

トラッキングデータからのサッカー実況生成

染谷 大河^{1,2} 石垣 達也² 高村 大也²¹ 東京大学 ² 産業技術総合研究所

taiga98-0809@g.ecc.u-tokyo.ac.jp ishigaki.tatsuya@aist.go.jp

takamura.hiroya@aist.go.jp

概要

サッカーの試合の臨場感や興奮を直接的に伝える主要な手段として、アナウンサーなどによる試合中の実況がある。一方で、サッカーの試合実況は専門人材やコスト面での制約が大きく、下位リーグやアマチュア、育成年代といったあらゆるカテゴリにおいて実施することは容易ではない。そこで、本研究では、サッカーにおける多様なカテゴリの試合に実況を付与することを目指し、試合映像や選手・ボールの位置情報等の情報から自動的に実況を生成する下流タスクを新たに提案する。Jリーグの試合映像に対して実況を付与したサッカー実況生成モデル構築のための大規模なデータセット **Live Football Commentary (LFC)** を作成した上で、選手・ボールの位置情報データを入力として、大規模言語モデル (LLM) を用いて実況を生成するベースラインモデルを構築し、評価を行なった。

1 序論

スポーツ観戦の臨場感や興奮を直接的に伝える上で、試合中の実況は非常に重要な手段である。実況者 (アナウンサーや解説者) が試合の状況や選手の動きをリアルタイムで伝えることで、視聴者が目の前のプレーを正確に把握できるだけでなく、さらなる没入感や高揚感が得られる。しかし、このような実況を提供するには、専門的な人材や制作体制が必要となり、コスト面での負担が大きい。そのため、下位リーグやアマチュア、育成年代などのカテゴリの試合で実況を付与することは容易ではない。そこで本稿では、特にサッカーの試合に対する実況の自動生成に取り組む。試合映像の解析は容易ではないため、昨今の GPS ジャケットなどの機器の普及を鑑みて、選手のトラッキングデータから実況を生成する。実況は、試合の進行と合わせて読んだり聞いたりするのに適した発話でなければならない。し

たがって、視聴者に没入感を与える自然な実況を自動で生成するためには、試合の流れやシーンのうち最も着目すべき部分を正確に把握する必要がある。

本研究ではまず、サッカーにおける自動実況生成タスクの確立に向けて、Jリーグ 2021 年シーズンの 40 試合に対して実況を付した、サッカー実況生成モデル開発のための大規模なデータセット **Live Football Commentary (LFC)** を構築する。また、対象となる試合に対応するトラッキングデータ (各選手とボールの位置座標の時系列データ) を入力として実況を生成する LLM ベースのモデルを構築し、評価を行なった。

2 関連研究

これまでサッカーデータを対象にしたテキスト生成研究の対象となっていたのは、主に**テキスト速報**と呼ばれるものであった。テキスト速報は、映像を視聴できない環境においても試合の展開を素早く把握でき、その臨場感を共有できる有用な手段である。しかし、テキスト速報は文章として読む対象であり、試合の進行とともに音声として聞いたり、サブタイトルとして読んだりする実況とは異なる。

研究例としては、プレーデータ (パス、シュートなどのイベントと、その位置や関わった選手からなる時系列データ) からテキスト速報を作成する試みと [1]、試合映像から作成する試みとがある [2, 3]。

一方で、スポーツの臨場感や興奮を直接的に伝えるもう一つの重要な手段である、アナウンサーや解説者による試合中の実況音声やテキストを対象にした研究は未だ少ない。ゲームや ActivityNet などの動画データに対する実況の試みは存在するが [4, 5, 6]、サッカーなど実際のスポーツの試合に対する実況データは存在しない。本研究では、サッカーの試合映像やその他のデータから実況を生成するタスクに取り組む、そのベンチマークとなるデータセットを初めて公開する。

統計値	前処理前	前処理後
総発話数	35,375	12,440
一試合あたりの平均発話数	884.4 ± 137.5	311.0 ± 106.6
一発話あたりの平均発話時間 (秒)	4.7 ± 4.1	2.9 ± 1.1
一発話あたりの平均発話文字数	29.9 ± 27.4	18.5 ± 9.2

