

特定選手に注目した試合結果記事見出しのパーソナライズ

佐々木泰河† 工家 昂之†† 山本 岳洋†† 莊司 慶行†††

† 兵庫県立大学 社会情報科学部 〒 651-2197 兵庫県神戸市西区学園西町 8-2-1

†† 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

††† 静岡大学 情報学部 〒 432-8011 静岡県浜松市中区城北 3-5-1

E-mail: †fa20e036@guh.u-hyogo.ac.jp, ††t.yamamoto@sis.u-hyogo.ac.jp, ad22l024@guh.u-hyogo.ac.jp,

†††shojiy@inf.shizuoka.ac.jp

あらまし 本研究では、試合結果記事における特定選手に注目した見出しを生成する手法を提案する。見出しの生成には T5 という文生成モデルを用いる。記事全文を入力として、T5 に見出しの最初の文字を注目している選手名にする制約をかけた単純な手法では、注目していない選手の活躍を、注目している選手の活躍のように見出し生成してしまうという問題がある。そこで我々は、試合結果記事から記事の主題に関する文、注目している選手の活躍に関する文のみを抽出して、T5 に入力する手法を提案する。このアプローチにより、指定した選手に注目した自然で事実に基づいた見出しを生成させる。評価実験により、提案手法は、記事全文を入力する単純な手法よりも、特定選手に注目した見出し生成の精度が良いことが示された。

キーワード 見出し生成, パーソナライズ, T5

阪神は打線が爆発し、2桁11得点の猛攻で連敗を「5」でストップさせた。0・5ゲーム差で首位だったDeNAが敗れたため、1試合で首位に返り咲いた。0-0の三回にビッグイニングを作った。先頭・梅野が中前打を打ち、その後1死二、三塁と好機を作ると、近本の右権飛で先取点をゲット。さらに中野&前川の連続適時打で3点を奪った。なおも2死一塁で、大山が左越え9号2ランをマーク。この回に一挙5点を挙げた。2点差に迫られた五回には、先頭・前川が中越え三塁打をマーク。その後、1死満塁の絶好機で梅野の平凡な打球を三塁・石川がファンブルして、その間に1点を追加。続く木浪が中前2点適時打を打ち、再びリードを広げた。終盤にも着実に加点し、前川がプロ初の猛打賞をマークした。先発の西勇は5回7安打3失点で今季5勝目を挙げた。

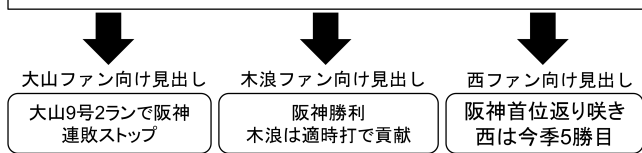


図1 目標とする生成例。

1 はじめに

プロ野球では、試合が行われるとその試合結果を報じる記事が各新聞社から配信される。その記事には試合の勝敗や試合内容、活躍した選手に関する情報などが記される。この記事をクリックするかという判断基準として、見出しに自分の興味を引く内容が含まれているかが考えられる。例えば、興味のある選手の選手名が見出しに含まれていると、読者はその記事が目にとまり、クリックする可能性が高い。具体的には、「阪神が打線爆発で連敗ストップ! 1試合で首位返り咲き 高卒2年目の前川がプロ初猛打賞で牽引 中日戦6連勝」という見出しの記事があるとき、前川選手のファンは記事をクリックする可能性が高い。しかし、見出しに自分の好きな選手の選手名が含まれていなかったとしても、記事の中身を読んでみるとその選手に関する情報が書かれていることもある。例えば、上記の例で示した記事本文には大山選手が9号2ランを放ったことが書かれている。この場合、読者が仮に大山選手のファンで見出しに「大山」

と含まれていない理由により、記事をクリックしなければ、この情報を見逃すことになる。

そこで本研究では、試合結果記事から指定した選手に注目した見出しの生成を行う。本研究が目標とする生成例を図1に示す。ある試合結果記事があったときに仮に大山選手に注目したときには「大山9号2ランで阪神連敗ストップ」と生成する。木浪選手に注目したときには「阪神勝利 木浪は適時打で貢献」と生成する。また、西選手に注目したときには「阪神首位返り咲き 西は今季5勝目」と生成する。このような見出し生成をすることができれば、記事を表示する際に読者が興味を引きやすい見出しに書き換えて表示することも可能になる。

この技術があることにより、読者が自分の興味のある情報を見逃しにくくなる。また、記事の配信者がより多くのクリック数を集めることができる。

試合結果記事見出しの生成において、生成型言語モデルの1つである T5 [9] を用いる。T5 は、文章を入力として与え、解いているタスクに応じた文章を出力するモデルである。大規模なデータで事前学習した T5 を少量のデータでファインチューニングすることで、様々なタスクに対して応用が可能である。本研究では、あるニュース記事本文を入力すると、その見出しを出力するように T5 をファインチューニングする。しかし、見出し生成の際、記事全文を入力すると、生成される見出しにおいて利用すべき情報を誤って選択してしまう問題がある。例えば、大山選手に注目して見出しを生成しようとしたとする。そして本文には「なおも2死一塁で、大山が左越え9号2ランをマーク。」という記述や「木浪が中前2点適時打を打ち、再びリードを広げた。」という記述がある場合を考える。このとき、仮に記事全文を入力したとすると、大山選手が適時打を放っていないかとしても「大山の適時打で阪神連敗ストップ」というような見出し生成をしてしまう可能性がある。大山選手に注

