

ピッチャーの配球を考慮した部分再帰型ニューラルネットワークに基づく 2M-03 プロ野球打者の初球戦略の最適化モデルの提案

谷畑 耀† 山下 遥‡
上智大学† 上智大学‡

1 はじめに

近年、野球の試合をより有利に進めるため、また、勝利する確率を上げるための打者に対するための支援に関して、様々なアプローチが提案されている。打者にとって難しい意思決定の1つとして投手が打者に投げる1球目(初球)に対して積極的にバットを振っていくべきなのか、慎重にボールを見極めるべきなのかの戦略が挙げられる。そのため、状況(アウトカウント、イニング、ランナーの有無、相手チームの情報、相手投手の配球、初球を打ったのか打ってないのかなど)の変化を考慮しながら、初球に対して最適な戦略を提示する方法が求められる。これまでの研究で、初球の重要性に関する議論は存在している。しかしながら、実際の試合の中で初球に対してどのようなアプローチを取れば試合を有利に進められるのかを明らかにするための数理モデルは提案されていない。

そこで本研究では、対峙する投手の配球やその場面での状況を考慮して打撃結果を予測するモデルを構築する。さらに、初球から積極的にバットを振っていくのか、いかないのかを変化させ、打席結果に対して好ましい戦略を決定するモデルを提案する。その際、相手投手が投げる配球は、前の時点での配球に大きく影響される時系列の構造を有するものである。また、状況は前打者までのプレイに依存して流れが大きく変化し、初球から積極的にバットを振るべきかどうかは変化するものと考えられる。このような複雑な状態に関して、精度よく結果を予測するために、時系列データを分析する再帰型ニューラルネットワーク(RNN[1])とニューラルネットワーク(NN)を組み合わせた部分再帰型ニューラルネットワーク(PRNN[2])を構築する。さらに、実際のプロ野球のデータを用いて、モデルの精度を確認する。また、提案モデルを活用した場合の予測得点を算出し、妥当性を考察する。

2 準備

2.1 分析対象データ

本研究では、日本統計学会スポーツ統計物科会の第11回スポーツデータ解析コンペティションから提供された野球データを使用する。実際、年度ごとの初球打ちの割合とチームごとの初球打ちの割合を集計すると、大きく異なることが確認された。そこで、本研究では、2018年から2019年の2年を通して成績が安定しており、かつ初球打ちを積極的に行っている西武ライオンズを対象にする。

Proposal of an optimization model for the first ball strategy of a professional baseball batter based on partial recurrent neural network considering the pitcher's throwing ball

†Akira Tanihata · Sophia University

‡Haruka Yamashita · Sophia University

2.2 部分再帰型ニューラルネットワーク (PRNN)

時系列データを分析する際、時系列による状態の変化を表すデータのみならず、データの特徴を表す非時系列データが存在するものと考えられる。このとき、非時系列データを時系列データの分析の中に入れてしまうと、変化がない非時系列情報を重視した学習を行ってしまうため、精度の良いモデルの推定が困難となってしまう。そこで、入力データのうち、過去の値を参照すべきデータおよびその時点だけを考慮すればよいデータへと分割し、前者はRNNの構造を中間層として、後者はNNの構造を中間層とし、2種類の中間層からの出力をまとめて出力層とするPRNNが提案され、その有効性が示されている。

3 提案モデル

今回は、PRNNの構造を基礎として、ピッチャーのバッターに対する初球の配球情報を過去の値を考慮する時系列入力変数、打席ごとの状況を過去の値を考慮しない非時系列入力データ、そして打席結果を出力変数として予測モデルを構築する。ここで、出力変数を打席結果とした場合、三振、凡打、単打、など12種類の打席結果があり、かつ好ましい出力が複数あると考えられるため、最適な戦略を決定することが困難になる場合がある。そこで本研究では、打席結果に対して得点の観点から0~1のスコア(例 凡打=0, 単打=0.25, 本塁打=1)を与え、出力変数を1次元の変数で表すことにする。これにより、"大きな値であるほど好ましい結果"という状況を作り出すことができるため、意思決定が容易になり、実応用性の向上が期待される。

まず、時刻 t ($t = 1, \dots, T$) における s 時点前までを考慮した L 種類の配球データを $\mathbf{x}_{lt} \in \mathbf{R}^s$ ($l = 1, \dots, L$) とする。また、 t 時点における試合状況を表す p 個の変数をもつデータを $\mathbf{z}_t \in \mathbf{R}^p$ とし、この中に初球を積極的に打ったのかを表す2値変数 $f_t \in \{0, 1\}$ を組み込む。 \mathbf{x}_{lt} はRNNの入力変数、 \mathbf{z}_t をNNの入力変数、 \mathbf{y}_t を本研究におけるPRNNの出力変数とし、 W_l^{inRNN} は l 番目のRNNの入力から k 個のニューロンをもつ中間層へのエッジの重み行列、 W_l^{RNN} はRNNの中間層の出力から次の時点でのRNNの中間層へのエッジの重み行列、 \mathbf{w}_l^{outRNN} は l 番目のRNNの入力から k 個のニューロンをもつ中間層へのエッジの重みベクトル、 \mathbf{h}^{lRNN} は t 時点の l 番目のRNNの中間層からの出力、 f_{RNN} はRNNの活性化関数とする。同様にNNの構造について、 W^{inNN} , \mathbf{w}^{outNN} , \mathbf{h}^{tNN} , f_{NN} を定義し、活性化関数

g を用いた次式のようなネットワークを想定する.

