

大規模マルチモーダルモデルを用いたロボカップサッカー標準プラットフォームリーグにおける自動実況システム

Development of an Automatic Live Commentary System for the RoboCup Soccer Standard Platform League using a Large Multimodal Model

大橋 玲音¹ 坪倉 和哉² 小林 邦和^{1*}

Reon OHASHI¹, Kazuya TSUBOKURA², Kunikazu KOBAYASHI¹

¹ 愛知県立大学 情報科学部

¹ School of Information Science and Technology, Aichi Prefectural University

² 愛知県立大学 大学院 情報科学研究科

² Graduate School of Information Science and Technology, Aichi Prefectural University

Abstract: ロボカップの試合は、人間のサッカーのルールと比べ複雑であり、競技者ではない観客にとって見どころやルールの把握が困難である。試合の実況・解説を行うことで、見どころやルールの把握が容易になり、観客がより楽しむことができるようになると考えられる。そこで本研究では、大規模マルチモーダルモデルを用いたロボカップサッカーの実況・解説を行うシステムを提案した。本システムによって生成された実況・解説文章に対して面白さ、明瞭さ、適切さの観点からアンケートを行った。アンケートのいずれの項目においても、肯定的な結果が半数以上得られた。

1 はじめに

ロボカップは人工知能やロボット工学の発展に貢献するプロジェクトとして知られている。その中でも、ロボカップサッカーでは、ロボットが自律的にサッカーを行い、西暦 2050 年までに「サッカーの世界チャンピオンに勝てる自律型ロボットのチームを作る」という目標に向けて研究や開発が進められている [1, 2]。

人間同士のサッカーにおいては、人間により実況や解説が行われ、サッカーに詳しくない人でも観戦を楽しむことが可能である。同様に、ロボカップサッカーにおいても実況や解説を行うことで、競技やルールに詳しくない人でも観戦を楽しむことができるようになると考えられる。ロボカップサッカーは人間のサッカーのルールを参考している部分も多いが、文献 [1, 2] でも説明されているように、各リーグ独自のルールが定められているため、人間のサッカー以上にルールの把握が困難である。ロボカップでは、一般の来場者が観戦することもを想定しているため、実況・解説を行うことで観客に試合を楽しんでもらうことは、ロボカップの活動への注目や興味を喚起するうえでも重要である。

これまで、シュミレーションリーグで実況・解説シス

テムが提案されていた [4, 5]。また、標準プラットフォームにおいても提案されているが、パスやシュートなどのロボットの行動に対する認識および自然な実況・解説文の生成が課題となっている [6]。そこで、本研究では、大規模マルチモーダルモデルを利用して、行動に対する認識および自然な実況・解説文の生成が可能なロボカップサッカー標準プラットフォームを対象とした実況・解説システムを提案する。これらの実況・解説文の生成及びロボットの行動に対する認識を大規模マルチモーダルモデルを用いて解決する。

以下、2 章では提案手法について述べる。3 章では提案手法の評価を行う。4 章で今後の課題について述べ、5 章でまとめを述べる。

2 提案手法

本章では、大規模マルチモーダルモデルを利用した試合状況の認識および実況・解説文の生成手法について述べる。

2.1 大規模マルチモーダルモデル

大規模言語モデル (Large Language Model, 以下 LLM と略記) は、膨大なテキストデータから言語の規則性を

*連絡先: 愛知県立大学 情報科学部
〒480-1342 愛知県長久手市茨ヶ廻間1522-3
E-mail: kobayashi@ist.aichi-pu.ac.jp

