

プロ野球選手育成のためのデータ解析システム

木下 涼† 北 直樹† 斎藤 隆文†

東京農工大学 工学部情報工学科†

東京農工大学 工学部情報工学科†

1. 背景と目的

近年、データ解析は医療、ビジネス、技術開発など様々な分野で用いられている。データ解析には既存のデータから未知データを推定する機械学習、データの傾向を探る回帰分析などが存在し、目的に応じて使い分けることができる。しかしながら、データ解析にはプログラミングの知識だけでなく、各分析手法の知識が必要になり、専門知識がない人には難易度が高い。データサイエンティストの人数が世界的に不足しており[1]、データ解析を行える人員を増やすことが必要とされている。そこで、本研究は初学者でも扱えるデータ分析システムを構築する。研究対象として、データの収集を行いやすい野球を対象とする。

メジャーリーグ(MLB)では、後述のようなプロ野球選手がデータ解析を利用したトレーニング支援を受けることができるサービスを提供している企業があり、日本のプロ野球(NPB)の一部球団でも利用している。しかし、個人的なデータを球団側が把握することができず、チーム全体の練習を効率的にできているとは言えない。本研究では、トラックマンという試合中のデータを集計する機械で得られるデータを、球団が選手にデータを提供し、選手自身が自分でデータ解析を行い、自分の弱点を発見できるようにすることが目標である。発見した弱点を基に選手が自分で練習内容を考え、克服すること(セルフコーチング)に本研究が役立つと考える。

2. 既存のデータ解析システム

本章ではプロ野球の球団が利用しているトラッキングシステムのデータを用いた選手強化サービスとタブレット上で動作するデータ解析アプリケーションを挙げる。

2.1. 練習用トラッキングシステム

Rapsodo Baseball[2]は、2016年にアメリカで発売された野球・ソフトボール用のトラッキングシステムである。このシステムの特徴は、小型でどこにでも持ち運んで使える点、一球ごと

Data analysis system for training pro baseball player

†Ryo Kinoshita, Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology.

†Naoki Kita, Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology.

†Takafumi Saito, Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology.

のデータをその場でタブレットを用いて確認できる点の二点である。

2.2. トレーニング支援サービス

データを用いた指導を行う外部企業のサービスとして、Driveline baseball[3]がある。この会社は、トラッキングシステムで集めた投球及び打球のデータとともに、選手につけたセンサーから得られる体の動作のデータを活用して指導を行っている。

2.3. データ解析アプリケーション

タブレット上でデータ解析を行うことができるアプリケーションはあまり存在しない。データ解析を行うことができる Scientific Data Analysis for the iPad[4]というアプリケーションも、使うことができる解析モデルに制限があり、利便性はあまり良いとは言えない。

3. データ可視化システム

本章では、提案システムの説明を行う。

3.1. システムの目的

MLBに大きな影響力を持つ The Book[5]では、チーム強化のためにプロ野球関係者全員がデータ解析に触れる必要性が説かれている。しかし、現状ではデータ解析を行うのは各球団のデータ解析班のスタッフだけであり、全員がデータ解析に触れてはいない。

選手やコーチがデータ解析を行わないデメリットとして次の二点があげられる。一つ目の短所として、データ解析班のスタッフの中にはあまり野球経験がない人もおり、適切なアドバイスを行えない点がある。二つ目の短所は、データ解析を気軽に行えないことにより、欲しい情報をすぐに得ることができない点である。

前章で挙げた練習用トラッキングシステムとトレーニング支援サービスなどの外部サービスに共通する短所としては、トラックマンから得られる試合データを利用できないことによる、練習の結果が試合に生かされているか不透明な点がある。

本研究では、球団が所持するデータを用いて選手やコーチがデータを解析できるようにすることで、効率の良い選手育成をできるようにすることが目的である。そのための工夫として、

データ解析班のスタッフが解析モデルやデータ群を作る作業と、選手やコーチが解析を行う作業を分割する。これにより、解析を行う人の知識や作業環境にとらわれず、データ解析が可能になる。

3.2. システムの流れ

システムは図 1 のように、まずデータ解析班のスタッフがパソコンで作業を行い、その後選手やコーチがタブレットで解析を行う。

データ解析班は、データの集計、データ群作成のための条件設定、データ解析モデルの実装、データ群と解析モデルの説明文記述の四つの作業を順番に行う。そして、作成したデータ群条件と解析モデルをタブレットにインポートする。

