

深層学習を用いた打球予測

深井 亮登† 小林 正弘‡ 中川 智之† 田畑 耕治† 松澤 智史†

東京理科大学 理工学部 情報科学科†
東海大学 理学部 情報数理科‡

1. 研究背景

近年、各分野でデータ分析が盛んに行われている [1]。NPB(日本野球機構)では、打者の打球傾向や、試合状況による守備シフトがセオリーの一つとして存在する。

しかし、使用されている守備シフトは経験に基づくことが多い。そこで、深層学習を用いることで守備シフトの陣形を考える際の一つの手段として提示する。

2. 研究の目的

本研究では、深層学習を用いて野球の試合の局面や選手から打球の落下点または処理する位置や種別、強度を予測するシステムを提案する。

3. 基礎知識

3.1 ニューラルネットワーク

ニューラルネットワークとは、人間の脳の神経回路構造を模した計算モデル [2] であり、複数のニューロンモデルを結合することで構成される。ニューロンモデルは、生体ニューロンの動作を簡単な数式でモデル化したものである。また、多層のニューラルネットワークを用いた学習方法を深層学習という。

3.2 F 値

F 値 (F-score) とは、再現率 (Recall) と適合率 (Precision) の二つの評価指標を組み合わせた指標である。予測結果と実際の答えの対応を表 1 に 3 つの指標の導出式を式 1、2、3 に示す。

表 1: 予測結果と実際の答えの対応表

		予測結果			
		1	2	...	n
実際の答え	1	T_1	F_{21}	...	F_{n1}
	2	F_{12}	T_2	...	F_{n2}

	n	F_{1n}	F_{2n}	...	T_n

$$Precision(k) = \frac{T_k}{T_k + \sum_{i=1}^n F_{ki}} \quad (1)$$

$$Recall(k) = \frac{T_k}{T_k + \sum_{i=1}^n F_{ik}} \quad (2)$$

$$F-score(k) = 2 \frac{Precision(k) \cdot Recall(k)}{Precision(k) + Recall(k)} \quad (3)$$

4. 提案手法

4.1 概要

本研究では、一打席における四つの予測をそれぞれ四つのニューラルネットワークを用いて行う。

- 打球方向の区域を予測するニューラルネットワーク (打球方向 NN)
ここではレフトを左、センターを中、ライトを右とする。区域に関しては、図 1 に示す。
- 打球の強度 (無、弱、強) を予測するニューラルネットワーク (打球強度 NN)
無は、バントしたときの打球の強度である。
- 打球の種別 (ゴロ、フライ) を予測するニューラルネットワーク (打球種別 NN)
- 打球の区域距離 (内野、外野) を予測するニューラルネットワーク (打球区域距離 NN)

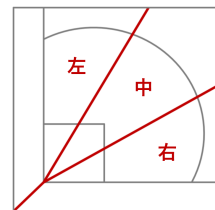


図 1: 打球方向の区域

これら四つの NN の評価指標は F 値を用いる。

4.2 データセット

本研究ではデータスタジアム様から頂いた日本プロ野球の公式試合データ (2016、2017、2018 年) を用いる。

5. 実験

5.1 パラメータ

方向 NN、強度 NN のパラメータを表 2、種別 NN、区域距離 NN のパラメータを表 3 に示す。

5.2 結果

5.2.1 打球方向

打球方向 NN の適合率、再現率、F 値を表 4 に示す。3 値の打球方向の予測では、左方向と右方向が中方向に比べ性能が高かった。テストデータの (左:中:右) 比は 1:1:1 であるので、この結果から中方向に打球が飛ぶ場面が特定できない。

The hit prediction on deep learning
†Akito Fukai, ‡Masahiro Kobayashi, †Tomoyuki Nakagawa,
†Kouji Tahata, †Tomofumi Matsuzawa
†Faculty of Science and Technology, Dept. of Information Sciences, Tokyo University of Science
‡Faculty of Science, Dept. of Mathematical Sciences, Tokai University

表 2: 方向、強度 NN の詳細パラメータ

パラメータ	方向 NN	強度 NN
NN	全結合	全結合
訓練データ	118240	
テストデータ	13142	
活性化関数	ReLU	
損失関数	categorical_crossentropy	
出力関数	softmax	
最適化手法	Adam	
入力層のユニット数	56	56
出力層のユニット数	3	3
隠れ層の数	5層	5層
学習率	0.001	
エポック数	50	50

表 3: 種別、区域距離 NN の詳細パラメータ

パラメータ	種別 NN	区域距離 NN
NN	全結合	全結合
訓練データ	118240	
テストデータ	13142	
活性化関数	ReLU	
損失関数	binary_crossentropy	
出力関数	softmax	
最適化手法	Adam	
入力層のユニット数	56	56
出力層のユニット数	2	2
隠れ層の数	5層	4層
学習率	0.001	
エポック数	50	50

表 4: 打球方向の適合率, 再現率, F 値

	適合率	再現率	F 値
左	0.49	0.62	0.55
中	0.36	0.17	0.23
右	0.47	0.56	0.51
平均	0.44	0.45	0.43

5.3 打球強度

打球強度 NN の適合率、再現率、F 値を表 5 に示す。

表 5: 打球強度の適合率, 再現率, F 値

	適合率	再現率	F 値
無	0.72	0.75	0.74
弱	0.60	0.50	0.55
強	0.61	0.69	0.65
平均	0.65	0.65	0.64

強度が無では高い評価のため、特定の試合状況である。

5.4 打球種別

打球種別 NN の適合率、再現率、F 値を表 6 に示す。

表 6: 打球種別の適合率, 再現率, F 値

	適合率	再現率	F 値
ゴロ	0.59	0.60	0.59
フライ	0.60	0.61	0.62
平均	0.61	0.61	0.61

ゴロと予測されるときは投球のコースが高く、フライのときはコースが低い。

5.5 打球区域距離

打球区域距離 NN の適合率、再現率、F 値を表 7 に示す。

表 7: 打球区域距離の適合率, 再現率, F 値

	適合率	再現率	F 値
内野	0.64	0.37	0.47
外野	0.59	0.82	0.69
平均	0.62	0.59	0.58

内野と外野の比が 5:6 であるから、再現率に大きな差異が出た。

6. まとめ

本研究では、ニューラルネットワークを用いて打球予測を実現した。

手法としては、打席情報のデータを入力として、方向、強度、種別、距離を予測させた。

その結果、F 値はランダムの結果を超えることを示せた。

また、この研究は「日本統計学会 スポーツ統計分科会の第 9 回スポーツデータ解析コンペティション」の成果である。

7. 謝辞

本研究で使用したデータを提供して下さったデータスタジアム株式会社様、並びにこのような機会を与えて下さった『情報・システム研究機構 統計数理研究所 医療健康データ科学研究センター』に深く感謝の意を表します

参考文献

- [1] Rossi, A., Pappalardo, L., Cintia, P., Iaia, F. M., Fernández, J., Medina, D. 2018. Effective injury forecasting in soccer with GPS training data and machine learning. PLoS ONE 13, e0201264.
- [2] 岡谷貴之, "機械学習プロフェッショナルシリーズ 深層学習", 講談社, p.175, 2015.