

2023年度学士特定課題研究論文

# サッカーのプレーデータを用いたテキスト速報の自動生成手法の提案

東京工業大学 工学院 経営工学系

指導教員 市瀬 龍太郎

学籍番号 20B12036

氏名 田中 智可良



## 論文要旨

スポーツサイト等で目にすることができるテキスト速報の作成は、スポーツデータ会社が手入力で行っており、直接そのスポーツの試合の観戦はできずとも、試合をリアルタイムタイムで追いたい人にとっては大変便利である。殊にサッカーにおいては、選手の自由度の高さから、多様なプレーが90分の間で展開されるため、テキスト速報の入力には大変な労力が伴う。これに対する省力化のための手段として、プレーデータを用いたAIによるテキスト速報の自動生成システムを本論文で提案する。選手のプレー内容を、時系列で数値データとしてまとめているプレーデータは、テキスト速報の自動生成システムの入力データとして利用できる。したがって、数値データからテキストを生成する (Data2Text) タスクに取り組む先行研究を参考に、既存手法であるNNベースによるシステムと、新手法であるルールベースによるシステムの2つを作成し、機械的評価と人手による評価でそれぞれ精度を比較した。結果として、ルールベース手法の精度がコンテンツ選択や文法の正確さ、固有名詞の出力の点でNNベース手法よりも大きく上回ったことから、テキスト速報を自動生成するシステムとして、ルールベースによる手法が有効だということが示された。



# 目次

論文要旨	iii
表目次	vii
図目次	ix
<b>1 はじめに</b>	<b>1</b>
1.1 研究背景	1
1.1.1 テキスト速報	1
1.1.2 プレーデータとは	3
1.2 研究目的	4
1.3 本論文の構成	4
<b>2 関連研究</b>	<b>5</b>
2.1 NN ベース型の Data2Text に関する研究	5
2.2 NN ベース型でない Data2Text に関する研究	7
<b>3 提案手法</b>	<b>9</b>
3.1 データセット作成	9
3.2 NN ベース型システムの手法	10
3.2.1 モデルの概観	11
3.2.2 Placeholder と Attention について	12
3.2.3 訓練・推論の流れ	12
3.3 ルールベース型システムの手法	14
3.3.1 モデルの概観	15
3.3.2 ルールベースによる Data2Text	15
3.3.3 T5 による Text2Text	16
<b>4 実験</b>	<b>17</b>
4.1 評価手法	17
4.1.1 機械的評価手法	17
4.1.2 人手による評価手法	18

---

4.2	実験結果 . . . . .	19
4.2.1	機械的評価による結果 . . . . .	19
4.2.2	人手による評価での実験結果 . . . . .	20
4.3	例文紹介 . . . . .	21
5	まとめ	25
	謝辞	27
	参考文献	29
A	テンプレート文	31
B	人手による評価の評価項目	33

# 表目次

1.1	テキスト速報データの例 . . . . .	2
1.2	プレーデータの主なコラム . . . . .	3
1.3	テキスト速報のタイプ . . . . .	4
3.1	テキスト速報内の選手名の数の分布 . . . . .	10
3.2	データセットの Statistics . . . . .	11
4.1	機械的評価の結果 . . . . .	19
4.2	人手による評価：意味的な正確さについての実験結果 . . . . .	20
4.3	人手による評価：選手名の出力精度に関する評価の結果 . . . . .	20
4.4	正解文と NN ベース型システムの出力とルールベース型のシステム の出力の比較 . . . . .	24
A.1	テンプレート文 . . . . .	32
B.1	人手による評価：意味的な正確さについての評価項目 . . . . .	34
B.2	人手による評価：選手名の出力精度についての評価項目 . . . . .	35





## 図目次

3.1	2つの提案手法の概観 . . . . .	10
3.2	LSTM アーキテクチャの概観 . . . . .	12
3.3	ルールベース型システムの概観 . . . . .	15



# 1

## はじめに

### 1.1 研究背景

本節では生成対象であるテキスト速報と，生成のために用いられるプレーデータの説明を行い，併せて，テキスト速報作成の対象とするスポーツの中で，サッカーを選んだ理由についても詳しく説明する．

#### 1.1.1 テキスト速報

テキスト速報とは，インターネット上のサイトに更新される，リアルタイムで行われているスポーツの試合の内容をまとめたテキストのことである．例えば，サッカーであれば以下の表1.1のように，試合に出場するチームの戦績やフォーメーション，あるいは特定の選手のプレー内容についての情報などが，テキスト速報として更新されていく．テキスト速報の入力はデータ会社が請け負っており，現在も全て手入力で行われている．そのため，表1.1にあるテキスト速報のように，ボールを蹴った足やボールの軌道など，細かい表現が可能であるが，その分，時間・金銭面においてコストがかかり，さらに，誤表記をしてしまう危険性もある．したがって，テキスト速報の入力を全てAIによって行うことができれば，これらのコストの削減に大いに貢献できる．これが研究としてテキスト速報の自動生成を目的とする最大の動機となっている．

表 1.1: テキスト速報データの例

試合 ID	ホーム	ア ウェ イ	前半:1 後半:2	時 間 (分)	テキスト速報
2021040201	C 大阪	鳥栖	1	0	鳥栖ボールでキックオフ、試合開始
2021040201	C 大阪	鳥栖	1	0	C 大阪 予想フォーメーションは 4-4-2. 前節からのスタメン変更は 2 人. 坂本, 原川が外れ西川, 藤田が入る
2021040201	C 大阪	鳥栖	1	0	鳥栖 予想フォーメーションは 3-1-4-2. 前節からのスタメン変更は 3 人. 島川, 山下, 本田が外れ松岡, 中野, 林が入る
2021040201	C 大阪	鳥栖	1	4	C 大阪 右サイドの敵陣中央で FK を獲得する. キッカーの西川は右足でクロスを上げると, 豊川がヘディングでそらし, 大久保がファーでヘディングシュートを放つ. ゴールネットを揺らす, 副審はオフサイドフラッグを挙げる
2021040201	C 大阪	鳥栖	1	4	主審が耳に手を当てるシグナルを見せ, VAR のチェックを待つ
2021040701	鹿島	柏	2	0	鹿島ボールでキックオフ, 後半開始
2021040701	鹿島	柏	2	0	柏 35 細谷 OUT → 31 I シノズカ IN
2021040701	鹿島	柏	2	4	柏 仲間にイエローカード

## サッカーを研究対象にした理由について

サッカーは誰もが知っているスポーツであり, その分テキスト速報の需要も高い. 野球やバスケットボール, テニスといったスポーツもサッカーと同様人気のスポーツではあるが, テキスト速報の作成はサッカーよりも簡易である. 例えば, 野球はサッカー同様プレイヤーの人数が多いスポーツではあるものの, テキスト速報においてはサッカーと異なり選手のプレーの具体性は求められない. どのようなプレーをしたかよりも, どのような結果になったかが重要視される. 対して, サッカーはどの選手が, どのようなプレーをしたかについての具体的な表現が重要視される. サッカーはフィールド上の選手の自由度が高いため, 得点までの過程が野球と異なり多様である. したがって, テキスト速報の入力者は, 複雑な試合の流れをテキストに落とし込まなければならず, その作業は野球と比べて高度な課題となる. バスケットボールも選手の自由度が高いスポーツであるが, 人数がサッカーより少ないため, その分複雑さは軽減される. したがって, サッカーのテキ

表 1.2: プレーデータの主なコラム

コラム	説明
試合 ID	試合を分類するためのラベル
攻撃履歴	一連の攻撃に割り当てられる数字 (ボールを奪ってから完全に奪われるまで)
時間	試合の経過時間
チーム名	特定のアクションを行った選手が所属するチームの名前
選手名	特定のアクションを行った選手の名前
位置	フィールドを縦 4, 横 6 に分割し, 縦横ごとにラベル付与 (縦:1~4, 横:1~6)
成功	選手のアクションが成功:1, 失敗:0
アクション	各アクション (ゴール, パス, シュートなど) を行った:1, 行わなかった:0
部位	シュートやパスなどをあてた体の部位 (左足:1, 右足:2, 頭:3, その他:4)

スト速報の自動生成は高度な技術を要すると同時に, スポーツ全般のテキスト速報の自動作成に貢献できる有用性を秘めている. これが著者がサッカーをテキスト速報の生成対象に選んだ理由である.

