

データの可視化を用いた野球観戦システム

加藤雄一[†] 澤田光明[†] 齋藤友彦[†]

湘南工科大学[†]

1. はじめに

プロ野球は娯楽として多くの人に親しまれている。近年では、選手の詳細なデータを可視化させた野球放送も実施されている。しかしながら、この可視化システムは一部のテレビ局が放送する試合のみに使われ、ユーザの好みの試合に使用することは不可能である。例えば熱狂的なプロ野球ファンは2軍の試合も好み、詳細なデータが表示された映像を望むが、現在のシステムではそれを実現することはできない。

本研究では、機械学習による画像認識とスクレイピングを用い、データを可視化させた新しい野球観戦システムを開発する。これによりユーザが好みの試合映像に自由にデータを可視化させることができる。システムの具体的な流れは次の通りである。まず、画像認識の一つである You Only Look Once (YOLO) [1]を使い試合動画から選手の認識を行う。次に、試合に付随したデータがあるプロ野球-スポーツナビ[2]、プロ野球データ Freak[3]からスクレイピングを行う。認識した選手の情報を抽出し、画面に表示する。

本研究では、著者が独自に用意した教師データを使い機械学習を行う。その際の選手の認識率を実験により評価する。更に、本システムの使用感をアンケートにより評価する。

2. 提案システム

2.1 概要

提案システムの概要を図1に示す。以下、提案システムの要素技術である画像認識 YOLO, スクレイピング, 可視化システムの詳細を述べる。

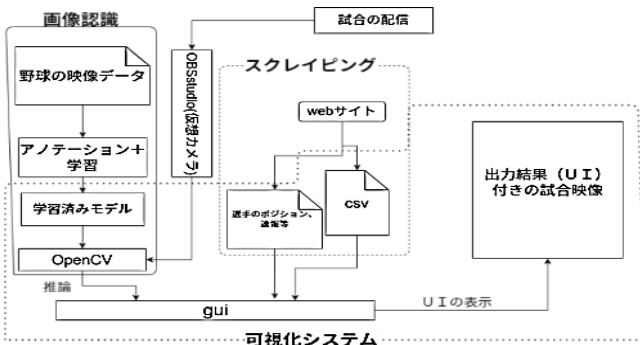


図1 提案システムの概要

2.2 画像認識

YOLO はリアルタイムオブジェクト検出アルゴリズムである。このアルゴリズムでは検出窓をスライドさせるのではなく、画像を一度に CNN (Convolutional Neural Network) に通すことでオブジェクトの検出を行う。

YOLO では、画像をグリッドセルとして扱う。各グリッドセルはバウンディングボックスを持ち、それらのボックスの信頼度 (ボックスにオブジェクトが含まれているかを表す確率) を予測する。また、各グリッドセルはクラスに対する条件付き確率を予測する。この条件付き確率とバウンディングボックスの信頼度を掛け合わせることで信頼度スコア (Confidence score) が得られる。信頼度スコアがある閾値以上の時、対象のオブジェクトとして検出する。

本研究では、YOLO を使い、投手、捕手、打者の認識を行う。まず、野球の試合録画データを用意し、アノテーションソフトを使用して教師データを作成する。そして、Darknet と呼ばれるニューラルネットワークフレームワークを用いて学習を行う。最後に OpenCV を使い推論を行う。推論結果の一例を図2に示す。図2では打者に対してバウンディングボックスを表示させている。

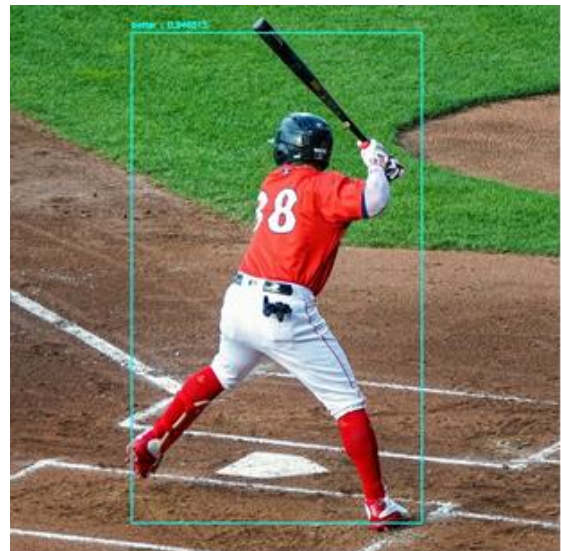


図2 文献[4]に対しての画像認識結果

2.3 スクレイピング

本研究ではPythonのライブラリであるpandas, Requests, BeautifulSoup4 を使用してスクレイピングを行い、認識した選手の情報を取得する。プロ野球-スポーツナビ[2]から、試合の速報デー