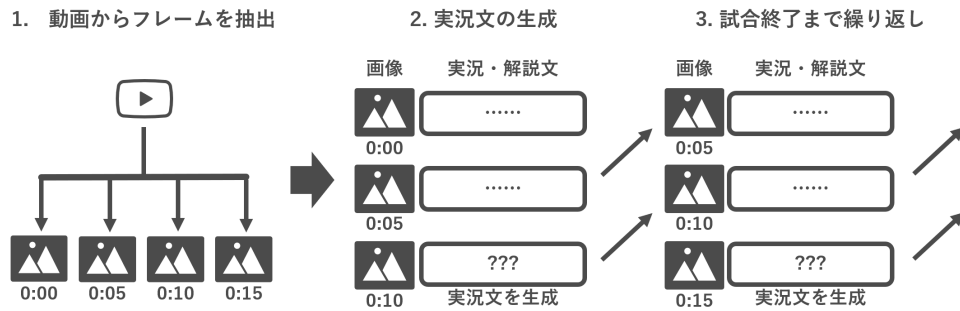


図 1: 実況・解説ロボットシステムの概要

学習することで、様々な言語タスクを実行できる人工知能モデルである。代表的な LLM として、GPT(Generative Pre-trained Transformer) シリーズ [7, 8], BERT(Bidirectional Encoder Representations from Transformers) [9] などが存在する。

また、大規模マルチモーダルモデル (Multimodal Large Language Model, 以下 MLLM と略記) は、テキストだけでなく画像などの複数のモダリティからも学習することで、テキスト以外に対するタスクにも対応した言語モデルを指す。代表的な MLLM は OpenAI の GPT-4v[10] や Google の Gemini シリーズ [11], Anthropic の Claude 3 シリーズ [12] が存在する。これらの MLLM はテキストに対するタスクだけではなく、画像に対するタスクも実行することができる。

本システムでは、これらの MLLM に対して、試合の画像を入力として与え、状況に適した実況・解説文を生成する。

2.2 実況・解説文の生成手法

本節では、提案手法の具体的な手順 (図 1) を示す。

まず、試合映像を定期的なインターバル (例えば 5 秒) でフレームを抽出し、各フレームは画像として保存する。

次に、実況・解説文の生成を行うために、システムプロンプト、対象フレームの直前の 2 フレームの画像、および直前の 2 フレームに対応する実況・解説文を入力として使用する。

具体的には、まず時刻 $t = 0$ および $t = 1$ の画像、実況・解説文を入力とし、 $t = 2$ 時点の実況解説文を生成する。その後、 $t = 1$ および $t = 2$ 時点の画像、実況・解説文を入力とし、 $t = 3$ 時点の実況解説文を生成する。この処理を試合終了まで繰り返すことで、連続した実況・解説文を生成する。本手法の利点は、前フレームの情報と生成文を含めることで、より継続性のある実況・解説文を生成することが期待できる点にある。なお、 $t = 0$ フレームについては前のフレームの生成文を用いず、システムプロンプトのみを用いて文章を生成

し、 $t = 1$ フレームについては $t = 0$ フレームの生成文一つのみを用いて文章を生成する。

さらに、生成した文章を、別のプロンプトを用いてさらにモデルで生成することで、最終的な実況・解説文を得る。

3 評価・分析

本節では、提案手法を実際の試合映像に適用して実況・解説文の生成を行い、その品質の評価や特徴の分析を行う。また、生成文の正確性を人手で評価することで、システムの性能を評価する。使用した MLLM は GPT-4v および Claude Sonnet である。

評価に使用する映像は、RoboCup 2023 の決勝戦における B-Human 対 HTWK Robots のサッカー標準プラットフォームリーグの試合¹の最初の 1 分間を選定した。画像分割のインターバルは 5 秒である。用いたシステムプロンプトを表 1 に示す。また、GPT-4v によって生成された文章を表 2 に、Claude Sonnet[12] によって生成された文章を表 3 にそれぞれ示す。GPT-4v, Claude Sonnet の使用は API を用いて行った。一つの実況・解説文の生成には GPT-4v, Claude Sonnet ともにおよそ 20 秒かかった。