### 1.1.2 プレーデータとは

プレーデータとは, テキスト速報と同様, データ会社によって作成されている, サッカーの試合のイベント (ボールに関与した特定の 1 人の選手のアクション) を時系列で並べたものである. 主なコラムや, データセットのサイズについては表 1.2 に記した. これから特に重要なコラムである「攻撃履歴」コラムについて説明する.

#### 攻撃履歴

各イベントには, 「攻撃履歴」と呼ばれる特殊なラベルが付与されており, このコラムでは, 一方のチームが, 一連の攻撃をしているイベントに対して同じ攻撃履歴番号が割り振られている. ここでいう一連の攻撃とは, 相手のボールを完全に奪ってから, 相手にボールを完全に奪われるまでを指しており, 一度のパス失敗などだけでは攻撃履歴は途切れず, 連続した相手のプレーが行われるまでは続くものとしている.

攻撃履歴は一連のプレーに割り振られており, テキスト速報は一連の攻撃についての内容を端的に述べるものである. この関連性より, 特定の攻撃履歴を持ったイベント群と特定のテキスト速報の対応付けが可能となる. このデータアノテーションに関する詳細な説明は 3 章で詳しく行う.

表 1.3: テキスト速報のタイプ

コメントタイプ	プレーデータで表現できるか	データサイズ
特定の選手のプレーについて (ファウルも含む)	○	1521
試合の流れやチーム全体のプレーについて	△	815
試合のプレビュー	×	254
選手交代	×	345
審判について	×	68
その他 (試合開始など)	-	273

## 1.2 研究目的

第1.1章で説明したことを踏まえ、本研究では、テキスト速報を自動生成することを目的と定め、その手段としてプレーデータを用いることとする。本研究のタスクは、テレビ中継での試合実況を再現するタスクではないことを留意したい。また、いくつかのタイプのテキスト速報は、プレーデータからでは生成できないことも併せて留意したい(表1.3参照)。プレーデータは、「試合のイベントのみ」に注目するため、例えばテキスト速報の1番はじめに必ず言及される、それぞれのチームのプレビュー(過去の戦績やフォーメーション、注目選手など)についてのテキスト速報(表1.1参照)は、プレーデータから表現することは不可能である。他にも、選手交代や審判のVAR確認、ロスタイムについての情報はプレーデータに内包されていないため、表現することができない。そのため、本研究で生成するコメントのタイプは、「ある特定の選手のプレーに言及したテキスト速報」のみとする。このタイプのテキストは、テキスト速報全体の約半分を占めていた。

## 1.3 本論文の構成

本論文は、以下に示す流れで論じられる。第2章では、Data2Textに関する既存手法についての説明を行い、本研究との類似性、また相違点についてもみていく。第3章では、提案手法についての説明を行う。本研究は2つの手法を提案し、それぞれの手法を用いて実験・評価を行った。各々の手法についての説明を、関連研究との接点も交えて節ごとに行った。第4章では、提案した2つの手法の実験結果についてみていく。システムの評価は、機械的手法によるものと人手によるものの2つで行うこととし、それぞれの評価手法で用いた指標について詳しく説明した後、結果について考察を行った。第5章では、本論文の結びとして、本研究の貢献、そして今後の課題についてのまとめを行った。

# 2

## 関連研究

AI を用いた技術が加速度的に開拓されている現代において，例えば表データから読み取れることをテキストに変換するモデルのように，ある系列を，異なる種類の系列に変換するタスクに特化したモデルが続々と台頭してきている．本研究のタスクも，そのような Data2Text と呼ばれるタスクに特化したシステムの開発であり，すでに，Data2Text のためのモデルやベンチマークは多数存在している．本章では，Data2Text に関する既存研究を紹介しつつ，本研究との関連にも注目していく．

### 2.1 NN ベース型の Data2Text に関する研究

まずは NN ベース型による，Data2Text 手法を用いた関連研究を挙げる．

谷口らは，2015-16 に行われた English Premier League の 380 試合に関するプレーデータを用いて，本研究と同様，ある特定の選手のプレーに関する実況の生成をしている [1]．プレーデータは，本研究で用いるものと同様，各行にボールに関与した特定の選手の名前や，アクション内容などの情報がラベルとして付与されているデータである．彼らが提案したモデルは，MLP(多層パーセプトロン)を用いて，各イベントの特徴を抽出する Encoder と，エンコードされた実況の各単語をエンベディングしたものに加え，Encoder で抽出した隠れベクトルを入力として用い，Long short-term memory(LSTM) [2] と Attention 機構 [3] によって出力単語を予測する Decoder で構成されている．NN ベース型のこのモデルは，数値データを，Encoder に入力することで特徴抽出し，その特徴から，教師データである，テキスト速報の単語の文脈を考慮した文を生成するような仕組みとなっている．ま

た, Attention 機構によって, イベントの各コラムと, テキスト速報の各トークンとの関連を考慮した出力を生成している. しかし, NN ベース型のモデルは, 比較的高い頻度で教師データ内に出現する, 普通名詞(サッカーにおいてはクロスやシュートといったアクション名や, DFといったポジション名など)の出力は問題なく行えるものの, チーム名や選手名といった, あまり高い頻度で出現しない固有名詞の出力が不得意である. その課題を解決するために, 谷口らは, Placeholder と呼ばれる固有名詞出力のための工夫を行った. 具体的には, 実況に出てくる選手名やチーム名を, "Player-1" や "Team-1" というタグに置き換え, 変換したテキストを教師データとして用いた. もしモデルがタグを出力すれば, Attention Score を参照し, 最も Attention Score が高いイベントに属する選手名やチーム名を, タグと置き換える作業を行うようにした. さらに, 出力すべき選手名を保持しているイベントの Attention Score が相対的に高くなるように, 以下の式で計算される損失を Attention loss として損失関数に加えた.

$$Attention\ loss = \sum_{j=1}^m \sum_{k \in U_j} a_{kj} \quad (2.1)$$

$U_j$  は  $j$  番目のトークンに対する集合であり, 出力すべき選手名を保持していないイベントのインデックスを, この集合の要素としている. このロスを低くすることで, 出力すべき選手名を保持していないイベントの Attention Score を抑える効果が働き, 出力すべき選手名がタグと交換されやすくなる. 谷口らは, この手法によって, NN ベース型モデルが苦手とする固有名詞の出力の改善に成功している.

本研究では, 谷口らが提案したモデルや Placeholder を用いた固有名詞の出力手法を再現し, 構築したシステムを Baseline として扱う. ただし, 本論文で提案する Baseline モデルには, 彼らの手法とはいくつか異なる点がある. 1つは, 入力に用いるイベントの数が本研究においては可変長である点である. 谷口らは, モデルの入力として用いるイベントの数は5つと固定しているが, テキスト速報を表現するためには, 5つのイベントだけでは表現しきれない可能性がある. その点において, 本研究で用いるプレーデータでは, 第1.1.2章で紹介したように, 「攻撃履歴」コラムがイベントを, 一連の攻撃ごとに区別しているため, 攻撃履歴ごとに適切なテキスト速報を対応させることにより, 十分なイベントを用いてテキスト速報を表現できることが保証される. さらに2つ目の違いとして, 本研究で提案する NN ベース型のモデルにおいては, Encoder にも Decoder と同様に, LSTM を用いている. 谷口らは, 固定長のイベントを Encoder の入力として用いているため, MLP で一度に5つのイベントから特徴を抽出していたが, 本研究で提案する NN ベース型のモデルの Encoder には, 同じ攻撃履歴を持った可変長のイベントを入力として使用するため, 谷口らのモデルのように, MLP で一度に学習させることができない. 代わりに, LSTM を用いることで可変長に対応し, かつイベントが持つ時系列性を考慮したモデルとなっている. 以上が谷口らのモデルと本研究のモデルの異なる点である. より詳しい提案手法の詳細については, 第3章で紹介