表 1: 3 節に示す実況データと、4.1 節で示す前処理を施した後の実況データの各種統計値。

3 Live Football Commentary (LFC)

本研究では、サッカーのプロリーグである Jリーグの 2021 年シーズンの 40 試合に対する実況テキストから成るデータセット **Live Football Commentary (LFC)**¹⁾ を作成した。実況データを作成する際には、公益社団法人日本プロサッカーリーグの Jリーグの試合映像を用いた。この試合映像には実況音声が付随しているが、実況音声は著作権問題により研究に利用することはできなかつたため、あらためてプロのアナウンサーに実況を行うように依頼した。実況者に対する指示は以下の通りである：

- 発話は事実のみに基づいて行い、試合中に何が起きているかを実況音声のみで把握できるようにする。
- 不自然に長い沈黙が生じないようにする。
- 正式なサッカー用語に加えて、正式ではないが通常使われるサッカー用語も用いて実況してよい (例: 「マイナスの折り返し」「ポケット」「スルーパス」「アドバンテージ」)。
- 実況者は、サッカーの試合の実況の経験があるものに限る。
- 実況者は、事前に用意した登録選手リストを参照しながら、可能な範囲で選手名を用いて実況を行う。

実況音声は文字起こしした。さらに、テキスト化された実況データに対して、各実況の開始時刻と終了時刻を付与した上で、0.1 秒未満の間隔で連続する実況は一つの実況として統合した。作成されたデータの基本的な統計量や対象となった試合等の情報は、表 1 と付録の表 3 に示す。

1) 次の URL より入手可能：
<https://kirt.airc.aist.go.jp/corpus/ja/LFC>

4 実験

構築した LFC を用い、トラッキングデータから実況テキストを生成する言語生成課題の実験を行う。実験には試合映像は用いていない。

4.1 データ前処理

次に、3 節で作成した実況データに対して、対応するトラッキングデータを紐づける。トラッキングデータは、データスタジアム株式会社²⁾から研究目的で購入したものである。トラッキングデータと実況データの紐付けを以下のように行った。各実況に対応するトラッキングデータを抽出するため、実況終了時刻の 2 秒前を終点とし、そこから遡って連続する 50 フレーム (5 秒間に相当) を抽出した。ただし、抽出した 50 フレームの開始時刻よりも前に実況開始時刻がある場合、その実況は使用しないこととした。これは、抽出したフレームよりも前の状況を実況している可能性が高いためである。この処理により、合計で 12,440 件の実況とトラッキングデータ (50 フレーム) のペアが作成されたが、本実験ではそこから 700 件のペアをサンプリングし実験に用いた。具体的には、以下 500 件のデータを用いてファインチューニングした実験として報告する実験においては、700 件全体を train/dev/test=500 件/100 件/100 件に分けて実験を行なった。また、100 件のデータを用いてファインチューニングする実験においては、上記の train/dev/test のそれぞれからサブサンプリングすることで train/dev/test=100 件/20 件/20 件のデータを作成し、そちらを実験に用いた。そして、後者で作成された 20 件のデータセットを用いて、人手評価を行った (4.3 節)。

4.2 モデル

本研究では、OpenAI 社が提供する GPT-4o [7] ベースの実況生成手法を構築し、その性能を検証する。³⁾ トラッキングデータの入力方法について以下に詳述する 3 つのパターン (text, image, text+image) を比較検討する。これら 3 種類のモデルと 1 つのルールベースモデルの合計 4 種類のモデルの性能比較を行う。また、LLM ベースのモデルについては 4.1 節で述べた 100, 500 件のデータを用いてファインチューニングしたモデルも評価する。ファイン

2) <https://datastadium.co.jp>

3) 以下すべての実験において、執筆段階で GPT-4o シリーズの最新版であった gpt-4o-2024-08-06 を用いる。

チューニング後のモデルも含めて合計 10 種類のモデルを比較検討する。学習時のハイパーパラメタについては付録 A に示す。

Nearest Player: ルールベースの手法としては、入力されたフレームで最もボールに近い選手名を出力するモデル (Nearest Player) を使用する。入力されたフレームの中で最もボールに近い選手が切り替わった場合は、それらすべての選手名を出力する。

