記事を抽出し、先行記事とする。本研究では 1995 年から 2019 年までの 24 年間分の記事集合を用いた。次に、先行記事の本文を用いて 24 年間分の記事集合に対して検索をかけるためのクエリを作成する。クエリは記事本文の中から MeCab で形態素解析を行った時、品詞分類が名詞、動詞、形容詞、形容動詞、副詞となる語を原型にしてクエリとした。その後、作成したクエリで記事集合に対

し BM25 でランキング付けを行う。検索には Elasticsearch¹ の BM25 をデフォルト設定で用いた。Elasticsearch での検索時のトークナイザは bigram と trigram である。最後に、検索結果のトップ 10 件を続報記事候補とし、先行記事と比較を行い、正例と負例のペアを作成する。

検証データは、正例、語が表層的に共通する負例、語が表層的に共通しない負例の 3 種類から構成される。図 9 は検証データの作成手順の概要である。まずは、上位 10 件の中に正例ペアが存在する先行記事と続報候補記事のペアにおいて、正例記事ペアは全て正例として検証データに用いる。負例ペアの中で BM25 の検索結果上位 1 件を、語が表層的に共通している負例とする。また、上位 10 件が負例のみの記事ペアの先行記事のみを収集する。そしてそれらをランダムに組み合わせ、正例と同数のペアを語が表層的に共通しない負例とする。本実験では、BM25 を用いて作成したデータセットは 1170 件であり、最終的に正例となったものが 45 ペア、語が表層的に共通する負例となったものが 26 ペア、語が表層的に共通しない負例となったものが 26 ペアとなり、合計で 97 ペアとなった。

テストデータは、図 8 で比較したデータから正例を一定数選び、正例と同数の負例をランダムに選ぶ。本研究では、テストデータは 76 件であり、正例が 38 件、負例が 38 件であった。

3.4 段落間続報判定モデルの学習結果

3.3 で示したデータを用いて、段落間続報判定モデルの学習を行う。BERT の事前学習モデルは東北大学の乾研究室が公開しているモデル²を使用した。疑似訓練データで学習を行った結果を、通常段落間続報判定モデルは図 10 に、共通 MASK 段落間続報判定モデルは図 11 に示す。通常段落間続報判定モデルは、学習率は $2e-9$ 、バッチサイズは 4 で学習を行った。21epoch で早期終了をかけ 31epoch まで学習を進めた。共通語 MASK 段落間続報判定モデルは、学習率は $2e-9$ 、バッチサイズは 4 で学習を行った。21epoch で早期終了をかけ 31epoch まで学習を進めた。

