

# シミュレーションによる AI学習用データ作りの研究

佐々木 弘隆

人工知能、機械学習を試している方の多くが、学習用データの取得(生成)に苦慮しているようです。ここではPythonの良さを体感していただく事例として、機械学習のための学習データを作る方法を紹介します。

## データ構造を決める

### ●データの種類

前章ではPythonでリアルタイム制御によるゲームを作成しました。このゲームからデータを取り出していきます。ゲームに含まれる情報の系統を分析しますと、

- 1: プレーヤ
- 2: キャラクタ
- 3: ノンプレーヤ・キャラクタ
- 4: ステータス

に分類できます。

#### ▶プレーヤ

プレーヤとは、ゲームをプレイ(遊ぶ)している人間のことです。プレーヤはゲームの進行に合わせて何らかの要求をゲームに伝えます。このゲームの場合は「スペース・キーを任意のタイミングで押す」になります。

#### ▶キャラクタ

キャラクタとはプレーヤの身代わりとしてゲーム中に存在するものになります。ゲームのルールに従った行動が許可され、プレーヤの要求に従って行動を起こす存在です。

#### ▶ノンプレーヤ・キャラクタ

ノンプレーヤ・キャラクタとは、プレーヤとの関係を持たない存在になります。ゲームのルールに従った行動をします。与えられた役割によってはキャラクタや他のノンプレーヤ・キャラクタに影響を及ぼすこともあります。

#### <このゲームのノンプレーヤ・キャラクタ>

- 車: 移動を行いキャラクタとの接触で事故を起こせる
- 地面: 存在するだけで何もしない

#### ▶ステータス

ゲームのステータス(状態)を表すあらゆる情報の集合です。このゲームの情報は以下です。

- 1: 時間
- 2: 横断成功
- 3: 横断成功数
- 4: 横断失敗
- 5: 横断失敗数

### ●構造

文章の機械学習のデータの場合は小説などのテキスト・データで、画像の機械学習データでしたら大量の画像データになります。しかし、リアルタイム制御の場合は「タイム」という言葉が入っている通り、時間軸によるデータ群が連続して存在します(図1)。

このゲームの場合、0秒のプレーヤ→キャラクタ→ノンプレーヤ・キャラクタ→ステータス→0.016秒のプレーヤ→…と続くということです。

### ●出力形式

時間ごとの一連のデータが必要なのは確定しました。次にこのデータの出力先ですが、分かりやすく多くの環境で利用しやすいCSVファイルを使っています。

#### ▶CSVファイル

CSVファイルとは、Comma Separated Valuesの略で、カンマ「,」記号でデータが区切られたテキスト・ファイルです。

テキスト・ファイルなので、多くのテキスト・エディタで編集ができますし、Excelなどの表計算ソフトウェアで扱うことができるデータ構造です。さらにPythonでは、CSVファイルを読み込んで分析することが得意なライブラリが用意されているのもうれしいところです。

#### ▶横方向の項目

CSVは表計算ソフトウェアで使えることから分かる通り、データが横と縦に並べられる構造をとっています。ゲームの実行時間に従い縦に追加するようにし