text: 入力となるトラッキングデータをテキストにして入力する方法。各選手のポジションや背番号、各選手とボールの位置と速度に加えて、各チームの攻撃方向やユニフォームの色の情報を入力した。ボールと選手の位置情報は、ピッチ座標を [0, 1] の範囲に正規化した値とし (図 3), 0.2 秒間隔で 25 フレーム分 (5 秒間) 入力した。また、ユニフォームの色が複数の色からなる場合は、より顕著な色を表示した。追加して、適切な実況生成を促すためのその他の指示や注意⁴⁾、試合時間やスコアなど試合状況、テストデータ以外からサンプルした実況発話 4 例を含めた。

image: トラッキングデータをもとに作成したアニメーションを複数の画像フレームに区切って入力する方法 (図 3)。⁵⁾ 各チームのプロットに用いる色は、テキストプロンプトに含まれるチームのユニフォームの色に一致させた。また、実況時に選手名を含められるよう、ボール付近の選手名やポジション、背番号などの情報をテキストとして入力した。追加して、text と同様に、適切な実況生成を促すためのその他の指示や注意、試合時間やスコアなど試合状況、テストデータ以外からサンプルした実況発話 4 例を含めた。

text+image: text と image を組み合わせた手法。基本的にこれらの 2 つの手法におけるプロンプトに含まれる情報を合わせた情報をプロンプトに入力するが、image のうち、text の情報と重複するボール付近の選手の情報を削除したほか、指示に画像と位置データ両方を入力する旨を追加した点で異なる。

4) 生成する実況の長さの指示は、作成した実況データ全体の平均値が 18.5 文字、標準偏差が 9.2 であったことに鑑みて、平均 ± 標準偏差 に概ね匹敵する長さの範囲とした。

5) API の制約上入力画像の数は 10 に制限されているため、0.5 秒間隔で 10 フレーム (5 秒間) の画像フレームに区切って入力した。

指示:
以下に入力されるのは、あるシーンに対応するボールと選手の位置データです。
このデータを基に、自然で流暢な日本語の実況を生成してください。生成された実況には、チーム名や選手名を適切に含め、試合の状況を的確に表現してください。

例:
ここは、相手陣内深くまで入っていきまされたけれども、一旦戻ります中野ボールを追いかける
しかし、体を使ってボールを奪いました山本です
ファンゾッから、少し下がった位置に来た松岡

注意:
・パスやプレーの流れを正確に反映してください。
・敵味方を混同しないように注意してください。
・プレーの場所や方向は特に注意してください。
・実況 1 つのみを生成しそれ以外のことは何も言わないでください。
・実況は、10 文字から 30 文字程度かそれより短いものを生成してください。
・Frame 1, 2, 3 と時系列になっていることに注意してください。

前半 21 分 16 秒
スコア: 横浜 F C 0 - 1 サンフレッチェ広島

Frame 1
ボール: (0.32, 0.98) 速度: 3.27 m/s
サンフレッチェ広島 (攻撃方向: 左 ユニフォーム: 白):
東 俊希 (DF 24): (0.33, 0.98) 速度: 0.84 m/s (最もボールに近い選手)
森島 司 (MF 10): (0.34, 0.78) 速度: 2.07 m/s
エゼキエウ (MF 14): (0.24, 0.81) 速度: 1.41 m/s
川辺 駿 (MF 8): (0.43, 0.82) 速度: 3.39 m/s
今津 佑太 (DF 33): (0.52, 0.67) 速度: 0.59 m/s
ドウグラス ヴィエイラ (FW 9): (0.23, 0.49) 速度: 1.54 m/s
柴崎 晃誠 (MF 30): (0.27, 0.46) 速度: 2.27 m/s
荒木 隼人 (DF 4): (0.48, 0.40) 速度: 0.31 m/s
大迫 敬介 (GK 38): (0.75, 0.51) 速度: 0.32 m/s
野上 結貴 (DF 2): (0.38, 0.18) 速度: 1.48 m/s
浅野 雄也 (MF 29): (0.22, 0.15) 速度: 1.22 m/s