生成文の評価のために、面白さ (字幕の文章は試合を観戦するうえで面白みがあるか)、明瞭さ (字幕の文章は試合を観戦するうえでわかりやすいか)、および適切さ (字幕は試合の状況を適切に表現できているか) という 3 つの観点で大学生 6 名に生成した実況・解説文を字幕として動画に適宜挿入した映像を見てもらってアンケートを実施した。GPT-4v と Claude Sonnet によるアンケート結果を表 2 と 3 に示す。

GPT-4v, Claude Sonnet ともに、面白さ、明瞭さ、適切さすべての項目に対して半数以上がとても当てはまるもしくはやや当てはまると回答した。

¹<https://www.youtube.com/watch?v=5JhxYgBN6AA>

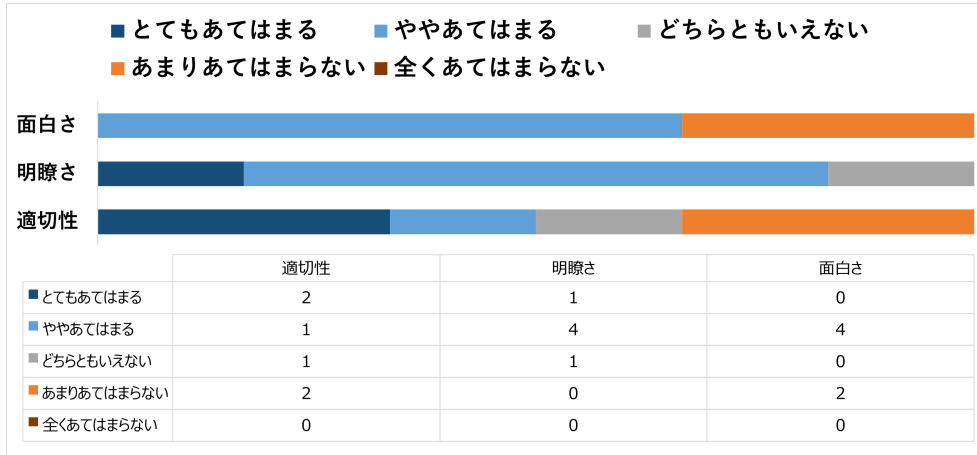


図 2: GPT-4v によるアンケート結果

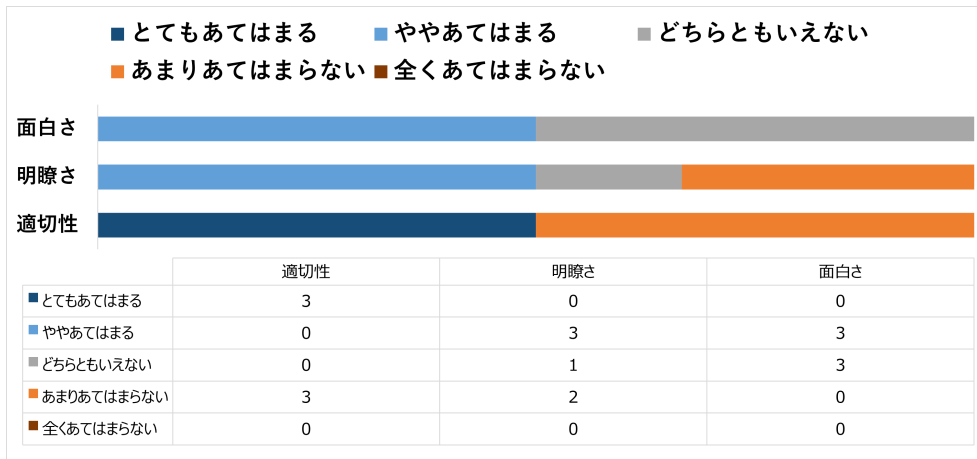


図 3: Claude Sonnet によるアンケート結果

4 まとめ

本研究では、ロボカップサッカーの標準プラットフォームリーグにおいて、MLLMを用いて自動で試合の実況・解説を行うシステムを提案した。また、提案したシステムによって生成された実況・解説文章に対して面白さ、明瞭さ、適切さの観点からアンケートを行い、いずれの項目においても、肯定的な結果が半数以上得られた。