3.5 記事間続報判定モデルの入出力とその結果

記事間続報判定モデルの入力は図 1 の通り先行記事全文と続

1: <https://www.elastic.co/jp/>

2: <https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>

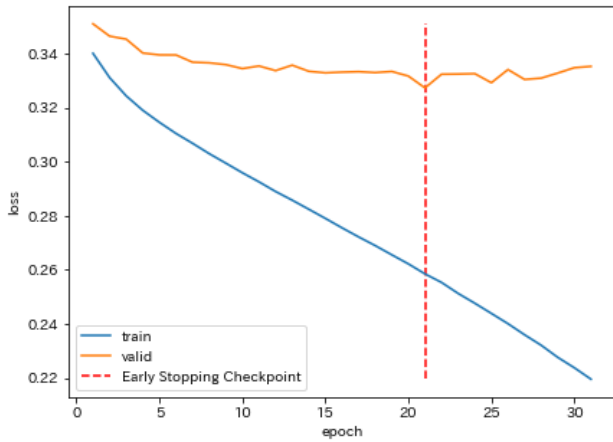


図 10 通常段落間続報判定モデルの学習結果

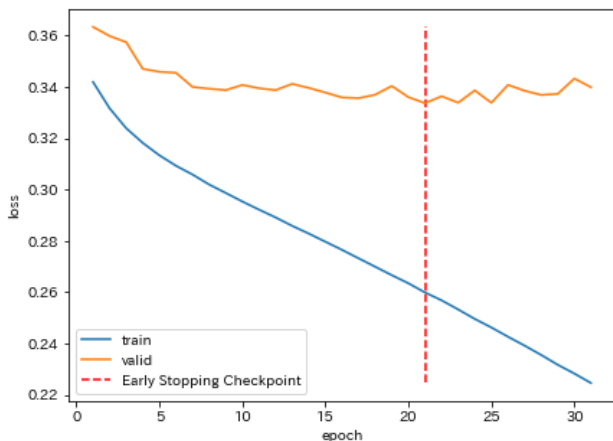


図 11 共通語 MASK 段落間続報判定モデルの学習結果

表 1 通常記事間続報判定モデルのテスト結果

		予測ラベル	
		続報	続報でない
正解ラベル	続報	25	13
	続報でない	16	22

報候補記事全文である。しかし、記事ペアの全文では入力長として長いので段落間続報判定モデルの入力は記事ペアからそれぞれ選んだ段落ペアとした。そのため、記事全体を表した段落をそれぞれの記事から抽出し、段落間続報判定モデルの入力として用いることができれば、段落間続報判定モデルの出力結果をそのまま記事間続報判定モデルの出力として扱うことができる。記事全体を表す段落として、それぞれの記事のリード段落を続報重視モデルの入力とした。それぞれの記事において第 1 段落をリード段落として扱った。

通常段落間続報判定モデルを使った記事間続報判定モデルを通常記事間続報判定モデル、共通語 MASK 段落間続報判定モデルを使った記事間続報判定モデルを共通語 MASK 記事間続報判定モデルとする。表 1 は、通常記事間続報判定モデルのテスト結果を示したものである。正解率は約 0.62 であった。表 2 は、共通語 MASK 記事間続報判定モデルのテスト結果を示したものである。正解率は約 0.54 であった。

表 2 共通語 MASK 記事間続報判定モデルのテスト結果

		予測ラベル	
		続報	続報でない
正解ラベル	続報	20	18
	続報でない	17	21

4 評価

作成した 2 種類の記事間続報判定モデルに対して、次に示すようなユーザによる評価を行う。

- 記事が与えられたとき、その記事の続報記事と先行記事をそれぞれ推薦し従来の推薦手法と比較してどの程度優れているかの評価
- 以前提案した記事アーカイブ検索インタフェース [12] [13] に記事間続報判定モデルを組み込み、ユーザに使用感について尋ねる評価

4.1 記事推薦の評価

記事間続報判定モデルを用いた記事推薦は次のように行う。まず、与えられた記事の本文中の自立語をクエリとして、記事の発行日前、発行日後の記事から BM25 を用いてそれぞれ上位 100 件の記事集合を得る。与えられた記事の発行日前の検索によって得られた記事集合を先行記事の推薦に、発行日後の検索によって得られた記事集合を続報記事の推薦に用いる。次に、与えられた記事と検索で得られた記事集合に対して続報関係があるかを記事間続報判定モデルを用いて判定し、与えられた記事に対する続報記事または先行記事を抽出する。最後に、抽出した記事集合の上位 5 件を推薦記事として扱う。抽出した記事集合のランキング順は BM25 で検索した時のスコア順である。

比較する従来の記事推薦方法として BM25 を用いた。与えられた記事の本文から自立語をクエリとして検索し、上位 5 件を推薦記事とした。

推薦に用いる記事は次の 5 つのトピックから選ぶ。それぞれのトピックから先行記事の推薦のために 5 件ずつ、続報記事の推薦のために 5 件ずつ記事を選ぶ。

- 阪神・淡路大震災
- 「もんじゅ」のナトリウム漏洩事故
- 消費税が 3% から 5% に引き上げ
- 明石海峡大橋開通
- 三宅島の噴火と全島避難

先行記事の推薦、続報記事の推薦のために用意されたそれぞれ 25 件の記事に対し、BM25、通常記事間続報判定モデル、共通語 MASK 記事間続報判定モデルを用いた 3 つの推薦方法で記事をそれぞれ推薦する。推薦された記事集合をユーザに見せて、続報である、続報ではないが関連がある、関連がないの 3 段階でスコア付けさせる。ユーザ 5 名に評価を行ってもらい、多数決によって推薦記事の最終的なスコアとする。ユーザの評価が 2 項目に 2 票、1 項目に 1 票と散らばった場合は、関連する記事に少なくとも 3 票入っているの、関連する記事である 1 のラベルを与える。3 段階のスコアを nDCG で評価して各推薦手