$$\begin{aligned} \mathbf{h}_i^{t_{RNN}} &= f_{RNN}(W_i^{in_{RNN}} \mathbf{x}_{tl} + W^{RNN} \mathbf{h}_i^{t-1_{RNN}}), \\ \mathbf{h}^{t_{NN}} &= f_{NN}(W^{in_{NN}} \mathbf{z}^t), \\ y_t &= g(\sum_{l=1}^L \mathbf{h}_l^{t_{RNN}} \mathbf{w}_l^{out_{RNN}} + \mathbf{x}_{tl} \mathbf{w}^{out_{NN}}) \end{aligned} \quad (1)$$

また、概要図を図2に示す. パラメータの最適化は誤差逆伝播法を用いる. さらに、得られたモデルに対して初球を積極的に打ったのかを表す2値変数 $f_t \in \{0, 1\}$ を変化させ、出力が最大になるものを選択することで最適な戦略を探索することができる.

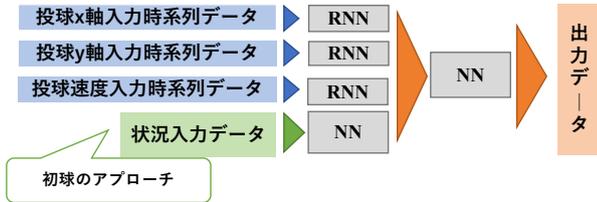


図1: 提案手法におけるモデルの概要

4 実データ分析

4.1 分析条件

提案した投手の配球を考慮した PRNN に基づくプロ野球打者の初球戦略の最適化モデルを用いた西武ライオンズの打者データの分析例を以下に示す. まず2018年から2019年後半、10000打席を学習データとし、それ以降の30時点の打席結果を予測する. このときの入力変数は、過去の値を考慮する (RNN に用いる) 時系列変数、過去の値を考慮しない (NN に用いる) 非時系列変数とし、目的変数はスコア化した打席結果とする.

4.2 予測精度による提案手法の妥当性の評価

まずテストデータである30時点を提案手法により予測した場合、LightGBMでの分析結果と比較し、提案手法の妥当性を評価する. 結果を表1に示す. LightGBMと比較して、打撃結果の変動を適切に捉えることができていることが分かるものの、十分な説明能力がないことが示唆される. ただし、ここでは、提案手法を用いた予測に元づく最適な戦略についても信頼ができる結果であると仮定し、妥当性の検証を行う.

表1: 提案手法および従来手法を用いたスコアの評価

手法	平均絶対誤差	相関係数
従来手法 (LightGBM)	0.139	0.181
提案手法	0.186	0.407

4.3 モデルの有効性の評価

次に提案手法及び従来手法を用いて戦略の最適化を行ったときの合計スコアを以下の式に基づき求め、提案手法の有効

性について検討する. 学習したモデルから算出される予測値のうち、初球から積極的に振っていくと仮定したときの予測値を $\hat{y}_i^{(0)}$, 実測値を y_i ($i = 1, \dots, n$), としたとき、合計スコアおよび上がり幅は (2)(3) 式で表される.

$$\text{合計スコア} = \sum_{i=1}^n \max(\hat{y}_i^{(0)}, \hat{y}_i^{(1)}) \quad (2)$$

$$\text{合計上がり幅} = \sum_{i=1}^n \{\max(\hat{y}_i^{(0)}, \hat{y}_i^{(1)}) - y_i\} \quad (3)$$

表2: 提案手法および従来手法を用いたスコアの評価

手法	合計スコア	合計上がり幅
従来手法 (LightGBM)	3.845	0.095
提案手法	6.726	2.976

以上の結果より、提案手法を当該データに適用することで、打撃力を表すスコアの値を大きく改善することができるものと思われる. よって、提案手法の妥当性および有効性の双方を示すことができたと考えられる. さらに、本研究の提案モデルを用いて打席ごとに打者の最適な初球に対するアプローチを選択すればチームとして打撃能力を高めうることが示唆される.

5 おわりに

本研究では、時系列データを含む複雑な入出力データの関係を精度よく推定するための分析方法の提案を目的とし、PRNNの枠組みを基礎としたモデル化を提案した. さらに、学習によって得られた高性能な予測器を用いて、初球に対して積極的に打つべきかを最適化する方法を提案した. また実際のプロ野球のデータを用いて、モデルの妥当性 (精度) および、有効性 (得点への影響) を考察した. その際 LightGBM と比較し、提案手法の従来手法に対する優位性を示した. 今後の課題は、より良い精度の結果を出すためのモデルの検討および、モデルを学習するタイミングの検討、西武以外ののチームに提案手法を適用したときの性能の評価、プロ野球ゲームを活用した提案手法の実用性の評価などが挙げられる.

参考文献

- [1] Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 404, 132306.
- [2] 福永峻, et al. (2019) ”部分再帰型ニューラルネットワークを用いたヘアサロンチェーンにおける顧客の離脱予測.” オペレーションズ・リサーチ= Communications of the Operations Research Society of Japan: 経営の科学 64.2: 87-94.

この研究は第11回スポーツデータ解析コンペティションで発表したものである. データの提供に感謝いたします.