する。

## 2.2 NN ベース型でない Data2Text に関する研究

一方で NN ベース型でない、Data2Text 手法に取り組んでいる研究者もいる。

Puduply らは、MLB(Major League Baseball の試合に関する Statistics-Table と、その試合の要約文のデータセット)と ROTOWIRE(Basketball の試合に関する Statistics-Table と、その試合の要約文のデータセット)の 2 つを Benchmark として用い、それぞれ Table から要約文を生成するタスクに当たっている [?]. MLB には 4 種類の Table があり、1 つ目は各イニングごとのチームの得点数、2 つ目は各バッターのスタッツ(バットに当たった回数、ヒットに貢献した回数など)、3 つ目は各ピッチャーのスタッツ(ストライクを取った回数、ボールを取った回数など)、4 つ目は得点が生じたイニングでの得点者やピッチャーに関するものである。前準備として、すべての表を用い、特定の時間の、特定の選手名やチーム名に関する行の情報を  $V(entity)$ 、MLB のみ特定のイニングに関する表現を  $V(event)$  と言語化し、要約文の各段落を  $V(entity)$  と  $V(event)$  の組み合わせで全て置き換えた (Macro Planning). 例えば MLB の要約文に以下のような段落がある。

“Keller gave up a home run to the first batter of the game – Cedric Mullins – but quickly settled in to pitch eight strong innings in the Kansas City Royals’ 9-2 win over the Baltimore Orioles in a matchup of the teams with the worst records in the majors.”

この文の内容から、段落を  $V(B.Keller)V(C.Mullins)V(Royals)V(Orioles)$  というように、 $V(entity)$  の組み合わせ (Macro Plan) で置き換えることができ、これを教師データとして用いる。以下、入力から出力の流れを説明する。まず、Macro Plan の候補 (Paragraph Plan) を Table から用意し、LSTM と Attention で構成された Content Selection モデルを通して、候補の中から 1 つを選びだす作業を行う。続いて、Macro Plan を、LSTM の Encoder-Decoder+Attention モデルを用いてテキストに変換するプロセスを経て、最終的な出力を得ている。

本研究で提案したルールベース型のシステムは、彼らの提案手法を参考にしていて、NN を用いて、直接データからテキストに変換するのではなく、データを一度簡易的なテンプレート文の時系列的な羅列に変換した後に、それらのまとまりから重要な部分を抽出したテキストを生成することで、時系列を考慮し、かつ幻覚と呼ばれる入力と矛盾するような内容を出力してしまう現象を防ぐ効果が望める。本研究で生成したい文は、段落で構成されていない短文であるが、一度プレーデータの各行をルールベースを用いて短文に変換し、それらのまとまりからテキスト速報を生成する方法は、彼らの手法と類似している。より詳しい手法の詳細については、第3章で紹介する。



# 3

## 提案手法

プレーデータからテキスト速報を生成するための手法として、2つのシステムを提案する。図3.1に、本研究で提案する2つの手法の概観を示した。本章でははじめに、2つの手法に共通している、攻撃履歴を用いたデータセットの作成法についての説明を行い、続いて、1つの提案手法であるNNベース型のシステムと、もう1つの提案手法であるルールベース型のシステムの手法の詳細について、それぞれ説明を行う。

### 3.1 データセット作成

本研究で用いたテキスト速報は、2021年のJ1リーグの40試合分のデータであり、プレーデータも同様の期間のものを使用した。

本論文で提案する2つのシステムに共通していることは、どちらも同じ「攻撃履歴」を持ったイベント群を入力とし、そのイベント群に対応したテキスト速報を教師データとして用い、これを再現したテキストを生成することである。したがって、イベント群とテキスト速報の対応付け、すなわちデータアノテーションを実験の前に行う必要がある。どのようにして、「あるテキスト速報がプレーデータ内の、どの一連の攻撃について説明しているのか」を探し出したのかについての説明をしていく。

本研究では、データアノテーションのためにデータセット内の2つのコラムを主に用いた。はじめに、プレーデータとテキスト速報データの両方に付与されている、「時間」コラムを参照した。テキスト速報データの「時間」コラムは、分のみの情報に対して、プレーデータの「時間」コラムは分と秒が記載されていたこ

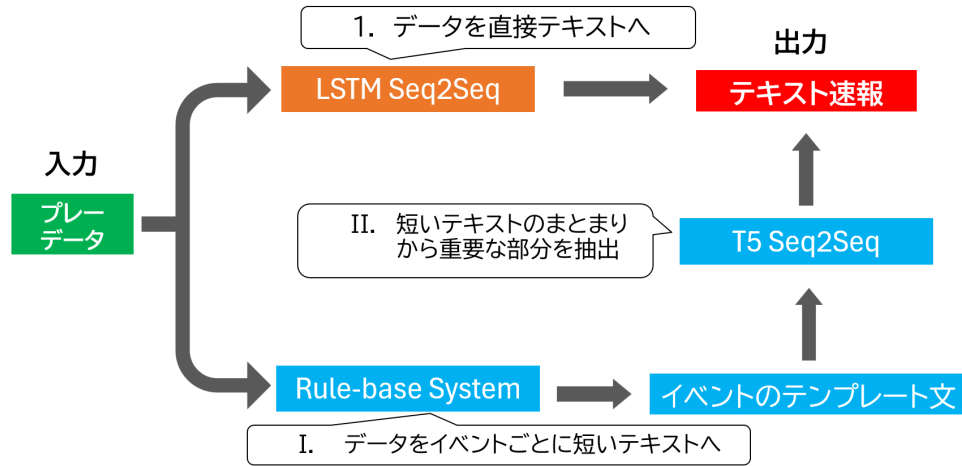


図 3.1: 2つの提案手法の概観

表 3.1: テキスト速報内の選手名の数分布

	1人	2人	3人	4人	5人	6人	7人
該当するテキスト速報の数	480	630	298	74	31	5	2

とから、プレーデータの「時間」コラムの分に当たる数字が、テキスト速報の「時間」コラムの値と等しいイベントを、データ対応付けの候補とした。次に、テキスト速報に出てくる選手名と、イベント内の「選手名」コラムを参照した。「時間」コラムを用いて、候補として絞った同じ攻撃履歴を持ったイベント群の中で、テキスト速報で言及された選手名を持つイベントがあれば、そのイベント群を最終候補とすることができる。表3.1が示すように、テキスト速報内に出てくる選手名数は、2人の場合が最も多いため、容易に候補を絞ることができた。もし候補が1つに絞れている場合は、その組み合わせを学習に用いるデータセットのペアとし、2つ以上の場合は、人力でプレーデータと照らし合わせながらペアを確定させた。この作業により、最終的に1522のデータセットが出来上がった。このデータセットを訓練データが85%、検証データが10%、テストデータが5%となるように分割し、2つの提案手法のシステムに使用した(表3.2参照)。

### 3.2 NNベース型システムの手法

本節では、Baselineとして用いるNNベース型システムの手法について詳細な説明を行う。はじめに、システム内で用いたモデルの概観について説明した後、モデルのEncoder, Decoderについての説明を行う。最後に第2.1章で説明したPlaceholderやAttentionといった手法の説明を行う。

表 3.2: データセットの Statistics

プレーデータのイベント数	87,528
テキスト速報のデータサイズ	3,276
学習・推論に用いたデータセットの数	1,522 訓練 : 1,294 (85 %) 検証 : 152 (10 %) テストデータ : 76 (5 %)
入力データ (プレーデータのコラム数)	NN ベース : 82 ルールベース : 75
語彙数	NN ベース : 1,128 ルールベース : 32,128

### 3.2.1 モデルの概観

LSTM と Attention 機構を用いた一般的な系列変換モデル (Sequence2Sequence) は, Text2Text のタスクに特化した構造となっている. Encoder は, Embedding Layer と LSTM ブロックで構成されている. トークナイズされ, エンコーディングされた入力系列を, Embedding Layer を通して単語間の特徴を反映したベクトルに変換し, それらをトークンの順序ごとに LSTM の各ブロックに入力していく. LSTM の最後のブロックの出力は, 系列の最初のトークンから最後のトークンの特徴を反映した隠れベクトルであり, これは続く Decoder の入力として使用される. Decoder も Encoder 同様, Embedding Layer と LSTM ブロックで構成されているが, Encoder と異なる点は, Attention 機構が LSTM の各ブロックの後に搭載されている点である. Attention 機構は, Decoder に入力された系列のトークンごとに, Encoder に入力された系列の各トークンとの関連度 (Attention Score) を求め, その関連度を考慮したベクトル (Context Vector) を生成する. さらに生成された Context Vector と, Decoder の LSTM ブロックの出力の両方の特徴を MLP など抽出した出力が, 最終的な Decoder の出力となる.