目した見出しを生成するために、木浪選手が適時打を放ったことは余計な情報である。そこで、ニュース記事全文を入力するのではなく、理想的な見出しを生成するために必要な情報を含む文のみを抽出して、入力する。これにより、モデルが目立っていない選手について言及している本文の情報を、見出し生成に利用することを防ぎ、見出しの精度を向上させることができる。

評価実験では、生成した見出しが本研究で目標とする見出しであるか評価する。

本論文の構成について説明する。第2節では、本研究と関連する研究を紹介する。第3節では、本論文で提案する手法について説明し、第4節では提案した手法の評価実験について述べる。第5節では実験で得られた結果をもとに提案手法による見出し生成の精度の変化について考察する。第6節ではまとめと今後の展望について述べる。

2 関連研究

2.1 大規模言語モデルを用いた文生成

文生成に関する研究は、盛んに行われている。過去にはRNN [2] や LSTM [3] による文生成が主流であったが、これらの手法は長い文章や長距離の依存関係をとらえることが難しいという課題がある。近年、このような課題に対処するためにTransformer [12] が提案され、その登場により自然言語処理の様々なタスクにおいて革新的な進展が見られた。Transformerは2017年にGoogleによって発表され、Attentionという機構を導入したモデルである。このTransformerをベースとして多くの大規模言語モデルが提案されており、その代表的な例にはGPT-n [8], BERT [1], T5がある。

江原らは、GPT-2のファインチューニングを工夫することにより主題語に応じたニュース記事見出しを生成する手法を提案している [15]。この研究で提案されている手法は、GPT-2が持つ「文の続きを推定する」という機能を利用している。黒木らはBERT2BERTを用いた広告文生成において特定のキーワードの出現率を高くする手法を提案している [17]。BERT2BERTとは事前学習済みBERTをテキスト生成タスクに応用したモデルである。福田らは文章の感情要約文をT5によって生成する手法を提案している [19]。この研究では、学習時に入力とする文の接頭辞に感情タグをつけることにより精度を向上させた。

本研究では、見出しを生成するための大規模言語モデルとして、T5を利用している。

2.2 要約タスク

要約タスクとは、与えられた文書や文章を簡潔にまとめ、主要な情報や要点を取り出す作業である。要約タスクは一般的に生成型要約と抽出型要約に分類される。

生成型要約とは、与えられた文書や文章から新しい表現で要約を生成するタスクである。このタスクでは大規模言語モデルが用いられることが多い。生成型要約は、元の文章には現れない新しい表現や言い回しで要約を生成する点に特徴がある。木村らはエンコードの際に原文の単語単位のベクトルに加え、原

文の文単位のベクトルも導入することで、クエリ指向要約モデルを強化する手法を提案している [20]。またZhangらは、異なる事前学習方法、プロンプト、モデルサイズにわたる10個のモデルで生成された要約文を人手で評価することで、自動要約において重要な要素を分析した [14]。YangらはChatGPTが多様なテキスト要約タスクにおいてどのような特性を持っているかについて分析した [13]。

抽出型要約は、元の文章にすでに存在する文や文の一部を選択して要約を形成する。小川らは、ニューラルネットワークを用いてニュース記事の各文の重要度を計算し、重要度の高い文を要約に利用する文として抽出する手法を提案している [18]。さらにRuanらは、テキストが持つ階層構造の情報を抽出型要約モデルに注入することで、要約の精度を向上させる手法を提案している [10]。また高津らは、文の特徴量とアンケートで得られるプロフィール特徴量に基づいてニュース記事の各文に対するユーザの興味度を推定し、その興味度を用いて文を抽出することで、パーソナライズされた要約を作成する手法を提案している [16]。

生成型要約と抽出型要約は要約の目的や利用ケースによって使い分けられている。生成型要約は元の文章には現れない新しい表現で要約を生成するため、より柔軟で創造的な結果が得られる。また文のつながりも流暢性のあるものになりやすい。しかし、生成される要約が事実と合致していない可能性がある点に留意する必要がある。本研究は、要約タスクの結果として得られた要約文を、T5に入力している。要約タスクを解いて入力文を作る目的は、生成した見出しと事実との不一致の可能性を下げることである。そのため本研究では、記事本文中の重要な文を抽出して入力文を作る抽出型要約を使用する。

3 提案手法

3.1 問題定義

本研究における問題定義は以下のとおりである。

- **入力**：注目する選手名、試合結果記事
- **出力**：指定した選手に注目した見出し

本研究では、注目する選手名と、試合結果記事を入力し、指定した選手に注目した見出しを生成することを目標とする。

また、生成したい見出しを明確にするために、以下の5つの特徴を持つ見出しを理想的な見出しとして定義する。

- **事実と異ならない**：事実と異なった見出しを生成しては読者の誤解を招き良い見出しとはいえない。
- **指定した選手の選手名を含んでいる**：指定した選手の選手名が含まれていなければ、その選手に注目したニュース記事見出しを生成できていないといえない。
- **記事全体の要約を含んでいる**：記事全体の要約も含まれていなければ、その記事で伝えていることが分からない。
- **日本語として自然である**：文法的に正しくなければ、良い見出しとはいえない。
- **見出しとして自然である**：見出しとして扱う情報として妥当でなければ、良い見出しとはいえない。例えば、「中野選手