選手やコーチが利用する際には、データ群と解析モデル、そして出力形式をタッチ操作で選択することで解析を行うことができる。

本研究では実用性を検証するために、データ解析モデルを実装する代わりに、新たな指標を作成して検証できる環境を構築する。その理由は、正確な解析を行えているか検証する際に、独自指標を用いる方が分かりやすいためである。

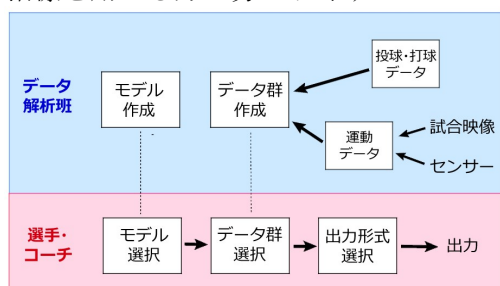


図 1: システムの流れ

3.3. 独自指標の作成

新たな指標を自作して独自の視点で選手を評価する手法は、プロ野球にデータ解析が持ち込まれたところから用いられている手法である。

本研究では、2019 年の MLB 公式戦データを用いて、選手が貢献した得点を表す独自指標 RC(TT)と選手があげた打点を比較した。図 2 は打点とチームへの貢献を表す指標である WAR の散布図、図 3 は打点と WAR の散布図である。二つの散布図を見比べると分かる通り、打点よりも RC(TT)の方が WAR との相関が強く、チームへの貢献度合いをより表していると考えられる。これにより、RC(TT)などの独自指標は既存の指標だけでは分からない、選手の隠れた能力を測

ることが可能になるということが出来る。

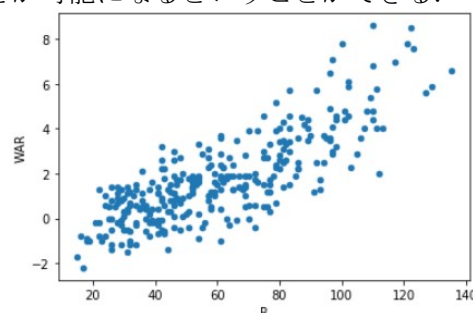


図 2: 打点と WAR の相関

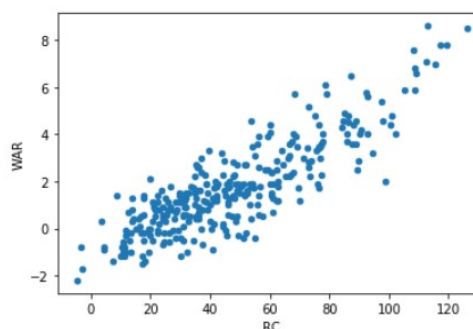


図 3: RC(TT)と WAR の相関

3.4. データ群

提案システムでは、データ群をあらかじめデータ解析班が生成するのではなく、選択された段階で全データから条件に応じて自動生成するようになっている。これにより、つねに最新のデータを用いて独自指標を算出することができ、選手の現状を正確に分析することが可能になる。

3.5. 出力形式

データの出力形式は散布図、棒グラフ、ランキングの三種の中から選択する。散布図は独自指標の正当性の評価、棒グラフは選手の比較、ランキングは独自指標でソートすることで新たな見方で選手の優劣を評価することが可能になる。

5. おわりに

本論文では、データ解析に慣れていない初学者のデータ解析を支援することを目的とし、タブレット上でデータ解析を行う手法の提案を行った。本論文では、データ解析モデルの実装には対応しておらず、まだ本格的なデータ解析を行うことができない。今後は、回帰分析モデルや機械学習を用いたモデルなどを実装し、様々な解析を行うことができるようにしていきたい。

参考文献

[1] 総務省 平成26年版 情報通信白書
 [2] Rapsodo baseball, (最終閲覧日: 2020/01/10) ,<https://rapsodo.com/ja/baseball/>
 [3] Driveline Baseball, (最終閲覧日: 2020/01/10) ,<https://www.drivelinebaseball.com/>
 [4] Scientific Data Analysis for the iPad, (最終閲覧日: 2020/01/10) ,<http://www.scidataanalysis.com/sda/main.html>
 [5] Tom Tango, Mitchel Lichtman, Andrew Dolphin, (2014), The Book: Playing The Percentages In Baseball