横浜 F C (攻撃方向: 右 ユニフォーム: 水):
ジャーマイン 良 (FW 14): (0.32, 0.88) 速度: 4.25 m/s
安永 玲央 (MF 15): (0.28, 0.79) 速度: 0.93 m/s
岩武 克弥 (DF 22): (0.21, 0.82) 速度: 0.70 m/s
渡邊 千真 (FW 39): (0.38, 0.74) 速度: 1.88 m/s
手塚 康平 (MF 30): (0.28, 0.62) 速度: 0.93 m/s
田代 真一 (DF 5): (0.21, 0.63) 速度: 1.83 m/s
中村 俊輔 (MF 10): (0.37, 0.56) 速度: 1.73 m/s
韓 浩康 (DF 26): (0.21, 0.50) 速度: 1.44 m/s
袴田 裕太郎 (DF 3): (0.21, 0.36) 速度: 0.98 m/s
松尾 佑介 (FW 37): (0.31, 0.32) 速度: 0.81 m/s
南 雄太 (GK 18): (0.04, 0.49) 速度: 0.52 m/s

Frame 2

...

図 1: text のプロンプト例。

4.3 評価方法

生成された実況テキストに対し人手評価を行う。評価の観点としては、テキスト速報生成の既存研究 [1] を参考にしつつ、正確性・関連性・理解の容易さ 3 つのアスペクトでそれぞれ 3 段階で評価する。⁶⁾ 人手評価は著者のうち 2 人で行い、その平均を報告する。

また、本研究では自動評価は行わない。これは、実況テキストのもつ性質上、一般にある場面に対して最も適切な「正解」となるような実況は存在せず、人間の実況者による実況テキストを「正解」として、そこからの逸脱度を評価することの意義は限

6) 評価アスペクトの詳細な説明については付録 C に示す。



図 2: モデルによる実況生成例と実況者による実況.

	正確性	関連性	理解の容易さ
Nearest Player	1.28	1.30	1.45
GPT4o-txt	2.18	1.85	3.00
GPT4o-img	1.53	1.48	3.00
GPT4o-txt+img	1.55	1.60	2.98
GPT4o-ft100-txt	2.03	1.85	2.88
GPT4o-ft100-img	2.00	1.88	2.68
GPT4o-ft100-txt+img	1.93	1.80	2.78
GPT4o-ft500-txt	1.78	1.70	2.60
GPT4o-ft500-img	2.15	2.03	2.70
GPT4o-ft500-txt+img	2.05	1.58	2.80
Gold	2.53	2.55	2.78

表 2: 各モデルの人手評価の結果. 評価者 2 人による 3 段階評価の平均の結果を示す. ft100, ft500 はそれぞれ 100 件, 500 件でファインチューニングを行ったモデルを示す. text は txt, image は img と略記した.

定的であり, 各実況テキストがある場面を記述しより楽しむためのものとしてどれほど適切であるかを評価するべきであると考えられるためである.

5 結果

各モデルの人手評価の結果と生成例をそれぞれ表 2 と図 2 に示す. 理解の容易さにおいては, ほぼ全てのモデルがスコアが 2.5 を上回っており, どのモデルも正しく理解のしやすい文章を生成することができていることが確認できた. また, 正確性においても 2 前後のスコアを達成したモデルも多く, 課題は残るもののある程度正確にゲームの状況の一端を捉えることができていることが示唆される. 一方

で, ほとんどのモデルが関連性においてはスコアが 2 を下回った. これは, ゲームの状況のうち注目して実況すべき部分を捉えることは必ずしもできていないことを示唆する. 実際のゲーム中の場面において, 注目すべきとされる場面を認識するための手法の開発が必要であることを示唆する. また, ファインチューニングを行っても精度が必ずしも向上しなかった点は特筆すべき点である. この結果は, 単純なテキストや画像ベースの入力のみでは LLM を用いても適切な実況を生成するのは難しく, データ入力方法の工夫や新たなアーキテクチャの考案が必要である可能性を示唆する.

6 結論

本研究では, サッカー実況生成モデル構築のための大規模なデータセット **Live Football Commentary (LFC)** を作成し公開した. また, LLM を用いてトラッキングデータから実況テキストを生成する手法を考案し, 実験を行った. 提案手法は, 状況を表現したテキストと, 試合状況を表す画像を入力とした. 提案手法では, 理解の容易さおよび正確性の観点では, 良好な結果が得られた. しかし, 関連性については全体的に改善の余地があった. これは, 試合内の注目すべき場面を特定し実況することは容易ではなく, サッカーの自動実況生成タスクにおいては, 更なるデータ入力方法の工夫や新たなアーキテクチャの考案が必要である可能性を示唆している. 公開モデルを用いたより大規模なファインチューニングやトラッキングデータのその他の入力手法の検討などは今後の課題である.