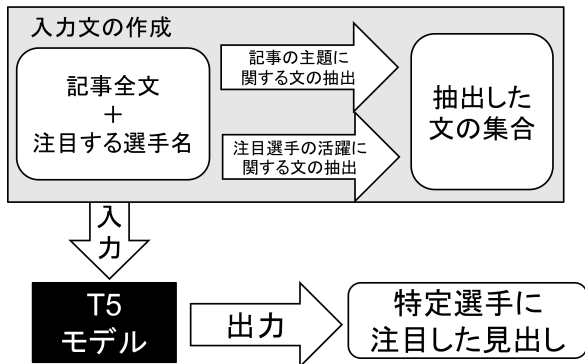


図2 手法の全体像.

がカーブを打ってホームランにした」という情報があるとき、見出しで「カーブを打った」ことのみを述べて、「ホームランを打った」ことに触れていなければ良い見出しとはいえない。

本研究では、上記の5点の特徴を含んだ見出しを生成することを目標とする。

3.2 手法の全体像

本節では、試合結果記事から特定選手に注目した見出しを生成するために、抽出型要約とT5を組み合わせた手法の全体像を説明する。

図2に本研究で提案する手法の全体像を示す。記事全文から、記事の主題に関する文の抽出と注目選手の活躍に関する文の抽出を行う。抽出した文を結合することにより、T5への入力文を作る。完成した入力文をT5に入力し、特定選手に注目した見出しを出力させる。

3.3 入力文の作成

通常、見出し生成をする際、モデルに記事全文を入力することが一般的である。しかし、記事全文をモデルに与えると、生成される見出しにおいて利用すべき情報を誤って選択する可能性がある。そこで、記事全文を入力しないことで、T5に余計な情報を与えない。これにより、注目していない選手の活躍を、注目している選手の活躍であるかのように見出し生成する可能性を下げることができる。

実際に、どのような特徴を持つ文を抽出するかについて、説明する。本手法において抽出する特徴を持つ文は2種類ある。1つ目は、記事の主題に関する文である。問題定義で定めた理想的な見出しが持つ特徴のひとつに「記事全体の要約を含んでいる」というものがある。記事の主題に関する文は、この特徴を生成した見出しに持たせるために抽出する。2つ目は、注目選手の活躍に関する文である。これは、生成した見出しに「指定した選手の選手名を含んでいる」という特徴を持たせるために抽出する。

3.3.1 記事の主題に関する文の抽出方法

記事の主題に関する文の抽出方法を説明する。図1の試合結果記事を具体例として実際に文を抽出していく過程を説明する。本研究では2つのアプローチを組み合わせることで、記事の主題に関する文を抽出する。

a) LEAD 法

1つ目のアプローチはLEAD法である。LEAD法とは、文書の冒頭に重要な情報が存在するという仮定のもと、文書内の先頭に位置する文や、先頭のいくつかの文を選択する抽出方法である。石川らはLEAD法は新聞記事の要約に特に効果的な手法として知られていると述べている[4]。試合結果記事においても記事の先頭にチームの勝敗など記事の主題となる重要な情報が書かれることが多い。そこで本手法では、記事の1文目を抽出し、入力文に含めることにする。図1の具体例に当てはめると「阪神は打線が爆発し、2桁11得点の猛攻で連敗を「5」でストップさせた。」という文を抽出することになる。

b) 本物の見出しと類似度が高い文を抽出する手法

2つ目のアプローチは、本物の見出しと類似度が高い文を抽出する手法である。本物の見出しは、記事の主題を読者に素早く伝えるために筆者が作成したものである。よって、本物の見出しと類似度が高い文は記事の主題を生成した見出しに含めるために必要な文であると考えられる。図1の本物の見出しは「阪神が打線爆発で連敗ストップ!1試合で首位返り咲き 高卒2年目の前川がプロ初猛打賞で牽引 中日戦6連勝」である。この見出しと類似度が高い文を本文から抽出する。そのために全ての文について本物の見出しとの類似度を求める。類似度は本物の見出しをクエリ、記事本文内の各文を文書と見なして計算する。ここでは「0.5ゲーム差で首位だったDeNAが敗れたため、1試合で首位に返り咲いた。」という文を具体例として、類似度の計算方法を説明する。

まずはMecab[5]を利用して記事本文と本物の見出しを形態素に分割する。その際、助詞と数値を省く処理をする。助詞を省くのは、「は」や「が」などの助詞は類似度をはかるのに必要無いためである。数値を省くのは、2桁と2年目では同じ「2」という形態素であったとしてもまったく意味が異なるためである。その結果、本物の見出しの形態素は「阪神;打線;爆発;連敗;ストップ;!;試合;首位;返り咲き;高;卒;年;目;前川;プロ;初;猛;打;賞;牽引;中日;戦;連勝」となる。「0.5ゲーム差で首位だったDeNAが敗れたため、1試合で首位に返り咲いた。」という文の形態素は「. ;ゲーム;差;首位;DeNA;敗れ;試合;首位;返り咲い」となる。そこから本物の見出しと類似度を計算している文で共起している形態素を抽出する。今回の例では「試合;首位」が抽出できる。

続いて、各形態素のIDF[11]を計算する。IDFは特定の単語が文書集合においてどれほど重要であるかを表す指標である。この指標を用いることにより、試合結果記事で現れることが少ない形態素が、本物の見出しの形態素と共起することの重要度を上げる。具体的には、「打」や「回」のような試合結果記事のあらゆる文で現れる形態素が、本物の見出しの形態素と共起することにより、「爆発」や「ストップ」のような、試合結果記事で現れることが少ない形態素が本物の見出しの形態素と共起することの重要度を上げている。式(1)を用いることで形態素 t のIDF値である idf_t を計算する。

$$idf_t = \log \left(\frac{10}{t \text{ が現れる文の数} + n} \right) \quad (1)$$

表 1 idf_t の計算結果 (一部).