第2.1章で説明したように, 本研究においてはこの一般的な LSTM と Attention 機構を用いた系列変換モデルを, Data2Text 用にカスタマイズしている (図3.2参照). 本研究で用いた入力データは, プレーデータ内の, 「テキスト速報が対応された同じ攻撃履歴を持つイベント群」である. これらのイベントは, 全て One-Hot ベクトルで入力するため, Embedding Layer(*nn.Embedding*) を通してベクトル化することはできない. これに対処するために, 本研究では Embedding Layer に MLP を用いて各イベントの特徴を抽出するようにした. これによって, Encoder の入力が, たとえ ID 化することが難しい数値データであっても, 各トークンの特徴を抽出したベクトルを得ることができ, それらのベクトルを一般の系列変換モデルのように, LSTM ブロックの入力として用いることができる.

損失関数は, クロスエントロピー誤差関数と第2.1章で説明した Attention loss の合計とし, オプティマイザとして AdamW を用いてロスの最小化を行った.

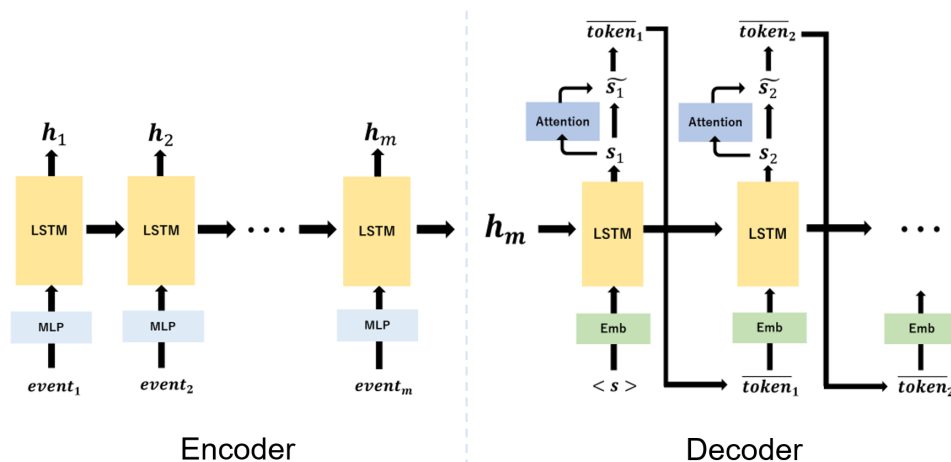


図 3.2: LSTM アーキテクチャの概観

その他，Decoder の入力系列として用いるテキスト速報の，トークナイズ用辞書を自作した．すべてのテキスト速報を，MeCab [5] と呼ばれる日本語に特化した形態素解析器によって分かち書きし，出現した語彙順に ID を割り振ることによって辞書を作成した．次節からは，固有名詞出力のために用いた Placeholder と Attention についての解説を行う．

### 3.2.2 Placeholder と Attention について

第2.1章でも説明したとおり，NN は固有名詞の出力を苦手としている．テキスト速報においても，多様な固有名詞(選手名やチーム名)が登場するが，その出現頻度は，ゴールやシュートといった普通名詞と比べて低い．そのため，データから直接固有名詞を含んだテキストを出力するのは難しく，何かしらのテクニックが必要となってくる．そこで，第2.1章でも説明したように，谷口らが提案した Placeholder [1] を用いて固有名詞の正確な出力を試みた．その手法の詳細については，第2.1章ですでに説明しているため参照されたい．Baseline である，谷口らの提案手法では，Attention Score を MLP を用いて計算していたが，より良い精度の出力を期待し，新たに Transformer [?] に用いられている Multihead Attention を用いた実験を併せて行った．

### 3.2.3 訓練・推論の流れ

ここでは実際の訓練・推論について，プレーデータ・テキスト速報の前処理から，テキスト速報出力までの流れを順にみていく．

### データの前処理

プレーデータで用いるコラムについては表1.2を参照されたい．これらのコラムの内，位置，成功，部位，アクションコラム(シュート，パス等)はすべて One-Hot ベクトルに変換し，それ以外のコラムは除去した．テキスト速報は先ほど説明した辞書を用いてトークナイズを行い，それらを Decoder の入力として用いた．

### Encoder について

図3.2が示すように，Encoder ではまず Embedding Layer である MLP に同じ攻撃履歴を持ったイベントを時系列順に入力していく．例えば，イベント数が  $m$  のイベント群を入力データとして用い，各イベントを  $\mathbf{e}_i (i = 1, 2, \dots, m)$  と表現すると決める．はじめに，各イベント  $\mathbf{e}_i$  を MLP に入力し  $\mathbf{p}_i$  を得る．

$$\mathbf{p}_i = \text{MLP}(\mathbf{e}_i) \quad (3.1)$$

パラメータとして，MLP は3層で構成し，MLP に入力するイベントのコラム数は76，MLP が出力するベクトル  $\mathbf{p}_i$  の次元数は128とした．MLP が出力したベクトルは，LSTM の各ブロックにイベントごとに入力され，それらは200次元の中間ベクトル  $\mathbf{h}_i$  として出力されるようにした．

$$\mathbf{h}_i = \text{LSTM}(\mathbf{p}_i, \mathbf{h}_{i-1}) \quad (3.2)$$

末尾の LSTM ブロックの出力  $\mathbf{h}_m$  は，Decoder の1つ目の LSTM のブロックの入力として用いられる．

### Decoder について

Decoder の Embedding Layer では，Pytorch の `nn.Embedding` を用い，テキスト速報の各トークンをベクトル表現に変換している．例えば，テキスト速報の1番目のトークン  $w_1$  は，文頭記号  $\langle s \rangle$  の ID であり，Embedding Layer によってベクトル表現  $\mathbf{u}_1$  に変換される．

$$\mathbf{u}_1 = \text{nn.Embedding}(w_1) \quad (3.3)$$

次に Embedding Layer で得た，ベクトル  $\mathbf{u}_1$  と Encoder の最終出力である  $\mathbf{h}_m$  を，1つ目の LSTM ブロックに入力し，1つ目の Decoder の隠れベクトル  $\mathbf{s}_1$  を得る．

$$\mathbf{s}_1 = \text{LSTM}(\mathbf{u}_1, \mathbf{h}_m) \quad (3.4)$$

次に Attention 機構で， $\mathbf{s}_1$  と Encoder の各中間層の隠れベクトル  $\mathbf{h}_j (j = 1, 2, \dots, m)$  との関連度を考慮した，コンテキストベクトルをつくる． $\mathbf{s}_1$  はクエリ，各隠れベクトル  $\mathbf{h}_j$  をキーとバリューとみなし，はじめにクエリとキーから MLP，または Multihead Attention を用いて類似度 (Attention score)  $a_{1,j}$  を計算した．前者ではま

ず，クエリ  $\mathbf{s}_1$  とキー  $\mathbf{h}_j$  をそれぞれ重みづけベクトル  $\mathbf{W}_{query}$ ,  $\mathbf{W}_{key}$  を用いて等しい次元数 (Attention MLP サイズ) にそろえ，新たなクエリ，キーを得る．得られたクエリとキーの和を求め，活性化関数を通し，さらに重みづけベクトル  $\mathbf{W}_{MLP}$  によってスカラー量に変換することで，Attention Score を得た．後者では，Pytorch の `nn.MultiheadAttention` 関数を用いて Attention Score を求めた．

$$MLP \text{ Attention Score } a_{1,j} = \mathbf{W}_{MLP}(\mathbf{W}_{query}(\mathbf{s}_1) + \mathbf{W}_{key}(\mathbf{h}_j)) \quad (j = 1, 2, \Gamma, m) \quad (3.5)$$

$$Multihead \text{ Attention Score } a_{1,j} = nn.MultiheadAttention(\mathbf{s}_1, \mathbf{h}_j) \quad (j = 1, 2, \Gamma, m) \quad (3.6)$$