今後の課題について、現在の生成文は、一般的な説明的な文体に偏っており、口調や表現が自然ではない。特に、スポーツ実況のような出来事を伝える文には、感嘆詞などを用いたより自然な表現が必要であると考えられる。このため、言語モデルの学習データの拡充や、より自然な表現を推定するためのファインチューニングなどによるモデルの改善が必要である。また、実況文章を音声で出力する場合、適切な長さの文が必要であり、生成された文が長すぎる場合や短すぎる場合には理解しにくい文章となる。そのため、文章の長さの制御を行

う方法が必要である。さらに、生成された実況文の正確性を高めるためには、画像に対する認識精度を高める必要がある。現在の認識技術では、認識精度が不十分な可能性がある。最後に、推論速度についても課題が存在する。実況文をリアルタイムで生成するためには、モデルの推論速度が重要である。現在のモデルでは、リアルタイムで実況を行うほど高速に推論することができない。

参考文献

- [1] 野田 五十樹, 南方 英明, 小林 邦和, 杉浦 藤虎, 武村 泰範, 秋山 英久, 岡田 浩之, ロボカップ西暦 2050 年を目指して (その 1), 知能と情報, Vol.29, No.1, pp.2-13, 2017.
- [2] 奥川 雅之, 伊藤 暢浩, 岡田 浩之, 植村 渉, 高橋 友一, 杉浦 孔明, ロボカップ西暦 2050 年を

表 1: システムプロンプト

プロンプト	内容
文章生成	あなたはロボットサッカーを実況・解説を行う AI です。5 秒ごとに試合の画像が与えられるので、その間の実況・解説の文章を生成してください。文章はボールの動きに注目して簡潔に出力してください。なお、画面下部に残り時間及びチームの得点が表示されています。
文章変形	“プロンプト 1 で生成した文章 “ 上記の文章を実際の実況のように変換してください。文章は 5 秒で発話できるぐらい短くしてください。文章だけを出力してください。

- 目指して (その 2), 知能と情報, Vol.29, No.2, pp.42-54, 2017.
- [3] 浅田 稔, ロボカップサッカーにおける科学技術チャレンジ, 日本ロボット学会誌, Vol.38, No.4, pp.323-330, 2020.
- [4] K. Tanaka, H. Nakashima, I. Noda, K. Hasida, I. Frank and H. Matsubara, “MIKE: an automatic commentary system for soccer,” Proceedings International Conference on Multi Agent Systems, pp. 285-292, 1998.
- [5] E. André, K. Binsted, K. Tanaka-Ishii, S. Luke, G. Herzog and T. Rist, “Three RoboCup Simulation League Commentator Systems,” AI Magazine, Vol.21, No.1, pp.57-66, 2000.
- [6] 大橋 玲音, 坪倉 和哉, 小林 邦和. ロボカップサッカーにおける実況・解説ロボットシステムの提案. 第 62 回 人工知能学会 AI チャレンジ研究会, No.Sig-Challenge-062-03, pp.11-14, 2023.
- [7] BROWN, Tom, et al. Language models are few-shot learners. Advances in neural information processing systems, 33: 1877-1901, 2020.
- [8] OUYANG, Long, et al. Training language models to follow instructions with human feedback. Advances in neural information processing systems, 35: 27730-27744, 2022.
- [9] DEVLIN, Jacob, et al. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.
- [10] GPT-4 Technical Report, <https://cdn.openai.com/papers/gpt-4.pdf>, (cited-2024-04-09).
- [11] TEAM, Gemini, et al. Gemini: a family of highly capable multimodal models. arXiv preprint arXiv:2312.11805, 2023.
- [12] The Claude 3 Model Family: Opus, Sonnet, Haiku. <https://www.anthropic.com/news/claude-3-family>. (cited 2024-04-09)