形態素	値
爆発	0.51
ストップ	0.51
試合	0.51
首位	0.51
中前	0.36
打	0.11
回	0.11

式 (1) 中の n は定数である. n の値を大きくすると別の文で現れることが少ない形態素の重要度が上がる. 今回は n に 5 を代入する. また通常の IDF では, 対象記事の文の数が分子となる. しかし, 記事によって文の数が異なることにより, 形態素登場 1 回分の重要度が変動することを防ぐため, 分子は 10 で固定する.

図 1 の例における idf_t を計算した結果の一部を表 1 に示す. 1 つの文でしか登場しない形態素の idf_t の値は 0.51 であるのに対し, 「打」や「回」など複数の文で登場する形態素は 0.11 と低い値である.

続いて, 共起している形態素の idf_t の値の合計 idf_{sum} を計算する. idf_{sum} は式 (2) を用いて計算する.

$$idf_{sum} = \sum_{i=1}^k idf_{t_i} \quad (2)$$

この式における k は類似度を計算している文と本物の見出しで共起している形態素の数である. 「0.5 ゲーム差で首位だった DeNA が敗れたため, 1 試合で首位に返り咲いた。」という文と本物の見出しで共起している形態素は [‘試合’, ‘首位’] である. よって k の値は 2 となる. 式 (2) を用いて idf_{sum} の値を計算すると $0.51 + 0.51 = 1.02$ となる.

式 (2) で得られた idf_{sum} を用いて文の類似度を式 (3) で求める.

$$\text{rank} = \alpha \cdot idf_{sum} - \beta \cdot \text{morphemes} \quad (3)$$

この式では長い文にペナルティをかけている. 文が長ければ長いほど形態素が共起する可能性が上がるため, ペナルティをかける必要がある. morphemes は, 類似度を計算している文の形態素の数である. また α と β は, 値の重要度を表すパラメータである. 今回は α に 5, β に 0.1 を代入する.

今回は式 (3) で値が 3 以上になったものを本物の見出しと類似度が高い文として抽出する. 「0.5 ゲーム差で首位だった DeNA が敗れたため, 1 試合で首位に返り咲いた。」という文の形態素は [‘.’, ‘ゲーム’, ‘差’, ‘首位’, ‘DeNA’, ‘敗れ’, ‘試合’, ‘首位’, ‘返り咲い’] である. したがって, 形態素の数を表す morphemes の値は 9 である. 式 (3) を用いて, 「0.5 ゲーム差で首位だった DeNA が敗れたため, 1 試合で首位に返り咲いた。」という文の類似度を計算すると $5 \cdot 1.02 - 0.1 \cdot 9 = 4.2$ となる. この値は 3 以上であるので, 本物の見出しと類似度が高い文として 「0.5 ゲーム差で首位だった DeNA が敗れたた

め, 1 試合で首位に返り咲いた。」は抽出対象の文となる.

しかし, 生成する見出しで注目している選手が本物の見出しで注目している選手と同じではないときは, 文の類似度が 3 以上であっても抽出すべきではない. 具体的に図 1 の記事の例で説明すると 「終盤にも着実に加点し, 前川がプロ初の猛打賞をマークした。」という文の類似度は 10.85 で 3 以上である. しかし, 生成した見出しを前川選手の活躍に注目させたい場合はこの文を入力文に含めても構わないが, 生成した見出しを大山選手の活躍に注目させたい場合はこの文は入力文に含めるべきではない. そこで, 文の類似度が 3 以上かつ, 注目している選手以外の選手名が, その文に含まれない文を抽出対象の文とする.

3.3.2 注目選手の活躍に関する文の抽出方法

注目選手の活躍に関する文の抽出方法を説明する. 試合結果記事においては, 選手の活躍について述べる際にその文に選手名が入っていることが多い. そこで, 注目している選手の選手名が含まれていたとき, その文を抽出することにする. 図 1 の記事の例において, 大山選手に注目した場合は 「なおも 2 死一塁で, 大山が左越え 9 号 2 ランをマーク.」, 木浪選手に注目した場合は 「続く木浪が中前 2 点適時打を放ち, 再びリードを広げた。」という文が抽出対象の文となる.

3.3.3 抽出した文の結合

これまでに抽出した文章を結合して, T5 への入力文を作成する. 結合する際には, 文の順番を変えないようにする. 文の位置関係は見出し生成の際に重要な情報の一部であり, これを変更すると生成される見出しの精度を低下させる可能性がある. したがって, 抽出した文の位置関係は元の記事と同じになるようにする. 以下に, 図 1 の記事を例に大山選手に注目した場合と木浪選手に注目した場合の入力文を示す.