表 3 先行記事推薦における有意差の有無

比較モデル (平均スコア大)	比較モデル 2(平均スコア小)	p 値	有意差あり
共通語 MASK 記事間 続報判定モデル	通常記事間 続報判定モデル	0.469	
共通語 MASK 記事間 続報判定モデル	BM25	0.537	
通常記事間 続報判定モデル	BM25	0.987	

表 4 続報記事推薦における有意差の有無

比較モデル 1(平均スコア大)	比較モデル 2(平均スコア小)	p 値	有意差あり
通常記事間 続報判定モデル	共通語 MASK 記事間 続報判定モデル	0.254	
共通語 MASK 記事間 続報判定モデル	BM25	0.055	
通常記事間 続報判定モデル	BM25	0.013	○

法の評価を比較する。

nDCG による各手法の評価結果を述べる。先行記事推薦において nDCG による評価の平均値は、通常記事間続報判定モデルの平均スコアが 0.80、共通語 MASK 記事間続報判定モデルの平均スコアが 0.82、BM25 の平均スコアが 0.80 となった。続報記事推薦において、ラベルが全て 0 になった記事推薦があったため最終的な評価結果が 24 件となった。続報記事推薦において nDCG による評価の平均値は、通常記事間続報判定モデルの平均スコアが 0.84、共通語 MASK 記事間続報判定モデルの平均スコアが 0.80、BM25 の平均スコアが 0.74 となった。

3 つ推薦手法の平均値が有意性を持つのかを t 検定を用いた多重比較法によって確認する。ここで、多重比較法は複数回 t 検定を行うので最初に設定した有意水準が最終的な有意水準を超えることが考えられる。従って、有意水準を両側 5% としボンフェローニ補正により補正した。補正後の有意水準は共に $0.05/3 = 0.0167$ である。それぞれの検定結果は表 3 と表 4 に示す。続報記事の推薦において、通常記事間続報判定モデルと BM25 の nDCG の平均スコアに有意差があるという結果になり、通常記事間続報判定モデルの nDCG の平均スコアは BM25 の nDCG の平均スコアを上回っている。従って、続報記事の推薦において、通常記事間続報判定モデルは BM25 を用いた推薦よりも続報記事を推薦できていることが分かった。

4.2 インタフェースの使用感による評価

我々は以前、探索者ビューを用いた記事アーカイブ検索インタフェースを提案した [12] [13]。探索者ビューはブックマーク記事や閲覧記事など、ユーザが必要としている記事間の続報関係をグラフ化したものである。探索者ビューを用いることで、ユーザは記事間の続報関係を確認しながら検索を行うことができ、スムーズな検索と効率的なまとめ作成ができると考える。そこで、提案したインタフェースに通常記事間続報判定モデルを組み込み、ユーザに使用感を尋ねる実験を行うことでモデルと提案インタフェースの有用性を示す。

探索者ビューを用いて記事アーカイブでの検索やまとめを行うと効率的に検索とまとめが行えるはずである。そこで、指示したトピックを一定時間内で検索しまとめを作成させるタスク

表 5 実験のグループ分け

グループ	前半タスク	後半タスク
A	佐用トピック 探索者ビューなし	姫路トピック 探索者ビューあり
B	姫路トピック 探索者ビューなし	佐用トピック 探索者ビューあり
C	佐用トピック 探索者ビューあり	姫路トピック 探索者ビューなし
D	姫路トピック 探索者ビューあり	佐用トピック 探索者ビューなし

表 6 インタラクティブシステムを評価する質問紙 [11] より必要部のみを抜粋した質問項目

質問	中心概念	大分類
1 目標達成のために必要かつ十分な機能がある	機能の有効性	役立ち度
2 仕事や生活を向上させる	効果・効率の実感	
3 私の力になってくれる	エンパワー	
4 賢く応答する	知性	
5 使い方の自由度が高い	柔軟性	
6 困ったときに頼りになる	信頼性・レジリエンス	
7 使うことに夢中になれる	没入感	共感度
8 良い意味での驚きを与える	インパクト	
9 使いこなしていく楽しさを感じさせる	期待感	

