Determining Baseball Performance Quality with an Elo Rating

Mac Bagwell

Stanford Computer Science

Abstract

Recent advances in data collection for MLB games has made quantitative analysis of individual player performance easier and richer. However, analysis of minor-league players lags behind that of major-leaguers in sophistication, despite the fact that franchises place greater value on the individual performance of their prospects than on the performance of the teams they happen to be playing on. In this project, we attempt to develop a elo-style rating system, similar to that employed by FIDE for chess players, to function as a point-in-time indicator of skill. In contrast to the FIDE system, we employ a modified version of L_2 regularization, inspired by the winning submission of a Kaggle competition. We apply this elo system to the task of predicting Baseball America's top prospects for the following season. We see minor but significant improvement over baseline models with the incorporation of an elo rating and other evidence that our rating system contains information about point-in-time performance.

Stanford University

Background

When MLB franchises draft players, they are not sent straight to the MLB; they start playing in smaller development leagues. Players may be called up to their MLB team's roster, but only if the quality of their play is deemed high enough.

Data

Play-by-Play Data

- ~14 million plate appearances for all games in Rookie, A-, A, A+, AA, AAA, MLB leagues since 2007
- Scraped from MLB Advanced Media's Gameday servers

Pitching Prospect Data

- ~45,000 observations of pitching prospects since 2007
- Observed draft round, Baseball America ranking, years of experience
- Draft round one-hot encoded
- Train-Dev-Test split on 70-10-20 ratio
- Dataset provided by Pando Pooling

Hyper-parameter Search

Table 1: Grid-Searched Hyperparameters

K	λ	β	O
0.1	0.1	0.95	0
0.01	0.01	0.9	0.125
0.001	0.001	0.85	0.25
0.0001	0.0001	0.8	0.5

Selected to minimize log-loss of expected outcome

$$L = (o_{ij} - \log p) + (1 - o_{ij})\log(1-p) + \frac{\lambda}{2}[(r_i - a_i)^2 + (r_j - a_j)^2]$$