- **大山選手に注目した場合の入力文:** 阪神は打線が爆発し, 2 桁 11 得点の猛攻で連敗を「5」でストップさせた. 0.5 ゲーム差で首位だった DeNA が敗れたため, 1 試合で首位に返り咲いた. なおも 2 死一塁で, 大山が左越え 9 号 2 ランをマーク.

- **木浪選手に注目した場合の入力文:** 阪神は打線が爆発し, 2 桁 11 得点の猛攻で連敗を「5」でストップさせた. 0.5 ゲーム差で首位だった DeNA が敗れたため, 1 試合で首位に返り咲いた. 続く木浪が中前 2 点適時打を放ち, 再びリードを広げた.

これらの入力文は, 理想的な見出しを T5 に生成させるため, 記事全文から記事の主題に関する文と注目している選手の活躍に関する文のみを抽出している.

3.4 T5 による見出し生成

見出し生成に利用する文生成モデルである T5 について述べる. 本研究では T5 の事前学習済みモデルとして `sonoisai/t5-base-japanese1` を用いる. このモデルを見出し生成に適した形にするためにファインチューニングをする.

見出しの生成には, ビームサーチを用いる. ビームサーチは

1: <https://huggingface.co/sonoisai/t5-base-japanese>

生成の各ステップにおいて次の単語として出現する可能性の高い、いくつかの単語を選択していく手法である。この手法を用いて、上位 100 件の見出しを T5 に出力させる。その後、1 位から順に生成した見出しを確認し、指定している選手名を含んでいる見出しを、指定した選手の活躍に注目した見出しとする。上位 100 件の中に指定している選手名を含んでいる見出しがひとつも存在しない場合は、生成失敗として処理する。

4 評価実験

生成した見出しの精度が実際に提案手法により向上したかどうかを確認するために、評価実験を行った。

4.1 評価基準

生成した見出しが良い見出しであるか判断するための評価基準を以下に示す。

- 事実と異なっていないか
- 指定した選手の選手名を含んでいるか
- 記事全体の要約を含んでいるか
- 日本語として自然であるか
- 見出しとして自然であるか

このように、問題定義で定めた理想的な見出しが持つ 5 つの特徴を、生成した見出しが持っているかについて評価した。事実と異なっていないか、指定した選手の選手名を含んでいるか、記事全体の要約を含んでいるかについては、基準を満たしているものを 1、基準を満たしていないものを 0 として、2 値で評価を行った。日本語として自然であるか、見出しとして自然であるかについては、基準を満たしているものを 5、基準を満たしていないものを 1 として 5 段階で評価を行った。

4.2 ファインチューニング

ここでは、評価実験に利用する T5 のファインチューニングについて述べる。ファインチューニングに利用した訓練データは 2021 年のデイリースポーツの阪神タイガースに関する記事 2,738 件である。このデータは Ceek.jp News² というサイトにより収集した。このサイトは、様々なニュースサイトをクロールして Web ニュースを収集し、まとめているサイトである。

デイリースポーツが配信する Web ニュースは、阪神タイガースに関連するニュースの場合、そのニュースの URL に tigers という文字が含まれる。そこで、Web ニュースの中から URL に tigers という文字が含まれるニュースを抽出した。これらのデータを利用してファインチューニングすることで、T5 に見出し生成のタスクを解かせることができるようになる。

ファインチューニングを実行する際に定めた、各種ハイパーパラメータの設定を表 2 に示す。また本研究では patience を 10 とした早期終了によりエポック数を決定した。検証データには、訓練データと同様の方法で収集した 2021 年のデイリースポーツの阪神タイガースに関する記事 152 件を用いた。

表 2 ファインチューニング学習条件.

項目	内容
学習率	3.0×10^{-4}
バッチサイズ	4
最適化アルゴリズム	Adam Weight Decay [7]

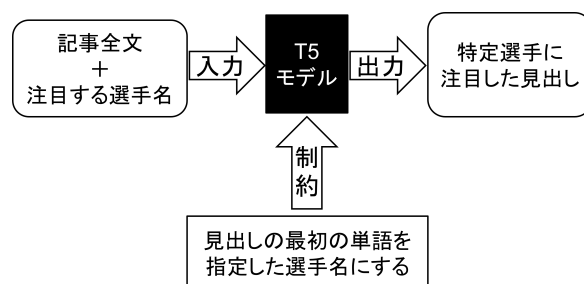


図 3 比較手法の全体像.

表 3 見出し生成時のハイパーパラメータ.

項目	内容
temperature	1
num beams	生成する見出しの数
diversity penalty	1
num beam groups	10
repetition penalty	10

4.3 比較手法

提案手法による見出し生成の精度の向上を確認するために、比較手法による見出し生成との比較を行った。比較手法では記事全文を入力文とした。文生成モデルの精度により生成される見出しの精度に差が出ないように、利用する文生成モデルは同じ訓練データで同じ学習条件によりファインチューニングした T5 とする。

比較手法では見出し生成時に、見出しの開始位置に注目する選手名のトークンを指定し、見出しを生成した。このように生成した見出しが、どの程度、評価基準を満たすことができるか確認した。比較手法の全体像を図 3 に示す。

4.4 見出し生成時のハイパーパラメータ

見出し生成をする際に定めた、ハイパーパラメータについて述べる。各種ハイパーパラメータの設定を表 3 に示す。temperature は生成にランダム性を加える温度パラメータである。値が高いほどランダム性が増す。num beams はビームサーチの探索幅である。diversity penalty は生成結果の多様性を生み出すためのペナルティである。num beam group はビームサーチのグループ数である。repetition penalty は同じ文の繰り返しを防ぐためのペナルティである。