次に Attention Score とバリューである，Encoder の各中間層の隠れベクトル  $\mathbf{h}_i$  との加重和を計算し，コンテキストベクトル  $\mathbf{c}_1$  を求めた．

$$Context \text{ Vector } \mathbf{c}_1 = \sum_{j \in 1, 2, \dots, m} a_{1,j} \cdot \mathbf{h}_j \quad (3.7)$$

コンテキストベクトル  $\mathbf{c}_1$  と， $\mathbf{s}_1$  を結合したベクトルを全結合層  $\mathbf{W}_1$  に入力し，2つのベクトルの特徴を抽出したベクトル  $\mathbf{s}'_1$  を得る．さらに，このベクトルを全結合層を通して最大語彙数と等しい次元のベクトルを抽出し，それに softmax 関数を通したベクトル  $\tilde{\mathbf{s}}_1$  を Decoder の1つ目のブロックの最終出力とする．

$$\mathbf{s}'_1 = \mathbf{W}_1([\mathbf{c}_1; \mathbf{s}_1]) \quad (3.8)$$

$$\tilde{\mathbf{s}}_1 = \text{Softmax}(\mathbf{W}_2(\mathbf{s}'_1)) \quad (3.9)$$

このベクトルの中で，値が最も大きいインデックスと，等しいIDを持った単語を辞書の中から探し出し，それを出力テキストとしての単語とする．この一連の流れを，モデルが文末記号  $\langle /s \rangle$  を出力するまで行う．

訓練時は70エポックで行い，バッチサイズは1として，逐次的に学習を行った．

次章からは，この提案手法とは全く異なるルールベースを用いた Data2Text 手法について説明する．

### 3.3 ルールベース型システムの手法

本節では，NN ベース型システムの対抗馬となりうる，ルールベース型システムについての詳細な説明を行う．本節の構成として，はじめにモデルの概観について説明した後，ルールベースによる Data2Text を行うシステムについての説明を行い，最後に事前学習済みモデルを用いて Text2Text を行うシステムについての説明を行う．



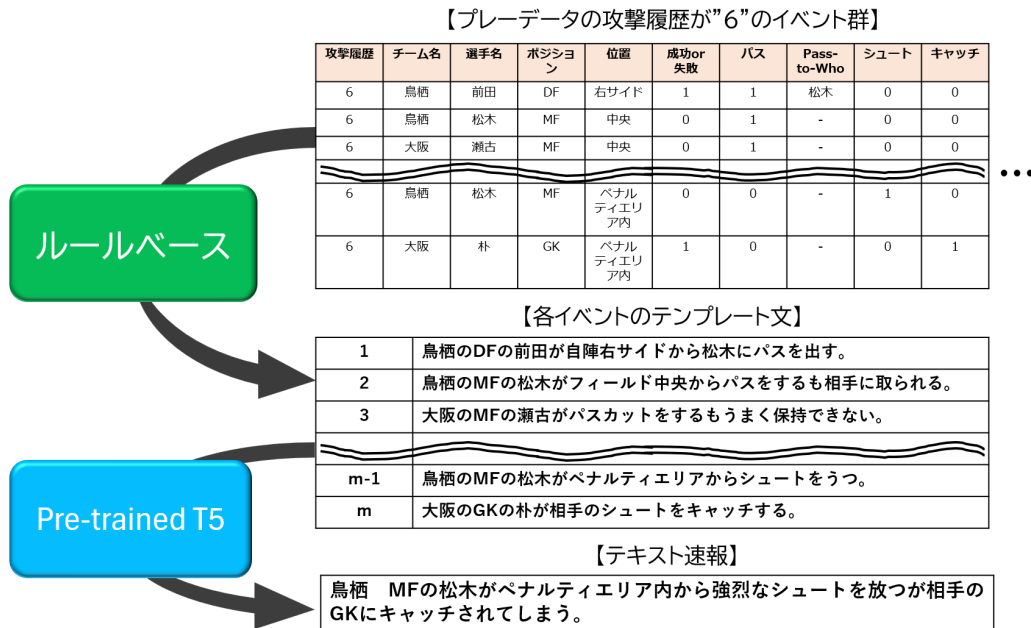


図 3.3: ルールベース型システムの概観

### 3.3.1 モデルの概観

ルールベース型システムの概観を、図3.3に示した。本手法では、2つのステップを経て、プレーデータからテキスト速報を生成する。第1ステップでは、各イベントをルールに従いながら、用意されたテンプレート文に変換していく。第2ステップでは、テキスト速報が対応付けられた、同じ攻撃履歴を持ったイベント群の各テンプレート文を、時系列にひとまとまりにした長い文として、それをT5(Text-to-Text Transfer Transformer) [7] の入力とし、要約のようなかたちでテキスト速報を生成していく。それぞれのステップの詳細について、次の節で順に説明していく。

### 3.3.2 ルールベースによる Data2Text

NN ベース型の手法ではモデルの性質上、用いるすべてのコラムを One-Hot ベクトルに変換したが、ルールベース型の手法では、データ加工を施さず、生のプレーデータからテキストを生成する。

まず、選手のアクションに応じたテンプレート文を用意した(付録A, 表A.1参照)。例えば、あるイベントのパスコラムの値が1の場合、パスに関するテンプレート文がイベントに対応付けられる。パスにもいくつか種類があるため、さらに細かくテンプレートを用意しておき、パスの種類に応じてテンプレート文を割り当てる。どのテンプレート文にも共通して選手名、チーム名、ポジションについて

の情報は必ず含むようにし、アクションに応じて、選手のフィールド上の位置など、追加して入れるべき情報を選択した。パスであれば、“パス ToWho”という誰にパスを出したかについてのコラムがあるので、テンプレート文にこのコラムも使用した。すべてのイベントを何らかのテンプレート文に変換した後、次のステップとして、これらのテンプレート文を用いてテキスト速報を生成する。

### 3.3.3 T5 による Text2Text

事前学習済み日本語 T5(retrieva) を用いて、ルールベースによってつくられたイベントのテンプレート文から、テキスト速報を生成していく。データセットの入力・出力のペアは NN ベース型システムと同様、同じ攻撃履歴をもったイベント群と、それに対応付けられたテキスト速報である。まず、同じ攻撃履歴を持ったイベント群の各テンプレート文を、時系列順に並べ、それを1つの長文とみなす。これを入力とし、T5 がその長文の中から、どの情報をテキスト速報に言及させるべきか（コンテンツ選択）を学習していく。

この手法の利点は、Text2Text に特化した事前学習済みのモデルを用いることができる点である。ルールベースによってイベントの重要な情報を漏らさずに、すでに日本語の文法を学習しているモデルにプレーデータの情報をインプットできることが、NN ベース型システムとの最大の違いである。

今回用いた T5 は、retrieva-jp/t5-base-long(パラメータ数：220M) と、retrieva-jp/t5-large-long(パラメータ数：770M) で、訓練時は8エポック、バッチサイズは8で学習を行った。

以上で、本研究で提案する2つの Data2Text 手法の説明を終わりとする。次章では2つのシステムの精度の評価を行う。

# 4

## 実験

Baseline である Attention Score 計算に MLP を用いた NN ベース型システム (NN-MLP) と, Multihead Attention を用いた NN ベース型システム (NN-MHA), さらに T5-base を用いたルールベース型システム (Rule-T5base), T5-large を用いたルールベース型システム (Rule-T5large) の 4 つのシステムの精度を, 機械的評価手法と人手による評価手法の 2 つを用いてそれぞれ評価する. 評価手法に関する節では, 評価手法として, 機械的評価手法と, 人手による評価手法の説明を行い, 続いて実験結果に関する節では, 各評価手法による結果についてみていく. さらにその評価手法による結果についての考察も併せて行った.

### 4.1 評価手法

#### 4.1.1 機械的評価手法

機械的評価手法とは, 数値計算によって生成システムの精度を計測する評価手法のことで, 本研究のタスクにおいては, ROUGE [8] という評価指標を用いた. 次項では ROUGE についての説明を行う.