謝辞

この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の助成事業 (JPNP20006) による支援の結果得られたものである。

参考文献

- [1] Yasufumi Taniguchi, Yukun Feng, Hiroya Takamura, and Manabu Okumura. Generating live soccer-match commentary from play data. **Proceedings of the AAAI Conference on Artificial Intelligence**, Vol. 33, No. 01, pp. 7096–7103, Jul. 2019.
- [2] Hassan Mkhallati, Anthony Cioppa, Silvio Giancola, Bernard Ghanem, and Marc Van Droogenbroeck. Soccernet-caption: Dense video captioning for soccer broadcasts commentaries. In **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops**, pp. 5074–5085, June 2023.
- [3] Jiayuan Rao, Haoning Wu, Chang Liu, Yanfeng Wang, and Weidi Xie. Matchtime: Towards automatic soccer game commentary generation, 2024.
- [4] Tatsuya Ishigaki, Goran Topic, Yumi Hamazono, Hiroshi Noji, Ichiro Kobayashi, Yusuke Miyao, and Hiroya Takamura. Generating racing game commentary from vision, language, and structured data. In Anya Belz, Angela Fan, Ehud Reiter, and Yaji Sripada, editors, **Proceedings of the 14th International Conference on Natural Language Generation**, pp. 103–113, Aberdeen, Scotland, UK, August 2021. Association for Computational Linguistics.
- [5] Yuki Saito, Shinnosuke Takamichi, and Hiroshi Saruwatari. SMASH corpus: A spontaneous speech corpus recording third-person audio commentaries on gameplay. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, H el ene Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, **Proceedings of the Twelfth Language Resources and Evaluation Conference**, pp. 6571–6577, Marseille, France, May 2020. European Language Resources Association.
- [6] Edison Marrese-Taylor, Yumi Hamazono, Tatsuya Ishigaki, Goran Topi c, Yusuke Miyao, Ichiro Kobayashi, and Hiroya Takamura. Open-domain video commentary generation. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing**, pp. 7326–7339, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.
- [7] OpenAI. Hello gpt-4o. <https://openai.com/index/hello-gpt-4o/>, 2024. Accessed: 2024-10-17.

A ハイパーパラメータ

GPT-4o の学習に使用した学習率は $1e-3$ 、エポック数は 3、バッチサイズは 5 である。

B 入力画像フレームの例

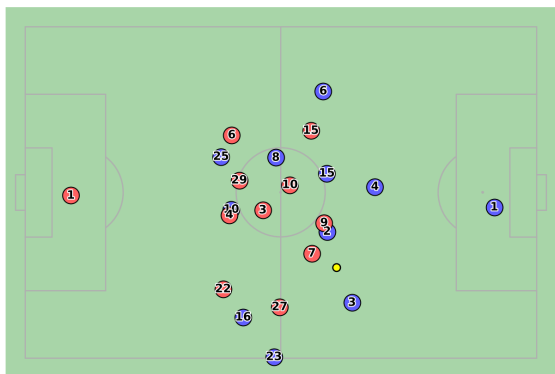


図 3: 入力画像フレームの例. 各選手とボールの位置座標と選手の背番号が表示されている。

C 人手評価のアスペクト

正確性

実況が、アニメーション、もしくはそこから推測される試合状況を正確に表していること。(ユニフォームの色については、色の系統が合っていれば良いものとする。例: 「青色」のところを「水色」と答えていても良い。)

評価基準 (3 段階評価):

- 1: 重大な誤りがある。
- 2: 軽微な誤りがある。
- 3: 誤りがない。

例:

- 高スコア: 「鈴木選手のヘディングが決まりました！」(鈴木選手がヘディングシュートを決めた場面において)
- 低スコア: 「田中選手のボレーシュートが決まりました！」(鈴木選手がヘディングシュートを決めた場面において)

