

1. 担当 PM

五十嵐 悠紀

(明治大学 総合数理学部 先端メディアサイエンス学科 専任准教授)

2. クリエータ氏名

黒田 和矢 (静岡大学 大学院総合科学技術研究科 情報学専攻)

3. 委託金支払額

2,162,000 円

4. テーマ名

深層学習による AI 実況プレイ動画生成

5. 関連 Web サイト

- YouTube
<https://www.youtube.com/channel/UC3ggg0ryKhUMSFaiCxKAwbw>
- ニコニコ動画
<https://www.nicovideo.jp/user/87155387/top>

6. テーマ概要

本プロジェクトは「マリオカート 8 デラックス」のプレイ動画を入力として実況プレイ動画を生成するシステムの開発を行った。システムは以下のステップに従い、入力されたプレイ動画をあたかも自分がプレイしているかのように実況する。

- (1) 画像認識／動作認識／物体検出を用いた画面認識
- (2) 画面認識と生体情報を用いた感情表現
- (3) 画面認識と感情表現を用いた実況内容の生成

7. 採択理由

ゲームは従来「プレイするもの」であったが、最近では「見て楽しむもの」と

いった娯楽の在り方も増えてきており、スポーツ観戦や囲碁・将棋などの実況と同様に、ゲームの実況動画も人気を集めている。本提案システムでは「こういうリアクションを取ったら楽しいのではないか」といったことをAIが考えながら実況する点で面白さがある。リアルタイムでイベント発生が多々起きる中で、重要度に応じた処理をどのように行うか、など、課題は多々あるが、開発に対する意欲も実力も十分であった。

開発する技術は、ゲームだけでなく、少年サッカーをはじめとした、現状では実況がつかないスポーツ観戦などにAI実況がつけられる可能性を秘めている。類似研究の中でどのように独自性を出していけるか期待した。

8. 開発目標

本プロジェクトでは、人間がプレイした動画を入力として、

- (1) 物体検出，動作認識
- (2) 感情の想起
- (3) 実況文の生成

のステップに従い実況文を自動で生成し、読み上げ用音声合成ソフトウェアを用いてそれを読み上げることを行うシステムの開発を目標とした。

深層学習による物体検出，動作認識を用いてゲーム中の出来事を認識する。例えば、レースゲームのプレイ動画を入力すると、物体検出では、レース中のキャラクターや自身の順位などを検出することが目標である。動作認識では、プレイヤーが操作するキャラクターの、ドリフトしている、加速しているといった動作を認識する。また、システムは人間の感情を模倣し、あたかもゲームをプレイしているかのように振る舞う。実況を行う際には、ゲーム内で起こっている出来事を伝えるだけでなく、リアクションなどを通じてエンターテインメント性のある実況を実現する。

9. 進捗概要

本プロジェクトでは人間がマリオカート 8 デラックスをプレイした動画を入力として、図 1 に示すような実況プレイ動画を生成するシステムを開発した。本システムにおけるゲーム実況は宮咲ふわら（図 1 右下）が担当する。システムは大きく分けて以下の 3 つのステップに従い実況プレイ動画を生成する。

- (1) 画像認識／動作認識／物体検出を用いた画面認識
- (2) 画面認識と生体情報を用いた感情表現
- (3) 画面認識と感情表現を用いた実況内容の生成



図 1. 実況プレイ動画

画面認識

システムは以下の 3 種類の深層学習のモデルを用いてゲーム画面中の情報を認識する。本プロジェクトでは各種モデル及び学習データ、図 2 に示す学習データの作成を支援するシステムの開発を行った。

- (1) 画像認識
- (2) 動作認識
- (3) 物体検出

画像認識は CNN（Convolutional Neural Network）というモデルを使って順位、アイテムの種類、コインの枚数、ラップ数、プレイヤーが操作するキャラクターの状態（無敵、透明等）を認識する。具体的には、順位を例にとると図 3 のようにゲーム画面上の順位が表示されている箇所について 1 位から 12 位のどの順位が表示されているのか推測する。

動作認識は LRCN（Long-term Recurrent Convolutional Network）というモデルを使ってプレイヤーが操作するキャラクターの動作を認識する。具体的には、図 4 のようなゲーム中の連続画像をもとにどのような動作が行われているか推測する。

物体検出は SSD（Single Shot Multibox Detector）というモデルを使ってコース上に存在するアイテム、アイテムボックス、敵キャラクターを検出する。具体的には、図 5 のようにゲーム画面全体に対してどの座標に何が表示されているか検出する。



