

# Deep Learning を用いた 競馬の勝ち馬予測プログラムの作成

谷聖一 研究室 宮脇 祐太  
Yuta Miyawaki

## 概要

近年、様々な分野で機械学習技術の 1 つであるニューラルネットワークを用いた Deep Learning が活用されている。本演習では、Deep Learning を用いて競馬の勝ち馬を予測するプログラムを作成した。

## 1 はじめに

### 1.1 競馬とは

競馬とは、大辞林によると、「競走馬に一定の距離を走らせ順位を競う競技。また、その勝馬や着順などを当てる賭け」([1]) である。競馬において「金を賭ける」とは、勝馬投票券 (通称、馬券) を購入することを言い、予測が的中したものには配当が支払われる。また、本演習における「競馬」は、日本中央競馬会 (JRA) が主催する「中央競馬」とする。

### 1.2 ディープラーニングとは

ディープラーニングとは「深い、すなわち、多くの層を持ったニューラルネットワークを用いた機械学習の総称」([2]) である。多層のニューラルネットワークを用いることでより多くの特徴を抽出することが可能である。現在、第 3 次人工知能ブームと呼ばれ様々な分野からディープラーニング (Deep Learning) が注目されている。ディープラーニングが話題になる背景には、コンピューターが層の深い DNN (Deep Neural Network) を学習できる能力を得た点とインターネットの普及により大量のデータが出現した点にある。

### 1.3 ニューラルネットワークとは

ニューラルネットワークとは「脳神経系をモデルにした情報処理システムのこと」([3]) である。ニューラルネットワークは入力層・中間層・出力層からなる。特に、中間層を複数持つニューラルネットワークモデルを DNN (Deep Neural Network) と呼ぶ。

### 1.4 背景・動機

[4] によると、ディープラーニング技術は画像認識・自動運転・ゲーム・株取引など様々な分野で活用され始めている。本演習では、このディープラーニング技術を用いて競馬の勝ち馬予測を行い利益を得られる予測モデル

を作成することにした。

### 1.5 演習目的・内容

Web ページから取得した情報を学習データとし、ディープラーニングを用いて競馬予測を行う DNN モデルを作成する。3 着内に来る馬とそうでない馬の分類を行う「3 着内予測モデル」、高い配当が得られる馬を予測する「配当期待値予測モデル」またそれぞれに対して「芝コース」、「ダートコース」専用の合計 4 種類を作成する。

## 2 演習方法

### 2.1 実行・開発環境

本演習における実行環境を表 1 に示す。

表 1: 実行環境

	MacBook Pro	Mac mini
モデル	Early2013	Mid 2011
CPU	Intel Core I5	Intel Core I7
メモリ	8GB	8GB

開発環境は以下の通りである。

- 使用言語
  - Python 3.5
- ライブラリ
  - Keras
  - Pandas
  - numpy
  - Beautiful Soup
  - Selenium
  - scikit-learn
  - fastText

## 2.2 学習データの収集

本演習では、競馬情報サイト「netkeiba」から Web スクレイピングを行いデータの収集を行う。2010 年 1 月から 2017 年 9 月までの新馬戦を除いた芝およびダートレースに出走した馬の以下のデータを取得した。取得したデータ数を表 2 に示す。

表 2: 取得したデータ数

芝	ダート	合計
122255 件	133552 件	253785 件

- レースデータ

「芝ダート」「右左周り」「距離」「競馬場」「開催何日目」「何レース」「馬場状態」「1 着賞金」

- 馬の情報

「枠番」「着順」「オッズ」「牝馬牡馬」「馬齢」「斤量」「騎手」「体重」「体重前走比」「所属」「脚質」「休み明け」「調教コース」「調教脚色」「調教評価」「1 走前から 5 走前着差・着順・上がりタイム・芝ダート・距離・グレード」「血統」

## 2.3 学習データの前処理・整形

収集したデータをディープラーニングの入力として扱えるようにするため、数値データに変換する処理を行う。

### 1. 文字列データの One-hot 値化

以下のデータに対して One-hot 値化を行う：「右左周り」「競馬場」「馬場状態」「牝馬牡馬」「騎手」「所属」「調教コース」「調教脚色」「1 走前から 5 走前芝ダート・グレード」

### 2. 血統図のベクトル化

[5] の手法を参考に、3 世代前までの血統図を 100 次元の数値ベクトルへの変換を以下の手順で行う。

- (a) 2010 年 1 月から 2017 年 12 月までのレースに出走した馬の血統図を空白区切り 1 行で表現したテキスト形式で取得する。
- (b) 子と親子の距離を近くするため、子を中心に親が左右に広がる様並べ替えを行う。
- (c) 並び替えを行なったテキストを入力として、fastText を用いて各馬 100 次元の数値ベクトルを得る。

### 3. 数値データの標準化

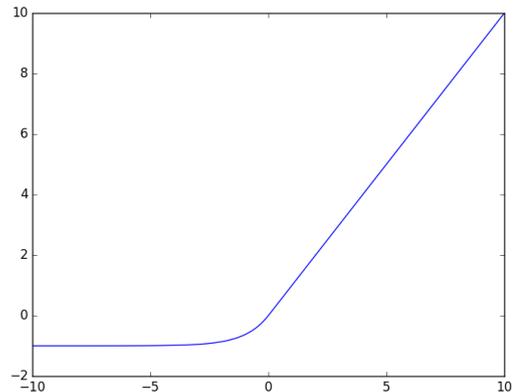
規格の揃わない数値データを学習の入力に用いることは適切ではないため、それぞれの特徴量が平均 0 分散 1 に分布するよう変換を行う。

