

深層学習による AI 実況プレイ動画生成 — ゲーム実況 AI: 宮咲ふわら —

1. 背景

近年ゲーム実況の市場が大きく発展を遂げている。ゲーム実況のスタイルには、人間が実況を行うスタイルに加え、AquesTalk や VOICEROID といった合成音声を用いたスタイルやバーチャル YouTuber と呼ばれる 3DCG を実況キャラクターとするスタイルが存在する。このように声やキャラクターが人間からシステムに置き換わっていく中、実況を行うのは常に人間である。

2. 目的

本プロジェクトはシステムが人間に変わってゲーム実況を行うことを目的として、ゲーム実況を自動で行う AI を開発した。

3. 開発の内容

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

3.1. 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



図 2 学習データ作成支援システム

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。



図 3 画像認識(順位)の例

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。



図 4 動作認識の例

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 5 物体検出の例

3.2. 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。

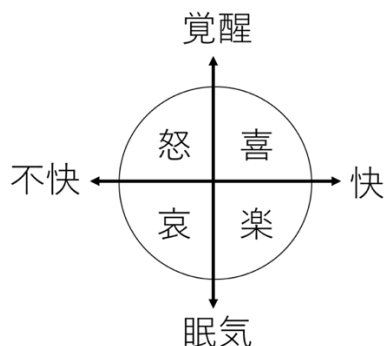


図 6 感情表現モデル



図 7 宮咲ふわら各種表情

3.3. 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。

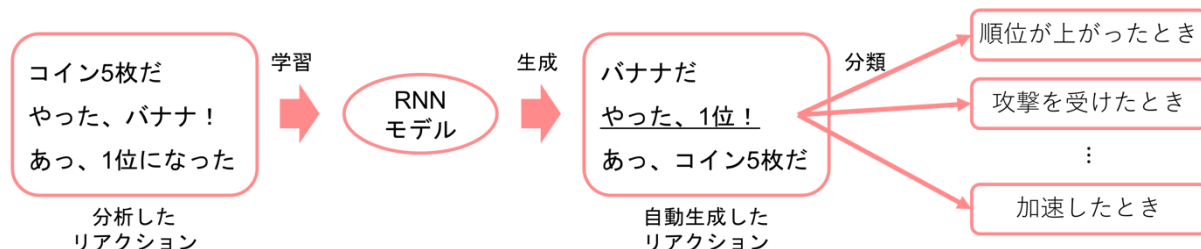


図 8 リアクション自動生成の例

4. 従来の技術(または機能)との相違

これまでゲーム実況は人間だけが行うものであったが、本プロジェクトを通してシステムがゲーム実況を行うことが可能となった。一方、実況という観点においてはスポーツの実況解説としての実況機能が組み込まれているゲームが既に存在する。しかし、それらのゲームはゲーム内部のパラメータをもとにルールベースでの実況を行っている。本プロジェクトで作成したシステムは深層学習を用いて画面上の情報を通して実況を行っているため、ゲームのみならず現実のスポーツの実況にも適応することが可能となる点において相違がある。

5. 期待される効果

本システムは人間がゲームをプレイするだけで実況プレイ動画を作成することを可能とした。これにより誰でも低コストで実況プレイ動画を作成することが可能となり、システムを通して今まで以上のエンターテインメントを創出することが期待される。また、本システムは一人称視点でゲーム実況を行うが、発話内容を三人称視点にするといった変更を加えることでゲーム実況のみならず e スポーツ等の競技の実況解説を行うことが期待される。

6. 普及(または活用)の見通し

現在 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。今後も動画投稿の継続やシステムの改良を繰り返すことで、より多くの人にエンターテインメントを届けるとともにゲーム実況 AI の新規開発者が現れることを目指す。

7. クリエータ名(所属)

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)

(参考)関連 URL

YouTube: <https://www.youtube.com/channel/UC3ggg0ryKhUMSFaiCxKAwbw>

ニコニコ動画: <https://www.nicovideo.jp/user/87155387/top>