関連性

実況がクリップ内の注目すべきプレーや状況に直接関連していること。

評価基準 (3 段階評価):

- 1: 実況すべきでない場面を実況している。
- 2: 直接プレーに関連するが、最も注目すべき場面ではない場面を実況している。
- 3: 最も注目すべき場面を実況している。

例:

- 高スコア: 「ゴールキーパーが見事なセーブを見せました！」
- 低スコア: 「鈴木選手がジョギングをしています。」(注目すべきプレーではない)
- 低スコア: 「ピッチには選手が 22 人いて、ボールがあります。」(クリップ内の特定のシーンやアクションに対応していない)

理解の容易さ

読み手が容易に理解でき、具体的な試合状況が想起されるか。(その試合状況が実際の試合状況と近い場合は関係がない。)

評価基準 (3 段階評価):

- 1: 状況が理解できない。
- 2: 一部不明瞭だが理解できる。
- 3: 状況がよく理解できる。

例:

- 高スコア: 「中村選手が右サイドからクロスを上げました。」
- 低スコア: 「彼がそれをして、ボールがあそこに行きました。」(具体的な状況が想起されない)
- 低スコア: 「試合がサッカーありました。」(文法性が低く理解できない)

D データセットに含まれる試合

LFC に含まれるのは表 3 に記載した試合の実況である。

節	試合開催日	ホーム	アウェイ
第 7 節	2021 年 4 月 2 日	C 大阪	鳥栖
第 7 節	2021 年 4 月 3 日	横浜 FM	湘南
第 7 節	2021 年 4 月 3 日	仙台	神戸
第 7 節	2021 年 4 月 3 日	名古屋	FC 東京
第 7 節	2021 年 4 月 3 日	広島	G 大阪
第 7 節	2021 年 4 月 3 日	福岡	札幌
第 7 節	2021 年 4 月 3 日	浦和	鹿島
第 7 節	2021 年 4 月 3 日	横浜 FC	柏
第 7 節	2021 年 4 月 3 日	川崎 F	大分
第 7 節	2021 年 4 月 4 日	清水	徳島
第 8 節	2021 年 4 月 6 日	横浜 FM	C 大阪
第 8 節	2021 年 4 月 7 日	鹿島	柏
第 8 節	2021 年 4 月 7 日	FC 東京	札幌
第 8 節	2021 年 4 月 7 日	川崎 F	鳥栖
第 8 節	2021 年 4 月 7 日	横浜 FC	広島
第 8 節	2021 年 4 月 7 日	湘南	名古屋
第 8 節	2021 年 4 月 7 日	清水	浦和
第 8 節	2021 年 4 月 7 日	G 大阪	福岡
第 8 節	2021 年 4 月 7 日	神戸	大分
第 8 節	2021 年 4 月 7 日	徳島	仙台
第 9 節	2021 年 4 月 10 日	広島	湘南
第 9 節	2021 年 4 月 10 日	C 大阪	福岡
第 9 節	2021 年 4 月 11 日	札幌	鹿島
第 9 節	2021 年 4 月 11 日	仙台	横浜 FM
第 9 節	2021 年 4 月 11 日	FC 東京	川崎 F
第 9 節	2021 年 4 月 11 日	鳥栖	横浜 FC
第 9 節	2021 年 4 月 11 日	大分	名古屋
第 9 節	2021 年 4 月 11 日	浦和	徳島
第 9 節	2021 年 4 月 11 日	柏	G 大阪
第 9 節	2021 年 4 月 11 日	神戸	清水
第 10 節	2021 年 4 月 16 日	札幌	横浜 FM
第 10 節	2021 年 4 月 17 日	横浜 FC	仙台
第 10 節	2021 年 4 月 17 日	徳島	鹿島
第 10 節	2021 年 4 月 17 日	福岡	FC 東京
第 10 節	2021 年 4 月 17 日	大分	柏
第 10 節	2021 年 4 月 17 日	湘南	神戸
第 10 節	2021 年 4 月 18 日	川崎 F	広島
第 10 節	2021 年 4 月 18 日	名古屋	鳥栖
第 10 節	2021 年 4 月 18 日	C 大阪	浦和
第 10 節	2021 年 4 月 18 日	G 大阪	清水

表 3: データセットに含まれる試合