を、提案インタフェースの有無で分けてユーザに行わせる。タスク後にインタフェースの使用感を尋ね、提案インタフェースの有無で使用感に差があったかを調べる。ユーザは次の 2 つのトピックについて検索とまとめを行う。

- 2009 年 10 月から姫路城の平成の大修理が行われた。この修理について、始まるまでに指摘されていた問題点や開催されていたイベントについてまとめよ。

- 2009 年 8 月に兵庫県佐用町で大規模な豪雨災害が発生した。この災害に対する支援活動についてまとめよ。

トピック、探索者ビューの使用の有無に関する組み合わせは表 5 に示す。各タスク終了後に表 6 に示す項目についてアンケートを行い、各ユーザの探索者ビューの有無によって評価がどう変化したかを比較する。

実験に参加したユーザは、同じ大学に所属する大学生 4 名、大学院生 4 名である。8 名のユーザは us01 から us08 までナンバリングした。8 名はそれぞれ表 5 において、us01 と us02 はグループ A、us03 と us04 はグループ B、us05 と us06 はグループ C、us07 と us08 はグループ D として実験を行った。2 つのインタフェースの使用感を尋ねたアンケートの回答結果において、各質問項目で平均したものを示したグラフを図 12 に示す。1 を「そう思わない」で 5 を「そう思う」とした時の評価尺度を縦軸に、表 6 における質問番号を横軸としている。

2 つのインタフェースの各項目において平均値に有意差があるのかを確認するため、t 検定を有意水準両側 5%で行った。複数回の検定となるので有意水準をボンフェローニ補正によって補正し $0.05/9 = 0.0056$ となった。表 7 は各項目の p 値と有意差があるかについて示したものである。「目標達成のために十分な機能がある」の項目において平均値に有意差があると確認できた。また評価値の平均において、探索者ビューを用いたインタフェースを用いた方が探索者ビューを用いないインタフェー

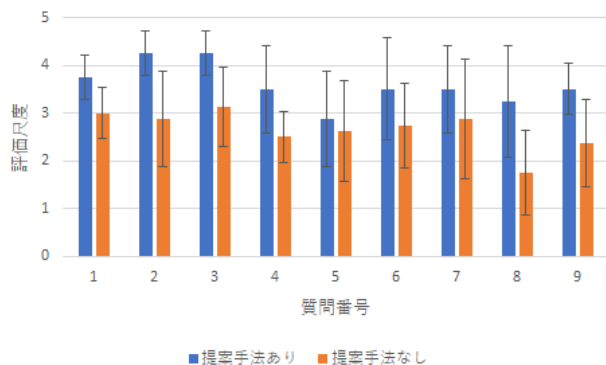


図 12 本実験における各インタフェースの使用感
(エラーバー：標準偏差)

表 7 質問項目ごとの有意差の有無

質問番号	質問内容	p 値	有意差あり
1	目標達成のために十分な機能がある	0.003	○
2	仕事や生活を向上させる	0.108	
3	私の力になってくれる	0.108	
4	賢く応答する	0.052	
5	自由度が高い	0.178	
6	困ったときの頼りになる	0.070	
7	使うことに夢中になれる	0.070	
8	良い意味での驚きを与える	0.034	
9	使いこなしていく楽しさを感じさせる	0.070	