#### ROUGE について

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) は, 2004 年に Chin-Yew によって発表された論文で提案された評価指標で, 「要約システムが作成した要約 (予測文)」と「人手で書かれた要約 (正解文)」の一致度を測る指標である. 名前の由来のとおり, 正解文の中のトークンと, どれだけ一致しているトークンが予

測文の中にあるかどうかを調べており、要約システムの評価によく用いられている評価指標である。本研究における、T5によるText2Textタスクは要約タスクと類似しているため、ROUGEを指標として用いることとした。ROUGEにはいくつか種類があり、今回用いたROUGEは以下の3つである。

- ROUGE-1: 式4.1が $N = 1$ のときの値。教師データとして用いられる実際のテキスト速報の1トークンと、どれだけ一致している1トークンが出力文内に存在しているかどうかを求める。
- ROUGE-2: 式4.1が $N = 2$ のときの値。実際のテキスト速報の、2つの連続しているトークンと、どれだけ一致している2つの連続しているトークンが出力文内に存在しているかどうかを求める。ROUGE-1と比べてより厳しく文法やコンテンツ選択の正しさを計測することができる。

$$ROUGE_N = \frac{\sum_{S \in \text{inreference}} \sum_{gram_n \in S} \text{Count}_{\text{match}}(gram_n)}{\sum_{S \in \text{inreference}} \sum_{gram_n \in S} \text{Count}(gram_n)} \quad (4.1)$$

- ROUGE-L: ROUGE-1,2と異なり、ROUGE-Lは生成した要約と人手で作成した要約との「一致する最大のシーケンス」(longest common subsequence=LCS)を評価する。したがって、たとえROUGE-2の値が高くても、ROUGE-Lの値が低いと文書全体の構造や文法が適切でないといえる。

各ROUGEにおける値をシステムごとに比較し、ROUGE-1からROUGE-2の変化に対する頑健性を見るなどして、コンテンツ選択や文法の精度を評価する。

#### 4.1.2 人手による評価手法

人手による評価手法として、テストデータを用いて出力したシステムの生成文の、意味的な正確さ・選手名出力の正確さの2つを、筆者が作成した尺度を用いてそれぞれ評価する。

##### 意味的な正確さの評価

まず、意味的な正確さを評価するために、各システムごとに、正解文と生成文を比較し、生成文が4段階の尺度(十分できている、最低限できている、一部誤りがある、全くできていない)のうち、どれに当てはまるかを分類し、数を求めた(詳しい評価項目の説明は付録B, 表B.1参照)。ROUGEによる評価と異なり、意味的な正確さの評価では、生成文が正解文に使用されていない語彙を使用していても、意味的に等しい説明をしていれば、その生成文は精度が高い生成文と評価されるような仕組みとなっているため、より実践的な評価手法となっている。なお、評価

表 4.1: 機械的評価の結果

		ROUGE-1	ROUGE-2	ROUGE-L
NN-	MLP	0.467	0.202	0.391
	MHA	0.442	0.172	0.359
Rule-	T5base	0.476	0.336	0.462
	T5large	<b>0.510</b>	<b>0.376</b>	<b>0.497</b>

にあたり，評価者によって結果のばらつきがでないと考えられることから，評価者としてサッカーに知見のある日本語話者 1 人に協力を仰いだ。

### 選手名出力の正確さの評価

固有名詞 (選手名) の出力の精度を比較するために，NN-MLP と NN-MHA，さらに Rule-T5large の 76 個の生成文が，3 段階の尺度 (十分できている，部分的にできていない，全くできていない) のうち，どれに当てはまるかを分類し数を求めた (詳しい評価項目の説明は付録 B，表 B.2 参照)。この結果より，2 つの NN ベース型システムのどちらが固有名詞の出力に長けているのか，さらにルールベース型システムが NN ベース型システムよりも，どのくらい固有名詞の出力において秀でているのかをみることができる。

## 4.2 実験結果

### 4.2.1 機械的評価による結果

機械的評価による結果を表 4.1 に示した。結果より，どの ROUGE においても Rule-T5large が最も精度がよいことがわかった。ROUGE-1 では，NN ベースとルールベースの値に大きな差がないように見えるが，ROUGE-2 においては，NN ベースがルールベースを大幅に下回った。評価手法の節で説明したとおり，ROUGE-2 は ROUGE-1 と比べてより厳しい評価を行うため，必ず値が ROUGE-2 の方が小さくなる。ROUGE-1 から ROUGE-2 への減少率を見ても，Rule-T5large では約 26 %，Rule-T5base は約 29 % 低下したのに対し，NN-MLP は約 57 %，NN-MHA は約 61 % と大幅に低下している。ROUGE-2 は，厳しく文法や，コンテンツ選択の正しさを計測しているため，NN ベースはその 2 点に関してルールベースよりも劣っていることが明白となった。また，NN ベースのみに注目すると，Baseline である NN-MLP の方が，NN-MHA よりも機械的評価においては精度がよかったこともわかった。なぜこのような結果になったのかについては，次の節で人手による評価の結果と併せて考察する。

表 4.2: 人手による評価：意味的な正確さについての実験結果

	NN- MLP	NN- MHA	Rule- T5base	Rule- T5large
十分に説明できている	6(7.9%)	6(7.9%)	<b>39(51.3%)</b>	<b>39(51.3%)</b>
最低限説明できている	4(5.3%)	8(10.5%)	10(13.2%)	<b>13(17.1%)</b>
少し説明に誤りがある	32(42.1%)	42(55.3%)	9(11.8%)	<b>8(10.5%)</b>
全く説明できていない	34(44.7%)	20(26.3%)	18(23.7%)	<b>16(21.1%)</b>
計	76(100%)			

表 4.3: 人手による評価：選手名の出力精度に関する評価の結果

	NN- MLP	NN- MHA	Rule- T5large
十分出力できている	11(14.5%)	13(17.1%)	<b>48(63.1%)</b>
部分的にできていない	21(27.6%)	34(44.7%)	<b>17(22.4%)</b>
全くできていない	44(57.9)	29(38.2%)	<b>11(14.5%)</b>
計	76(100%)		

#### 4.2.2 人手による評価での実験結果

人手による評価の結果についてもみていく。

まず、意味的な正確さについての結果を見ていく。表4.2に示した結果より、Rule-T5large の精度が最も高いことがわかった。「十分できている」の項目では、ルールベース型のシステムの T5base/large とともに全体のうちの 51.3 % の生成文が該当しているが、NN-MLP/MHA とともに、7.9 % とルールベースと比べて大きく下回った。「最低限できている」の項目では、T5large が T5base を 4 % ほど上回った。出力として許容できるテキストを、「十分できている」と「最低限できている」のいずれかに該当する生成文だとすると、最も精度が高かった T5large が 68.4 %、それにつづく T5base が 64.5 %、NN-MHA が 17.9 %、NN-MLP が 13.2 % と、やはり NN ベース型システムとルールベース型システムとでは、人間目線での評価は機械的な評価以上に明確にその精度の差を示す結果となった。さらに人手による評価においては機械的評価と異なり、NN-MHA の方が NN-MLP よりも高くなったことも注目すべきポイントである。

続いて、固有名詞の出力についての結果についてもみていく。表4.3に示した結果より、NN-MHA の方が NN-MLP よりも固有名詞の出力に長けていることがわかった。これらのことから、NN-MHA は NN-MLP と比較しコンテンツ選択や固有名詞の出力に長けていると考えられる。ROUGE においてのみ NN-MHA が下回ったのは、NN-MHA の方が、正解文で用いられている単語とは異なる単語を用いて、正解文と意味的に等しい文を生成しようとする傾向が強かったためと推測できる。ただ、いずれにしてもルールベース型システムの精度は、コンテンツ選択、文