### 4. 3 着外データ数の調整

機械学習において、学習データのラベルの偏りが小さいことが望ましい。競馬では、3 着外の馬が 3 着内の馬と比較してデータ数が大きくなってしまいうため、3 着外の馬のデータをランダムにドロップし数の調整を行う。

## 2.4 DNN モデルの構築

本演習では、800 次元の中間層を 5 層持つ DNN モデルを用いる。また、各中間層に対して Batch Normalization を適用する。活性化関数には「elu」(図 1)を用いる。



$$x = x(x \geq 0)$$

$$x = e^x - 1(x < 0)$$

図 1: elu

出力層はそれぞれ以下の通りである。

- 3 着内予測モデル  
出力層 2 層で [3 着内, 3 着外] の 2 値。
- 配当期待値予測モデル  
出力層 1 層で配当の期待値。

## 3 結果考察

### 3.1 実験

作成したモデルを用いて、2017 年 10 月から 2018 年 1 月までのレースの予測を行的中率・回収率の算出および購入のシミュレーションを行う。購入のシミュレーションは、初期資金 10 万円で 1000 円づつ勝馬投票券を

購入するとする。購入する勝馬投票券は「複勝式」とする。各モデルの回収率および的中率は表 3 に、購入のシミュレーションを図 2 から図 5 に示す。

表 3: 予測モデルの回収率および的中率

モデル	条件	回収率	的中率
3着内予測モデル	全レース購入	85%	49%
3着内予測モデル	出力最大と 2 番目の馬の値の差が 0.1 以上のレースのみ購入	80%	55%
配当期待値予測モデル	全レース購入	116%	26%
配当期待値予測モデル	出力最大と 2 番目の馬の値の差が 10 以上のレースのみ購入	140%	23%

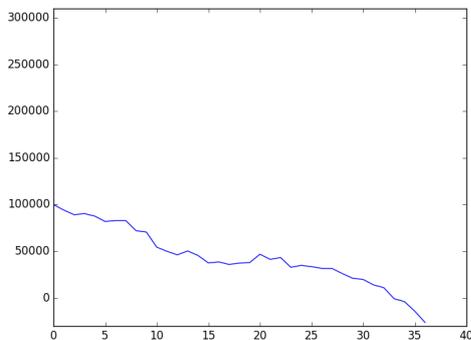


図 2: 3着内予測モデル全レース購入

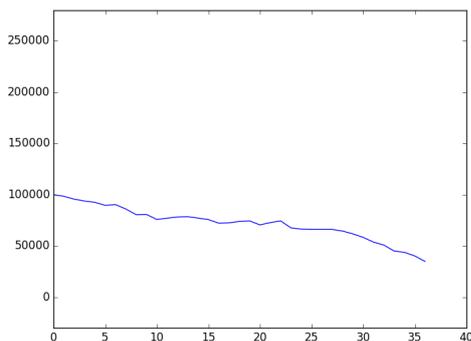


図 3: 3着内予測モデルの出力最大と 2 番目の馬の値の差が 0.1 以上のレースのみ購入

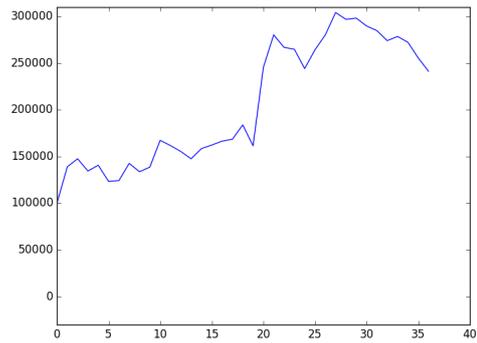


図 4: 配当期待値予測モデル全レース購入

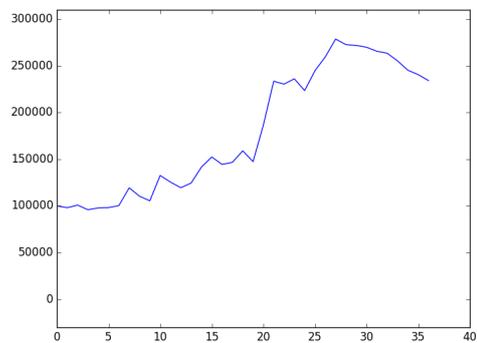


図 5: 配当期待値予測モデルの出力最大と 2 番目の馬の値の差が 10 以上のレースのみ購入

### 3.2 考察

表 3 に示した通り、3 着内予測モデルに関しては高い的中率を得ることができている。また、配当期待値予測モデルに関しては回収率で 100%を越す結果が得られた。

### 3.3 今後の課題

本演習における今後の課題は 2 点あげられる。1 点目は、本演習では本来用いるべきでない主観的なデータである「調教脚色」「調教評価」を用いているため、これらのデータを客観的なデータへ置き換えを行うことである。2 点目は、更新された新しいレース情報を自動的に取得し再度学習し、新しく作られたモデルを用いて自動的に予測するプログラムを作成することである。

## 参考文献

- [1] 「大辞林」三省堂
- [2] 「深層学習」第 4 版, 人工知能学会監修
- [3] 村上研究室 ニューラルネットワーク - 画像処理・理解研究室 - 愛媛大学 (参照日時:2018/2/12)  
<http://ipr20.cs.ehime-u.ac.jp/column/neural/index.html>

- [4] DeepAge 「AI ビジネスを考える上で押さえておきたい、ニューラルネットワークの実用例 30 選」 (参照日時:2018/2/13)  
[https://deepage.net/machine\\_learning/2016/08/30/nn\\_examples.html](https://deepage.net/machine_learning/2016/08/30/nn_examples.html)
- [5] fastText を用いた競馬血統のベクトル化 (参照日時:2018/1/28)  
<https://qiita.com/KTaskn/items/e76551191214593c278e>