スを上回っている。従って、探索者ビューを用いたインタフェースの方が探索者ビューを用いないインタフェースに比べて、機能の有効性において優れていることが分かった。

5 ま と め

本稿では、記事間の続報関係の有無を判定する記事間続報判定モデルを提案した。記事間続報判定モデル作成のために、記事ペアからそれぞれ1つずつ選択した段落ペアを入力とし、続報か否かが出力となる段落間続報判定モデルをBERTを用いて構築した。また、続報記事ペアの訓練データセットを手で作成することはコストが高いため、同一記事内のリード段落、後続段落の関係性に注目して、訓練データを自動で作成する疑似訓練データを提案した。作成した記事間続報判定モデルをユーザによって評価した。既存手法と記事間続報判定モデルによる続報記事推薦を評価して、どちらがよりよい推薦を行えたかを比較した結果、通常記事間続報判定モデルのほうがBM25よりも続報記事をうまく推薦できていることが分かった。また、記事間続報判定モデルを組み込んだ記事アーカイブ検索インタフェースと既存の検索インタフェースのそれぞれでまとめを伴う検索をユーザに行わせ、使用感を尋ねて比較した。その結果、探索者ビューを用いたインタフェースの方が探索者ビューを用いないインタフェースに比べて、機能の有効性において優れていることが分かり提案モデルとインタフェースの有効性が確認できた。

今後は、BERTの学習を工夫し記事間続報判定モデルの正解

率を向上させることを目指す。

謝 辞

本研究はJSPS 科研費 JP21H03775, JP21H03774 の助成を受けたものです。ここに記して謝意を表します。

文 献

- [1] Diego Calvanese, Tiziana Catarci, and Giuseppe Santucci. Laurin: a distributed digital library of newspaper clippings. *World wide web*, 2001.
- [2] Anup Anand Deshmukh and Udhav Sethi. IR-BERT: Leveraging BERT for Semantic Search in Background Linking for News Articles. *arXiv preprint arXiv:2007.12603*, 2020.
- [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] Ramesh C Gaur, Mahesh Chand, Kavita Gaur, and Amar Singh Yadav. Online newspaper clippings & news services for libraries: Experiences in Indian Libraries. *IFLA WLIC 2013 - Singapore - Future Libraries: Infinite Possibilities*, 2013.
- [5] Anna Glazkova, Maksim Glazkov, and Timofey Trifonov. g2tmn at Constraint@ AAI2021: Exploiting CT-BERT and Ensembling Learning for COVID-19 Fake News Detection. In *International Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation*, pp. 116–127, 2021.
- [6] Hao Li, Yang Wang, Xinyu Liu, Zhichao Sheng, and Si Wei. Spelling error correction using a nested rnn model and pseudo training data. *arXiv preprint arXiv:1811.00238*, 2018.
- [7] Jingang Liu, Chunhe Xia, Xiaojian Li, Haihua Yan, and Tengting Liu. A BERT-based Ensemble Model for Chinese News Topic Prediction. In *Proceedings of the 2020 2nd International Conference on Big Data Engineering*, pp. 18–23, 2020.
- [8] Ting Liu, Yiming Cui, Qingyu Yin, Weinan Zhang, Shijin Wang, and Guoping Hu. Generating and exploiting large-scale pseudo training data for zero pronoun resolution. *arXiv preprint arXiv:1606.01603*, 2016.
- [9] Luis Sanchez, Jiyin He, Jarana Manotumrukha, Dyaa Albakour, Miguel Martinez, and Aldo Lipani. Easing Legal News Monitoring with Learning to Rank and BERT. *Advances in Information Retrieval*, 2020.
- [10] Qi Zhang, Jingjie Li, Qinglin Jia, Chuyuan Wang, Jieming Zhu, Zhaowei Wang, and Xiuqiang He. Unbert: User-news matching bert for news recommendation. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 3356–3362, 2021.
- [11] 松本啓太, 善方日出夫. ユーザー体験 (UX) に基づいてインタラクティブシステムを評価する質問紙. *人間工学*, Vol. 53, pp. 46–50, 2017.
- [12] 松本直彰, 湯本高行, 山本岳洋, 大島裕明. ニュースアーカイブ探索のための記事間の関係抽出とその可視化. 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, F25-3, 2021.
- [13] 松本直彰, 湯本高行, 山本岳洋, 大島裕明. 記事の言及関係に基づく探索者ビューを用いたニュースアーカイブ検索. *電子情報通信学会技術研究報告*, pp. 13–18, 2021.
- [14] 水野充大, 岡留剛. 系列データにおける特徴空間上での疑似訓練データ生成. 2020年度情報処理学会関西支部 支部大会 講演論文集, 1884-197X, 2020.