法，固有名詞出力のどれにおいても NN ベース型システムよりも高いことが結果からわかった。

NN ベース型システムの精度が低くなった理由は主に 2 つ考えられる。1 つ目は，固有名詞の出力が Placeholder や Multihead Attention を用いても，なお完璧には行えなかったことだ。Attention Score に MLP を用いた NN ベース型システムによる出力を確認したところ，出力文に含まれる Placeholder のタグの数が多くなればなるほど，固有名詞の出力に誤りが生じる傾向にあることがわかった。特に，Player-1 と Player-2 のタグが出力されたとき，Player-1 の出力は正確に行われる場合が多かったものの，Player-2 の出力が Player-1 の対応する固有名詞と同じものになってしまうケースが頻発した。NN-MHA は，NN-MLP よりも同じ選手名が頻発するケースは少なかったものの，表 4.3 の結果からわかるように，十分に出力できたケースは少なかった。2 つ目は，幻覚(入力にすらない情報を出力してしまう現象)の起こり易さである。LSTM や T5 を含む NN モデルは，内部がブラックボックスのため，なぜ幻覚が起こるのかを理解するのは難しい。それを踏まえたうえで，Text2Text を行う T5 と比べて，NN ベースによる出力において幻覚が多く見られた理由として，NN ベース型システムの場合，数値を直接テキストに変換しているため，入力の特徴をテキストに反映するのが難しかったのではないかと推測。

一方で，ルールベース型システムの精度が高くなった理由は，事前学習済みのモデルを利用できた点にある。ルールベース型システムでは，大規模の日本語文章データセットから，日本語が持つ文脈のルールや語彙の特徴をすでに学習したモデルを用いることができるため，モデルをダウンロードした後は，自分のタスク用にパラメータの微調整 (*fine tuning*) を行えばよく，LSTM よりも短時間で，日本語の文脈の正しさを十分に考慮したテキストを生成することができる。これが結果を大きく分けた原因だと考えられる。

## 4.3 例文紹介

本章の最後に，正解文とそれに対応する NN ベース型システム，およびルールベース型システムの出力の組の例をいくつか紹介したい。表 4.4 には，「2 つとも (最低限) 正しい」，「2 つとも惜しい」，「NN ベースのみ誤っている」，「2 つとも誤っている」の 4 パターンに該当する出力の組の例を挙げている。

まず，「2 つとも (最低限) 正しい」例についてみていく。この正解文 (実際のテキスト速報) は，柏レイソルが右サイドでの CK を獲得し，キッカーである I シノヅカが右足でクロスを蹴り込むも，相手にニアサイドでクリアされた，といった内容である。それに対し，NN ベースの出力文は，右サイドからの CK を獲得した，という記述が含まれていなかったものの，キッカーの名前や，イベントの内容は矛盾なく説明できているため，最低限正しく出力できているといえる。ルールベースの出力文は，右 CK を獲得したことをきちんと表現できていたことに加え，内容も意味的に正しく説明できていた。

「2 つとも惜しい」例の正解文は，鳥栖のエドゥアルドが，敵陣の中央左寄り

のスペースから左足で浮き球を送り、林がそれに追いつこうとするも、GKに処理されてしまう、といった内容である。NNベースの出力文は、鳥栖の選手が、敵陣中央の位置から左足でクロスを上げるも、GKにキャッチされるといった内容を正しく表現できているものの、クロスを上げた選手とGKの選手名が適切でなく、同じ名前になってしまっている。これは人手による評価による実験結果に関する節で説明したように、Placeholderを用いた固有名詞表現のシステムに何らかの不備があったのではないと思われるが、その原因は明らかでない。また正解文では表現されていた、林がボールを受け取ろうとした描写もなかった。これは、ルールベースの出力文でも当てはまるエラーであった。この正解文に対応するプレーデータのイベント群を確認したところ、林に関しての情報はなかったことが判明した。つまり、今回の林のように、ボールに触ることができなかった場合は、その選手に関するイベントの描写や、"パス ToWho"コラムへの名前が記載されないため、プレーデータからではこういった選手に関する情報を出力するのは不可能であったことが明らかとなった。しかし、テキスト速報を入力するのは試合を直接目で見ている人間であるから、ボールに触れることができなかった選手の描写も行うことができる。これが原因で出力文に誤りを含んでしまう事例がいくつか見られた。

次に「NNベースのみ誤っている」例についてみていく。正解文は、横浜FCの松尾が、ドリブルで左サイドを駆け上がり、右足でペナルティエリア内へクロスを上げるも、DFの飯野にブロックされてしまう、といった内容が記されている。ルールベースの出力文では、松尾のクロスをブロックした選手の名前が記されていないものの、内容に矛盾はなく正しく描写できている。一方NNベースの出力文は、チーム名以外、正解文の内容と全く異なる内容の出力がなされている。このようなパターンもNNベースの出力文において散見された。

最後に「2つとも誤っている」例を紹介する。正解文は、名古屋の吉田が左サイドの敵陣中央からペナルティエリア左のディフェンスの裏にパスを入れるも、走り出した柿谷には合わなかった、という内容である。NNベースの出力文は、描写している内容自体は正解文と大きな乖離はないものの、パスの種類やパスを出した位置、加えて固有名詞のミスが見受けられた。ルールベースの出力文の内容は、NNベースの出力文よりは正解文と近い状況を説明できているものの、パスを出した選手や、パスを受けた選手の名前に間違いがあった。これはボールに直接関与していない柿谷の情報をプレーデータが含んでいないために、パスの受け手と出し手の選手名に混乱が生じて起きたエラーだと考えられる。これに対する改善策はプレーデータ自体を修正することであるが、再現性を考慮し修正は行わなかった。

ここで出力文の評価を経て得た知見をまとめる。まず、NNベース型システムがうまく出力できた文の特徴として、入力イベント数が少ないこと(すなわち一連の攻撃自体が短かった)、また出力すべき固有名詞も少ないことが挙げられる。これはイベント数が少ないことで、固有名詞の選択やコンテンツの選択が簡易になるためと考えられる。また、ルールベース型システムはプレーデータの情報不足により、どうしても説明できない部分を除けば、コンテンツ選択や固有名詞に



関して高精度な出力が可能であることがわかった．したがって，プレーデータの作成法をこのタスクにより適したものにするすることで，さらに精度が向上することが見込める．

表 4.4: 正解文と NN ベース型システムの出力とルールベース型のシステムの出力の比較

例文の種類	例文	
2 つとも (最低限) 正しい	正解文	柏 右サイドからの CK を獲得. キッカーの I シノヅカは右足でクロス をゴール前に蹴りこむ. だが相手にニア サイドでクリアされてしまう
	NN ベース	柏 キッカーの I シノヅカは右足でクロス を供給する. しかし, ゴールには繋がら ない
	ルールベース	柏 右 CK を獲得. キッカーの I シノヅカ はゴール前へクロスを入れるが DF に クリアされてしまう
2 つとも惜しい	正解文	鳥栖 エドゥアルドが敵陣中央の左から 左足で浮き球を送る. 林が駆け上がるが GK に処理されてしまう
	NN ベース	鳥栖 飯野が左サイドの敵陣中央の左足 で鋭いクロスを上げるが飯野に直接 キャッチされてしまう
	ルールベース	鳥栖 エドゥアルドが左サイドの敵陣 中央からペナルティエリア内へ浮き球の パスを入れる. しかし GK に処理 されてしまう
NN ベースのみ誤っている	正解文	横浜 FC 松尾がドリブルで左サイドを 駆け上がる. ペナルティエリア左まで 進入して右足でクロスを上げるが 飯野にブロックされてしまう
	NN ベース	横浜 FC 自陣深くまで前嶋が戻って アダイウトンからボールを奪う
	ルールベース	横浜 FC 松尾が左サイドの敵陣中央から ペナルティエリア中央へ 低い弾道の浮き球のパスを入れる. しかし相手に阻まれてしまう
2 つとも誤っている例	正解文	名古屋 吉田豊が左サイドの敵陣中央 からペナルティエリア左のディフェ ンスラインの裏へ縦パスを入れる. しかし走り出した柿谷へは合わない
	NN ベース	名古屋 稲垣が左サイドの敵陣中央から 左足でクロスをゴール前に供給する. しかし, 精度を欠いてボールはゴール ラインを割ってしまう
	ルールベース	名古屋 マテウスがペナルティエリア 手前の中央でボールを持ち左へパスを 出す. 吉田豊が受けて走り込むが精度を 欠いてボールはゴールラインを 割ってしまう