4.5 評価用データセット

ここでは、評価用データセットの構築方法について説明する。評価用データの収集にも訓練データの収集に利用したサイトを用いた。訓練データでは 2021 年のデイリースポーツの阪神タイガースに関する記事を収集したが、評価用データでは 2022

2 : <https://news.ceek.jp/>

年のデイリースポーツの阪神タイガースに関する記事を収集した。収集した記事から試合結果について報じている記事 50 件を抽出した。1 件の記事に対して 2 人の選手に注目する。注目する選手は記事本文に出てくる選手の中からランダムに決定した。提案手法で生成された最も上位の見出しと、比較手法で生成された最も上位の見出しを評価の対象とした。

評価している見出しが提案手法により生成した見出しなのか、比較手法により生成した見出しなのか評価者が分かると評価する際にバイアスがかかってしまう可能性がある。そこで、評価者がどちらの手法で生成された見出しなのか特定できないようにした。

4.6 評価者

見出しの精度を正確に評価するためには、プロ野球に関する知識や、プロ野球に関するニュース記事の見出しを頻繁に目している必要がある。そこで、評価者はプロ野球ファン 3 名とした。

4.7 評価結果

本節では、評価者による評価の集計結果を示す。本実験は、人手による評価を行っている。そのため、評価者の評価の信頼性を確認する必要がある。評価者の信頼性を担保するために、各評価者の一致度を確認するため、Fleiss の κ 係数を各評価項目ごとに計算した。計算結果を表 4 に示す。 κ 係数の評価に関して、本実験では Landis ら [6] によって提案された基準を使用した。この基準によれば、指定した選手の選手名を含んでいるかの項目は「Almost perfect」（ほとんど一致）と評価され、事実と異なっていないか、記事全体の要約を含んでいるかの項目は「Substantial」（かなりの一致）と評価された。また日本語として自然であるかの項目は「Fair」（ある程度の一致）、見出しとして自然であるかの項目は「slight」（わずかな一致）であった。2 値評価である 3 項目については十分な一致が得られたといえる。5 段階評価である 2 項目については、2 値評価である 3 項目より κ 係数は下がるという結果が得られた。

評価の集計方法について説明する。2 値評価である場合、3 名の評価者の評価の多数決により集計した。5 段階評価である場合、3 名の評価者の評価の平均により集計した。その後、比較手法で生成された見出しと提案手法で生成された見出しを手法ごとに集計した。2 値評価のものは評価基準を満たした見出しの割合を計算し、5 段階評価のものは平均得点を計算した。

このように集計した結果を表 5、表 6 に示す。事実と異なっていないかという評価基準については提案手法のほうが精度が良いという結果が得られた。指定した選手名を含んでいるかという評価基準については比較手法のほうが精度が良いという結果が得られた。記事全体の要約を含んでいるかという評価基準については提案手法のほうがやや精度が良いという結果が得られた。日本語として自然であるか、見出しとして自然であるかという評価基準についても提案手法のほうがやや精度が良いという結果が得られた。また、提案手法と比較手法の結果の差が統計的に有意であるか確かめるため、有意水準 5% で、Welch

表 4 κ 係数計算結果.

項目	κ 係数
事実と異なっていないか	0.788
指定した選手の選手名を含んでいるか	0.886
記事全体の要約を含んでいるか	0.621
日本語として自然であるか	0.212
見出しとして自然であるか	0.169

表 5 評価基準を満たした見出しの割合.

手法	事実か否か	指定した選手名	全体の要約
提案手法	0.56*	0.90	0.64
比較手法	0.25	0.99*	0.60

表 6 平均得点.

手法	日本語としての自然さ	見出しとしての自然さ
提案手法	3.59	2.99
比較手法	3.39	2.70

の t 検定を行った。事実と異なっていないか、指定した選手名を含んでいるかという評価基準については、提案手法と比較手法の結果の差に統計的に有意差が見られた。

5 考察

本実験で得られた結果をもとに提案手法による見出し生成の精度の変化について考察する。

本実験により、提案手法は事実と異なっていない見出しを生成する可能性を上げることが分かった。提案手法と比較手法の事実と異なっていないかという評価基準における結果の差は統計的にも有意であった。これは、提案手法が記事全文ではなく理想的な見出しを生成するために必要な文のみを入力として使用していることが影響していると考えられる。これにより T5 が別の選手の活躍を注目している選手の活躍であるかのように見出しを誤生成するリスクを下げている。表 7 は実際に提案手法が別の選手の活躍を注目している選手の活躍であるかのように見出しを誤生成することを防いだ例である。記事全文を入力する比較手法では、実際に白星を挙げたのは藤浪投手であるにもかかわらず、佐藤選手が白星を挙げたかのような見出しを生成している。それに対して、提案手法では事実即した見出しを生成することが出来ている。

しかし、提案手法においても事実と異なる見出しを生成することがある。これは注目選手の活躍に関する文を抽出する際に、別の選手の活躍に関する情報が含まれてしまう可能性があることが原因である。提案手法では注目選手の活躍に関する文を抽出するための基準を、注目選手の選手名が含まれているか否かとしている。そのため別の選手の活躍の情報も含まれてしまう場合がある。具体例として、「先頭の中野が右前打で出塁すると、続く熊谷は左翼線へ運ぶバスターエンドランを成功。」といった文が挙げられる。このような文は 1 つの文で 2 人の選手の活躍について述べられており、誤生成の原因となる可能性がある。

表 7 提案手法が別の選手の活躍を注目している選手の活躍であるかのように見出しを誤生成することを防いだ例.