表 2: GPT-4v による生成文

時間	生成文
0:00	キックオフ！B-Human がどんな戦略を見せるか、注目です！
0:05	キックオフから数秒、B-Human がボールを前進！ポジショニングが鍵を握る。HTWK Robots、防御態勢で対応！
0:10	B-Human が攻め込む！HTWK の守備を破れるか!? まだ決定機なし！ロボットたちが戦略的に動く！
0:15	B-Human がボール支配、HTWK が堅い守備！激しい攻防、次の得点チャンスが勝負の分かれ目！
0:20	HTWK が攻撃へ！B-Human は守備を固めるが、次の一手が勝負を決めるかも！
0:25	HTWK が攻撃！B-Human のディフェンスは堅い！ボール争奪戦が中央で続く！どちらがコントロールを握るか！
0:30	HTWK が攻めるも B-Human 堅守！ボールは中央で膠着、次の一手に期待！
0:35	HTWK が攻めるも B-Human 堅守！カウンター狙う！次の一手が勝負だ！
0:40	HTWK が攻めるも B-Human 堅守！ボール中央、両チーム得点狙う！
0:45	HTWK が前進、B-Human のゴール目前！ディフェンスを突破なるか？左サイドから攻撃激化！状況変わるかも！
0:50	プレイ再開間近！B-Human のキックオフ、HTWK Robots の対応は？攻撃の流れを掴めるか!?
0:55	HTWK Robots のキックオフ！攻撃のチャンスを狙う！B-Human は堅守！プレイ再開、どちらが優位か!?

表 3: Claude Sonnet による生成文

時間	生成文
0:00	試合開始間もなく、ボールは青チームのロボット側にあり、赤チームがディフェンスに構えています。両チーム相手を注視し、次の展開を待っています。
0:05	キックオフで赤チームがスタート。両チームのロボットが動き出しポジショニングを変える。青ロボットがボールを持ち前線で攻撃を狙う。互いにチャンスを窺う展開に。
0:10	ボールは青チームの手に、赤チームがディフェンスで包囲しています。青チームはパスを狙うがスペースが作れず、両チームが攻防を繰り返しています。
0:15	青のロボットがボール保持、赤チームがプレスをかけるも青はパスを通す。中盤でボールキープを続け、次の機会を窺う緊迫した攻防が続いている。
0:20	混戦状況が続く中で、ボール奪取の激しい攻防が交わされている。
0:25	青チームがボール確保、攻撃機会を狙う。ロボット達が機敏にポジショニングを変え、前線へパス！赤チームの守備陣が食い止めに動くが、青チームはシュートを狙う。両チームの緊迫した動きが場内の空気を引き締める。
0:30	ゴール前で青チームが攻めています。パスを回しながらシュート狙っています。赤チームは守備を固め必死です。双方のロボットがぴったりとマークし合い、得点の瞬間を窺っています。
0:35	ボールは中盤に、両チームが新たな構えへ。青は攻勢をにらみ、赤は守りを固めカウンター狙う。
0:40	青チームの素晴らしいパスワークによる猛攻が赤チームの守りを崩そうとしています。中盤でボールを奪い合う熱い攻防が続く、いつ形勢が変わるかわかりません。青チームの猛攻に注目です！
0:45	ミドルでボールがこちら行ったり行ったり、双方ロボットが押し込む形に。青チーム数的優位も赤チームの守備がタフ。両者ボール支配権を争い激しく当たり合い、白熱した攻防が続く。隙を突けば展開が動きそう。
0:50	ゴール前で熱い攻防が続く。青チームシュートに狙うが、赤ディフェンスがしのぐ。しかしキックオフで青に新たな攻撃のチャンス。試合はいつ動くかわからない混戦模様だ。
0:55	ムードル地帯でボールがかろうじて保たれている。青チームのロボットがパス回しで地域伸展をねらうが、赤チームの集中マーキングで前進が阻まれている。