図 2. 学習データ作成支援システム



図 3. 画像認識（順位）の例

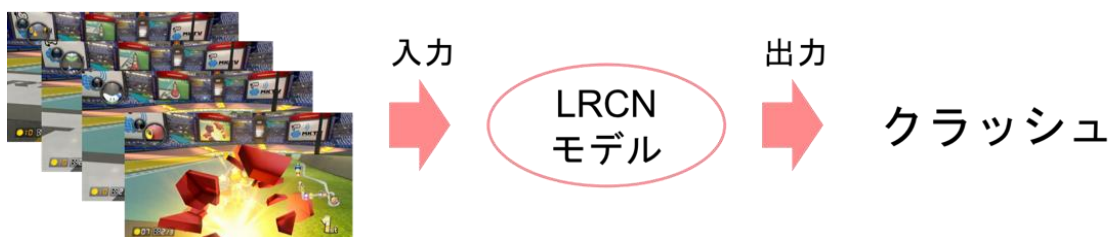


図 4. 動作認識の例



図 5. 物体検出の例

感情認識

システムは図 6 に示すモデルをもとに感情の表現を行う。モデルは快-不快を表す横軸と眠気-覚醒を表す縦軸から成っており、喜怒哀楽の 4 種類の感情を表現することができる。システムは画面認識の結果を用いて「快-不快」の値、ゲームのプレイ中に検出した心拍数から「眠気-覚醒」の値をそれぞれ決定し、実況プレイ動画中で図 7 に示す表情を以て感情を表現する。

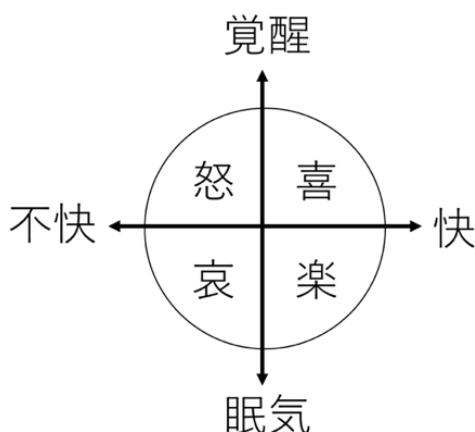


図 6. 感情表現モデル



図 7. 宮咲ふわら各種表情

実況内容の生成

システムは画面認識と感情認識で得られた結果をもとに、あらかじめ収集したリアクションを使用して実況を行う。リアクションの収集について、はじめに動画投稿サイトに投稿されているマリオカート 8 デラックスの実況プレイ動画から、人間がゲーム中のどのようなタイミングでどのようなリアクションを取っているのか分析を行った。また、RNN (Recurrent Neural Network) というモデルを使って分析したリアクションを学習し、分析した実況プレイ動画中に存在しない新しいリアクションを生成した。さらに、自動生成したリアクションについてどのようなタイミングで取ることができるか自動分類した。一連の処理の例を図 8 に示す。

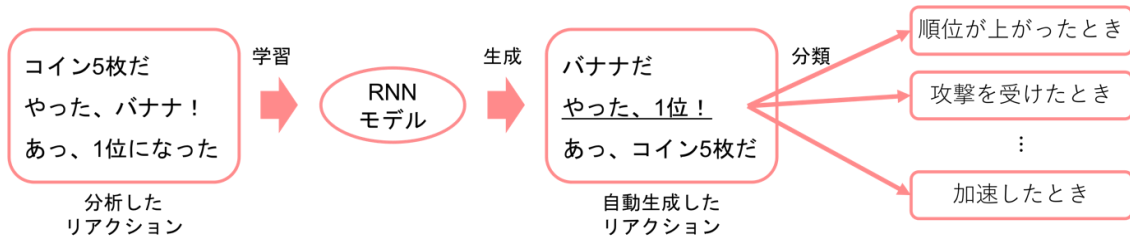


図 8. リアクション自動生成の例

クリエイターは合宿やミーティングが終わる都度、もらった意見とそれに対する自分の回答、考えられる開発目標とその中でも今回注力して行う具体的な開発事項、といった内容を PM と Slack で共有し、開発を進めることができた。

10. プロジェクト評価

本プロジェクトでは深層学習を用いてゲーム実況を行う AI「宮咲ふわら」を開発した。これまでにゲーム動画を入力として AI でゲーム実況を行うものではなく、未踏性は非常に高いと言える。

画像認識・動作認識・物体検出を用いて画面を認識する技術、画面認識と生体情報を用いた感情表現、画面認識と感情表現を用いた実況内容生成の 3 本柱で取り組み、どれも完成度の高い実用的なシステムが完成した。あらかじめゲーム制作者がコンテンツを用意するのではなく、プレイ動画からすべて全自動でゲーム実況を行うことができる技術を作成したことで、学習モデルの構築さえできれば今後さまざまなゲーム動画を対象としたゲーム実況や、e スポーツなどの競技の実況解説を行うことも可能になる。このように、動画実況の在り方を切り拓く技術を作成した点を評価する。

11. 今後の課題

現在マリオカートに特化したものではあるが、YouTube やニコニコ動画への動画投稿を通して本システムの周知に努めている。今後同じ技術を他のゲームや e スポーツなどに適用するなど、システムの改良を行うことで汎用性を示して欲しい。また手軽に第三者が使えるような仕組みも作り上げシステムを提供するなど、AI 実況自体を第三者が使いこなせるようなシステムになることを期待する。