# 5

## まとめ

サッカーの試合のプレーデータを用いてテキスト速報を生成するタスクにおいて、NNベース型のシステムとルールベース型のシステムの手法を比較した。その結果、ルールベース型の方が、情報の正確性や事前学習済みのText2Textモデルを再利用できる点で、優れた結果を出すことが分かった。この結果より、ルールベース手法による、プレーデータを用いたテキスト速報の自動生成の有用性が示された。しかし、プレーデータから全てのコメントを正確に表現することは難しく、さらに、リアルタイムでの出力を求められるテキスト速報を生成するためには、プレーデータをリアルタイムで順次生成しなければならないという実務上の課題もある。このボトルネックを解消する最も良い方法は、動画から直接テキスト速報をキャプションすることである。映像はプレーデータよりも豊富な情報を含んでいるため、詳細かつ多様なタイプのテキスト速報の出力が期待できる。映像を入力とするアクション分類器、そして本研究で提案したルールベース型システムをうまく織り交ぜることによって、映像からテキスト速報のようなキャプションが生成できないかについて思案し、今後の課題として、システムの開発に踏み込んでいきたい。



## 謝辞

本論文を作成するにあたり，日夜手厚いサポートをしてくれた家族に感謝の意を表す．



## 参考文献

- [1] Yasufumi Taniguchi, Yukun Feng, Hiroya Takamura, and Manabu Okumura. Generating live soccer-match commentary from play data. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, pp. 7096–7103, 2019.
- [2] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. Neural computation, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [3] Thang Luong, Hieu Pham, and Christopher D. Manning. Effective approaches to attention-based neural machine translation. In Lluís Màrquez, Chris Callison-Burch, and Jian Su, editors, Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 1412–1421, Lisbon, Portugal, September 2015. Association for Computational Linguistics.
- [4] Ratish Puduppully and Mirella Lapata. Data-to-text Generation with Macro Planning. Transactions of the Association for Computational Linguistics, Vol. 9, pp. 510–527, 05 2021.
- [5] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to Japanese morphological analysis. In Dekang Lin and Dekai Wu, editors, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, pp. 230–237, Barcelona, Spain, July 2004. Association for Computational Linguistics.
- [6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, Advances in Neural Information Processing Systems, Vol. 30. Curran Associates, Inc., 2017.
- [7] 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.

- 
- [8] Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In Text Summarization Branches Out, pp. 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics.



# A

テンプレート文

表 A.1: テンプレート文

アクションの種類		テンプレート文
パス	スローイン	<チーム名>の<ポジション><選手名>が<位置>から (<パス ToWho>へ) スローインを投げた/するもパスは通らない
	スルーパス, フリックパス, クロス, フィード, パス	<チーム名>の<ポジション><選手名>が<位置>から (<パス ToWho>へ)<部位>で<アクション>をした/するもパスは通らない
シュート	ゴール	<チーム名>の<ポジション><選手名>が<位置>から<部位>のシュートで (直接) ゴールを決めた
	ゴールならず	<チーム名>の<ポジション><選手名>が<位置>から<部位>のシュートで (直接) ゴールをねらうも決めることはできない
被ファウル	イエロー/レッドカード	<チーム名>の<ポジション><選手名>が<位置>で<ファウルプレーヤー>に倒された (倒され PK を得る). <ファウルプレーヤー>はイエロー/レッドカードの判定
	else	<チーム名>の<ポジション><選手名>が<位置>で<ファウルプレーヤー>に倒された (倒され PK を得る)
オフサイド		<チーム名>の<ポジション><選手名>がオフサイドをした
ドリブル	ペナルティエリア脇侵入	<チーム名>の<ポジション><選手名>が<位置>でドリブルしペナルティエリア脇に侵入した
	ペナルティエリア侵入	<チーム名>の<ポジション><選手名>が<位置>でドリブルしペナルティエリアに侵入した
	else(成功/失敗)	<チーム名>の<ポジション><選手名>が<位置>でドリブルし相手をかかわした/するも相手にボールを奪われた
インターセプト		<チーム名>の<ポジション><選手名>が<位置>で相手のパスをインターセプトした
クリア		<チーム名>の<ポジション><選手名>が<位置>からボールをクリアした
ブロック		<チーム名>の<ポジション><選手名>が<位置>から相手の<前回のアクション>をブロックした
タックル (成功/失敗)		<チーム名>の<ポジション><選手名>が<位置>で<相手選手>にタックルしボールを奪った/するもボールは奪えない
空中戦 (成功/失敗)		<チーム名>の<ポジション><選手名>が<位置>で<相手選手>と空中戦を行い競り勝った/負けた
キャッチ (成功/失敗)		<チーム名>の<ポジション><選手名>が相手の<前回のアクション>をキャッチした/しようとするもファンブルしてしまう
ハンドクリア (成功/失敗)		<チーム名>の<ポジション><選手名>が相手の<前回のアクション>をハンドクリアした/しようとしたが手を弾きボールはゴールの中へ
トラップ (成功/失敗)		<チーム名>の<ポジション><選手名>が<位置>でボールをトラップした/しようとしたがうまく収まらない
ボールアウト		ボールアウトしてしまう

# B

人手による評価の評価項目

表 B.1: 人手による評価：意味的な正確さについての評価項目

評価尺度	説明	正解文と生成文のペアの例	
十分できている	コンテンツ選択○ 意味的に矛盾なし	正解文	横浜 FM 左 CK を獲得. キッカーの M ジュニオールが右足でクロスを上げるが GK にパンチングされてしまう
		生成文	横浜 FM 左 CK を獲得. キッカーの M ジュニオールはゴール前へクロスを入れるが菅野にパンチングではじかれてしまう
最低限できている	コンテンツ選択△ 大きな矛盾なし	正解文	大分 敵陣でボールをつなぐと最後は町田からのパスを受けた福森が仕掛けペナルティエリア手前の左からシュートを放つ. しかし相手に当たり枠をとらえられない
		生成文	大分 福森が左サイドの敵陣中央から右足でカーブを掛けたシュートを放つ. しかし DF にブロックされてしまう
少し説明に誤りがある	コンテンツ選択△ 固有名詞一部△ 意味一部矛盾あり	正解文	G 大阪 L ペレイラが前線でボールをキープし中央へパスを送る. 受けた小野瀬がペナルティエリア手前中央から右足でシュートを放つが味方に当たってコースが変わってしまう
		生成文	G 大阪 黒川がペナルティエリア手前の中央から右足で浮き球のパスを入れる L ペレイラが走りこむがシュートには持ち込めない
全くできていない	コンテンツ選択× 固有名詞× 内容が大きく異なる	正解文	横浜 FC 松尾が左サイドの敵陣中央から前線にボールを送るとクレーベがペナルティエリア左で受ける. しかし DF に囲まれてボールを失ってしまう
		生成文	横浜 FC 小川が右サイドの敵陣中央からペナルティエリア内へ浮き球のパスを入れる. しかし精度を欠きボールはゴールラインを割ってしまう

表 B.2: 人手による評価：選手名の出力精度についての評価項目

評価尺度	説明	例文	
十分できている	ポイント○ 選手名○	正解文	鹿島 エヴェラウドがパスを受けると、フィジカルの強さを生かしてキープし前を向く。そのまま持ち込んでペナルティエリア手前の中央からシュートを放つも槇野にブロックされてしまう
		生成文	鹿島 エヴェラウドがドリブルで持ち上がるとペナルティエリア手前の中央から右足でシュートを放つ。しかしDFにブロックされてしまう
部分的にできていない	ポイント× 選手名○ or ポイント○ 選手名×	正解文	神戸 後方でのパス回しから、味方が右サイドの敵陣中央へ縦パスを送る。藤本が抜け出して収めるがオフサイドの判定となってしまう
		生成文	神戸 藤本が自陣から右サイドの敵陣深くへロングボールを供給するが藤本のポジションはオフサイドの判定
全くできていない	ポイント× 選手名×	正解文	湘南 ボールを運んだ舘がハーフウェーライン付近で後方から古橋に倒されてしまいファウルの判定となる
		生成文	神戸 菊池が山川に対してプレッシャーを掛けてミスキックを誘う