Baseball Viewing System Using Data Visualization

[†]YUICHI KATO Shonan Institute of Technology

[†]MITSUAKI SAWATA Shonan Institute of Technology

[†]TOMOHIKO SAITO Shonan Institute of Technology

タを取得し、現時点での投手、捕手、打者が誰なのかを特定する。そして、特定した選手の詳細なデータをプロ野球データ Freak[3]から取得する。

2. 4 可視化システム

OpenCV のみでも簡単な文字列であれば描画は可能だが、アニメーション等を加えることができない。そこで、本研究では Python の GUI ライブラリの一つである pygame を使って可視化システムを開発した。まず、試合映像の上に pygame を使って透明なウィンドウを表示させる。仮想カメラを通して試合の画面を OpenCV に入力し選手を認識する。スクレイピングによって、選手データを取得し、取得したデータを pygame で表示させる。概念図を図 3 に示す。

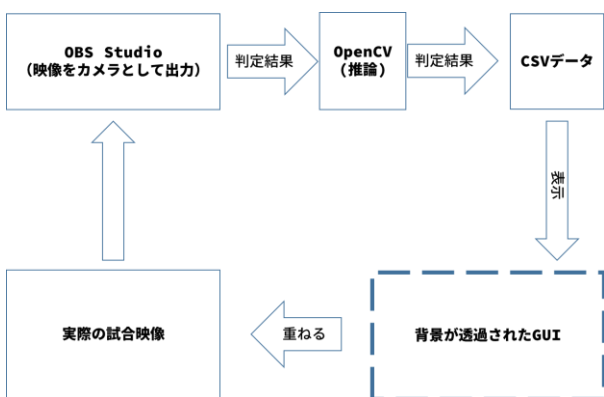


図 3 可視化システムの概念図

3. 評価実験及び考察

3. 1 認識率に関する実験

YOLO の認識率を実験により評価する。教師データとは別のテストデータ 126 枚に対し、投手、捕手、打者の認識を行う。信頼度スコアの閾値を変化させたときの適合率、再現率を表 1 に示す。表 1 より、閾値が 30%以上から適合率が高くなり、45%以上であれば、適合率は90%以上となる。また、閾値を 30%としたとき、再現率も高く、バウンディングボックスも十分に表示されることが確認された。

表 1 実験結果

閾値	適合率	再現率
15%	69%	83%
30%	83%	75%
45%	90%	68%
55%	95%	61%
70%	96%	51%

3. 2 アンケート評価

3.1 節の実験より、信頼度スコアの閾値を 30% にすれば、適合率及び再現率が十分に高いことが確認された。そこで閾値を 30%に設定し、可視化システムを構築した。このシステムの使用感を評価するため、被験者 5 名に対して、アンケート調査を行った。被験者には約 7 分の動画をシステム上で視聴してもらい、その後「表示されるデータに対して違和感がありますか？」という質問に対して 5 段階評価 (1. 悪い, 2. やや悪い, 3. どちらともいえない, 4. やや良い, 5. 良い) で回答してもらった。その結果、3 人が「3. どちらともいえない」、2 人が「4. やや良い」と回答した。5 名の被験者に対して、詳細な感想を聞いたところ、データの表示は適切になされていると回答した。従って、画像認識及び閾値の設定は適切に行われていると考えられる。また、「3. どちらともいえない」と回答した被験者にその詳細を聞いたところ、「表示される速度が遅く感じる」との感想があった。これは画像処理が重くなることを避けるため画像判定を一定周期で行うように実装したためと考えられる。

4. おわりに

本研究では、画像認識の一つである YOLO とスクレイピングを用いて、画面上に選手のデータを表示する観戦システムの開発と評価を行った。今後の課題として、リアルタイム観戦システムの構築と評価、見た目の改良によるエンターテインメント性の強化などが挙げられる。

謝辞

本研究は JSPS 科研費 JP17K00316, JP18K11585 の助成を受けたものです。

参考文献

- [1] Redmon, J., Divvala, S., Girshick, R. and Farhadi A.: You Only Look Once: Unified, Real-Time Object Detection, Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), pp.779-788 (2016).
- [2] Yahoo! JAPAN: プロ野球-スポーツナビ, スポーツナビ, 入手先<<https://baseball.yahoo.co.jp/npb/>> (参照 2021-12-29).
- [3] プロ野球 Freak: プロ野球データ Freak, 入手先 <<https://baseball-data.com/>> (参照 2021-12-29).
- [4] Pixabay: 野球 ゲーム プレイヤー - Pixabay の無料写真, 入手先 <<https://pixabay.com/ja/photos/%e9%87%8e%e7%90%83-%e3%82%b2%e3%83%bc%e3%83%a0-%e3%83%97%e3%83%ac%e3%82%a4%e3%83%a4%e3%83%bc-5239230/>> (参照 2022-01-28)