<p>記事全文： 阪神は3連勝を飾り、2年連続となる今季の巨人戦勝ち越しを決めた。先発・藤浪が21年4月16日・ヤクルト戦以来、491日ぶりとなる先発勝利で、今季初白星を挙げた。2試合連続の無四球で7回6安打1失点、5奪三振の好投。巨人戦の白星は16年4月5日以来、2328日ぶりだ。この日は打線も藤浪を強力援護した。二回に2死一、二塁から梅野の右前適時打で先制。三回には2死からロハスが中前打で出塁すると、続く佐藤輝がバックスクリーン左へ飛び込む18号2ランを放った。さらに3-0の四回には1死一、二塁で藤浪が三塁線への絶妙な犠打を決め、2死二、三塁と好機を演出。1番・中野が俊足を生かした三塁への適時内野安打で1点を追加した。なおも2死一、三塁では、島田がセーフティスクイズを決めて、五回までに5点のリードを奪った。</p>
<p>注目選手： 佐藤</p>
<p>提案手法の入力文： 阪神は3連勝を飾り、2年連続となる今季の巨人戦勝ち越しを決めた。巨人戦の白星は16年4月5日以来、2328日ぶりだ。三回には2死からロハスが中前打で出塁すると、続く佐藤輝がバックスクリーン左へ飛び込む18号2ランを放った。</p>
<p>提案手法： 阪神2年連続の巨人戦勝ち越し佐藤輝がバックスクリーン左へ18号2ラン</p>
<p>比較手法： 佐藤輝が今季初白星!藤浪が491日ぶり先発勝利</p>

表 8 指定した選手名を含まない例.

<p>記事全文： 阪神はまさかの逆転負け、横浜スタジアム7連敗となり、8連敗した98年以外の屈辱となった。同球場では今季10敗。これは78年の開場以降、ワーストの数字そして最短で11日にも優勝の可能性が完全消滅となる。6連戦の初戦を託された藤浪晋太郎投手(28)が突如、崩れて大量失点された。通算1000奪三振をこの試合で達成したが、素直に喜べない。2度の暴投が痛恨だ。1点差に詰め寄せられた五回、先頭から連打を許し、その後1死二、三塁とピンチを招いた。9番の代打・オースティンに対する初球150キロ・スプリットがワンバウンドし、梅野が捕球できず、三走・楠本が生還し、同点を許した。なおも1死三塁で、2ボールから投じた155キロ直球が外角低めに外れ、梅野のグラブをかすめる暴投に。三走・伊藤光が生還し、まさかの形で勝ち越しを許した。結局、オースティンに四球を与えると、桑原、佐野に連打を浴びて1死満塁と窮地に陥った。ここで3番・佐野が高々と打ち上げた左翼線への打球が、左翼・ロハスの前にポトリと落ちる適時打に。ベンチはたまらず、交代が告げられた。序盤3回は、DeNA打線を1安打無失点に抑える圧巻の投球を見せていただけに、まさかの事態。4回1/3を今季ワーストの7失点でKOとなった。「ストレートに狙いを絞られて、うまく打たれてしまったという印象です。そんな中でも、なんとか粘りながら、もうひと踏ん張りしたかったなと思います」とコメントしている。打線は初回に近本の右翼線への先制打、大山の右犠飛で幸先良く先取点を奪ったが、二回以降は相手先発・上茶谷から決定打が出ず。三回までは毎回安打も、四回からはばたつと安打も止まってしまう。打線の流れも悪く、藤浪が失点。悪循環が続いてしまい、敗戦につながってしまった。</p>
<p>注目選手： 伊藤</p>
<p>提案手法の入力文： 阪神はまさかの逆転負け、横浜スタジアム7連敗となり、8連敗した98年以外の屈辱となった。これは78年の開場以降、ワーストの数字そして最短で11日にも優勝の可能性が完全消滅となる。三走・伊藤光が生還し、まさかの形で勝ち越しを許した。</p>
<p>提案手法： 三走・伊藤光がまさかの逆転負け横浜スタジアム7連敗</p>
<p>比較手法： 伊藤光が今季ワーストの7失点「もうひと踏ん張りしたかった」</p>

生成された見出しに指定した選手名を含むかについて評価した結果では、比較手法のほうが良いという結果が得られた。この結果は、統計的にも有意であった。これは、提案手法は生成した上位100件の中に指定した選手名を含まない場合、生成失敗として処理されるためである。比較手法においても指定した選手名を含まない場合が存在する。その例を表8に挙げる。この例では伊藤将司投手と同じ名字の伊藤光選手に注目した見出しを生成してしまっている。このような同じ名字の選手をどのように対処するか検討する必要がある。

記事全体の要約を含む精度については提案手法のほうがわずかに高い結果が得られた。このことから、T5に記事全文から

記事の主題を見つけさせる手法よりも、記事の主題に関する文を事前に抽出しておき、その情報をT5に渡すほうが精度がわずかに高くなることが分かる。しかし提案手法と比較手法の結果の差は統計的には有意でないため、今後より多くの評価データで精度を検証する必要がある。

