

# 転移学習を用いたセイバーメトリクスに基づく年俵評価

慶應義塾大学大学院 松木 拓弥

慶應義塾大学 鈴木 秀男

## 1. はじめに

数年前から、メジャーリーグ (MLB) において選手年俵の高騰が話題となっている。それに伴い、資金力が少ない球団は主力選手の年俵を支払えないため放出せざるを得ず、チーム力が低下してしまうという状況になるケースがある。そのような課題に対して、オークランド・アスレチックスが「セイバーメトリクス」と呼ばれる選手の評価や戦略をデータで分析する手法を用いて、高年俵選手を放出して資金を作り、低年俵で優秀な選手を獲得するという戦略を用いて 2002 年にチームを優勝に導いたという例がある。本研究は、2015 年の MLB 公式戦全試合の投手の球質に関するデータ (トラッキングデータ) を用いて、選手の年俵に関する妥当性を評価し、球団経営に実践することを目的とする。

## 2. 分析内容

本研究では、共変量シフトに対応した転移学習 [1] [2] をパス解析に導入し、トラッキングデータ⇒セイバーメトリクス指標⇒年俵の因果関係に基づく予測モデル構築・分析を行った。共変量シフトとは入力分布が訓練、テスト標本で異なる状況を意味し、それに対応するために訓練、テスト標本の確率密度関数  $p_0(x_i), p_1(x_i)$  によって以下のように重み付けを行う。

$$\min_{\theta \in \Theta} \left[ \frac{1}{n} \sum_{i=1}^n \left( \frac{p_1(x_i)}{p_0(x_i)} \right)^\lambda l(x_i, y_i, \hat{f}(x_i; \theta)) \right]$$

但し、 $l(x, y, \hat{y})$  は損失関数、 $\lambda$  はパラメータを表す。分析の流れは以下の①～④である。

- ① 球種ごとのデータセットに分ける
- ② 予測する球種以外のデータセットで重み付けをしてパス解析 (トラッキングデータから求める新たなセイバーメトリクス指標を考案し、モデルに組み込む)
- ③ 予測する球種のデータをモデルに当てはめて投球のスコアを算出
- ④ 投手ごとのスコアを③で求めたスコアを基に算出し、年俵との関係性と妥当性を確認

## 3. 結果・考察

詳細な結果・考察は当日報告を行う。共変量シフトに対応した転移学習の有効性や実測年俵と予測年俵で差のある選手を抽出し、良好な球団経営に向けての考察について報告する。

## 参考文献

- [1] Shimodaira, H., “Improving predictive inference under covariate shift by weighting the log-likelihood function”, Journal of Statistical Planning and Inference Vol. 90, pp. 227-244, 2000
- [2] S.J. Pan and Q. Yang, “A survey on transfer learning”, IEEE transactions on knowledge and data engineering, 2010