日本語として自然さの生成精度についても提案手法のほうが比較手法よりわずかに高いという結果が得られた。比較手法が提案手法より日本語として不自然な見出しが多く生成された理由として、T5に見出しの最初の単語を指定した選手名にする制約をかけたことが考えられる。ビームサーチによる生成で本来最初に配置されるべき単語ではなく、制約により強引に選手

名を最初の単語として配置している。これが日本語としての不自然さを生む原因になっていると考えられる。ただし、統計的な有意差は無いため今後より多くの評価データで精度を検証する必要がある。

また見出しとしての自然さの項目で提案手法が比較手法を上回ったことにより、提案手法による見出しのほうが比較手法による見出しより扱っている情報が見出しとして妥当であることが分かる。これは重要な文を事前に抽出することで T5 に与えられる情報が見出しとして扱う情報として、妥当なものにほぼ限定されることが要因であると考えられる。しかし、この項目についても統計的な有意差は無いため、今後より多くの評価データで精度を検証する必要がある。

6 まとめと今後の課題

本研究では、T5 による特定選手に注目したニュース記事の見出し生成に取り組んだ。文生成モデルが事実と異なっていない見出しを生成する可能性を上げるため、入力文を記事全文ではなく重要な文を抽出した抽出文とする手法を提案した。

今後の課題として、今回の実験ではすべて試合結果記事を対象とした見出しを生成している点がある。本研究では対象記事が野球の試合結果記事に絞られているが、重要文の抽出方法を工夫することで、政治的な記事や経済に関する記事などの別ジャンルの記事を対象としても汎用的に精度の高い見出し生成が可能になると考えられる。

また、現時点では、見出しの生成というタスクまでしか到達していない。本研究で提案した手法を用いたシステムを構築し、実用化するためには、これからも多くの検証が必要である。

謝辞 本研究は JSPS 科学研究費助成事業 JP21H03774, JP21H03775, JP22H03905, による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] 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. Association for Computational Linguistics, 2019.
- [2] Jeffrey L Elman. Finding Structure in Time. *Cognitive science*, Vol. 14, No. 2, pp. 179–211, 1990.
- [3] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [4] Kai Ishikawa, Shinichi Ando, and Akitoshi Okumura. Hybrid Text Summarization Method based on the TF Method and the Lead Method. *NTCIR Conference on Evaluation of Information Access Technologies*, 2001.
- [5] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying Conditional Random Fields to Japanese Morphological Analysis. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp. 230–237, 2004.
- [6] J Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *Biometrics*, Vol. 33, No. 1, pp. 159–174, 1977.

- [7] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.
- [8] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving Language Understanding by Generative Pre-Training. 2018.
- [9] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *The Journal of Machine Learning Research*, Vol. 21, No. 1, pp. 5485–5551, 2020.
- [10] Qian Ruan, Malte Ostendorff, and Georg Rehm. Histruct+: Improving Extractive Text Summarization with Hierarchical Structure Information. *arXiv preprint arXiv:2203.09629*, 2022.
- [11] Karen Sparck Jones. A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *Journal of documentation*, Vol. 28, No. 1, pp. 11–21, 1972.
- [12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. *Advances in neural information processing systems*, Vol. 30, pp. 5998–6008, 2017.
- [13] Xianjun Yang, Yan Li, Xinlu Zhang, Haifeng Chen, and Wei Cheng. Exploring the Limits of ChatGPT for Query or Aspect-based Text Summarization. *arXiv preprint arXiv:2302.08081*, 2023.
- [14] Tianyi Zhang, Faisal Ladhak, Esin Durmus, Percy Liang, Kathleen McKeown, and Tatsunori B. Hashimoto. Benchmarking Large Language Models for News Summarization. *arXiv preprint arXiv:2301.13848*, 2023.
- [15] 江原駿介, 莊司慶行, 山本岳洋, Martin J.Durst. スタンスに合わせたニュースタイトルの自動生成によるニュース記事本文へのアクセスを促す情報提示. 第 15 回データ工学と情報マネジメントに関するフォーラム, 3a-9-1, 2023.
- [16] 高津弘明, 奥田真由, 松山洋一, 本田裕, 藤江真也, 小林哲則. 会話によるニュース記事伝達のための抽出型要約のパーソナライズ. 2020 年度人工知能学会全国大会 (第 34 回), 3Rin4-71, 2020.
- [17] 黒木開, 岩井大志, 川上孝介, 石塚湖太, 中田和秀. キーワードを考慮した BERT2BERT による広告文生成. 2021 年度人工知能学会全国大会 (第 35 回), 2D3-OS-7a-02, 2021.
- [18] 小川恒平, 永塚光一, 瀧美雅保. ニューラルネットワークを用いたニュース記事の抽出的三行要約. 2019 年度人工知能学会全国大会 (第 33 回), 4Rin1-11, 2019.
- [19] 福田芳樹, 伊東嗣功, 石井雅樹, 堂坂浩二. T5 を用いた感情要約文の生成. 言語処理学会 第 28 回年次大会, D8-2, 2022.
- [20] 木村輔, 田上諒, 宮森恒. 文エンコーダによるクエリ指向要約モデルの強化. 情報処理学会論文誌, Vol. 12, No. 4, pp. 66–79, 2019.