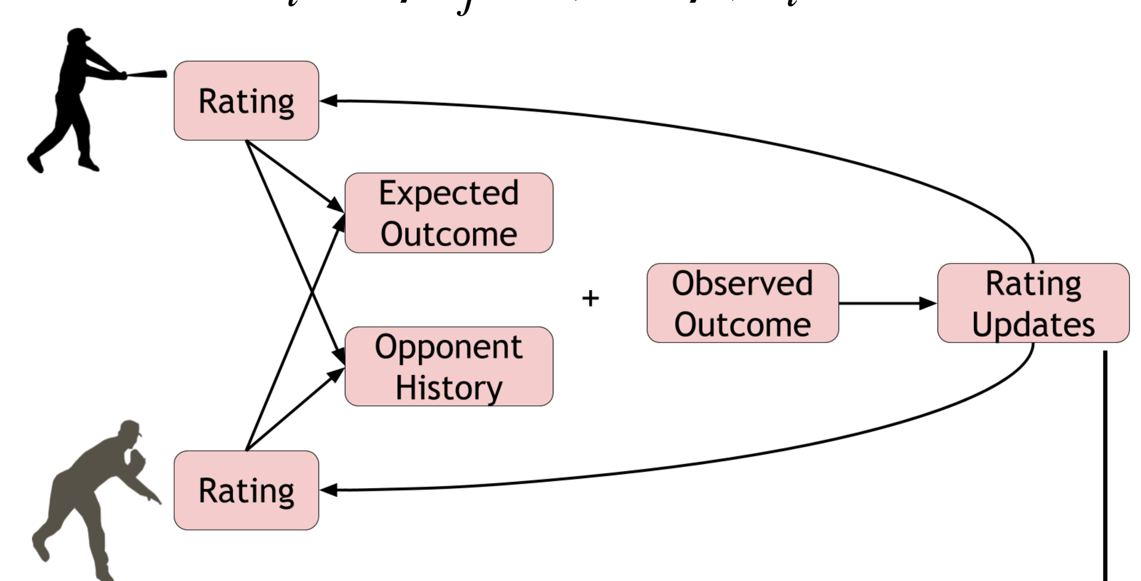
Elo Rating System

Overview

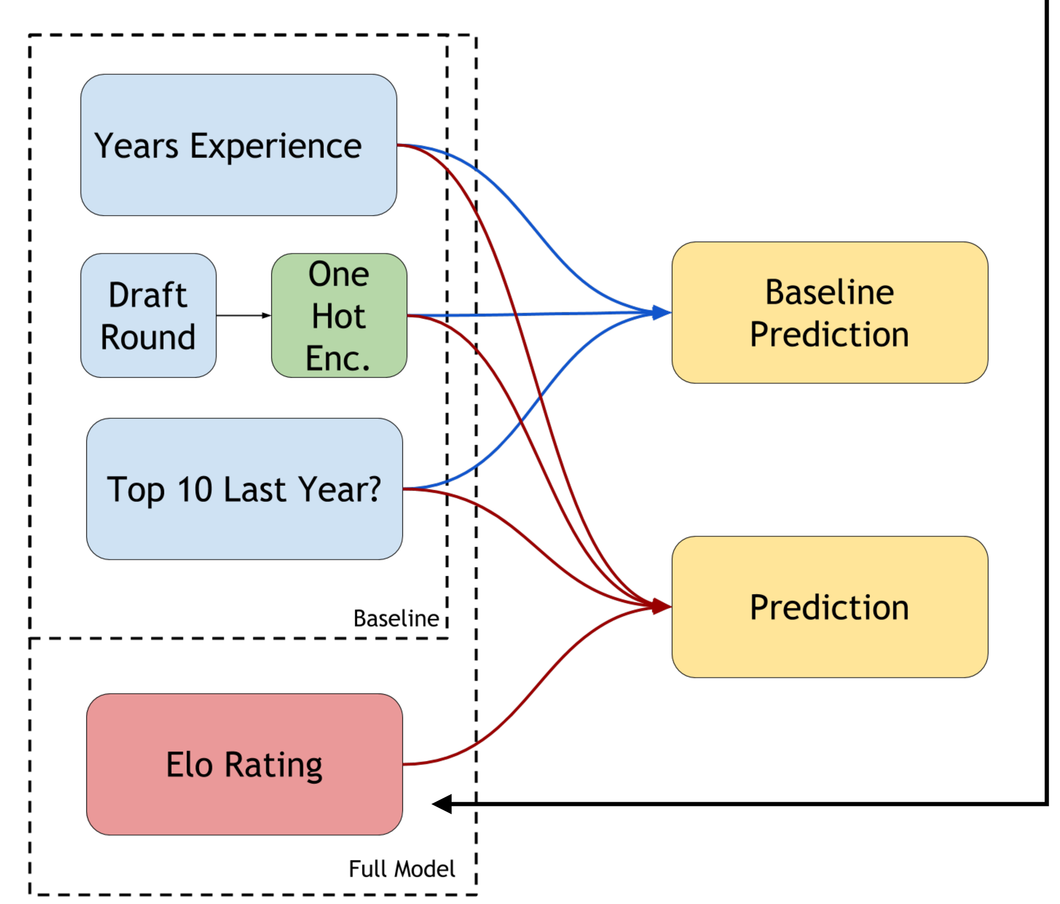
- Every player gets a rating of 100 to start, adjustment for level.
- A player's rating goes up when they "win" and down when they "lose".
- Size of movement determined by difference in ratings before contest
- Regularized towards EWMA of opponent ratings (better players tend to be playing with better players)

Update Rule

$$r_i \leftarrow r_i + K\left((o - p) - \lambda(r_i - a_i)\right)$$
$$a_i \leftarrow \beta r_j + (1 - \beta)a_i$$



Prospect Classification



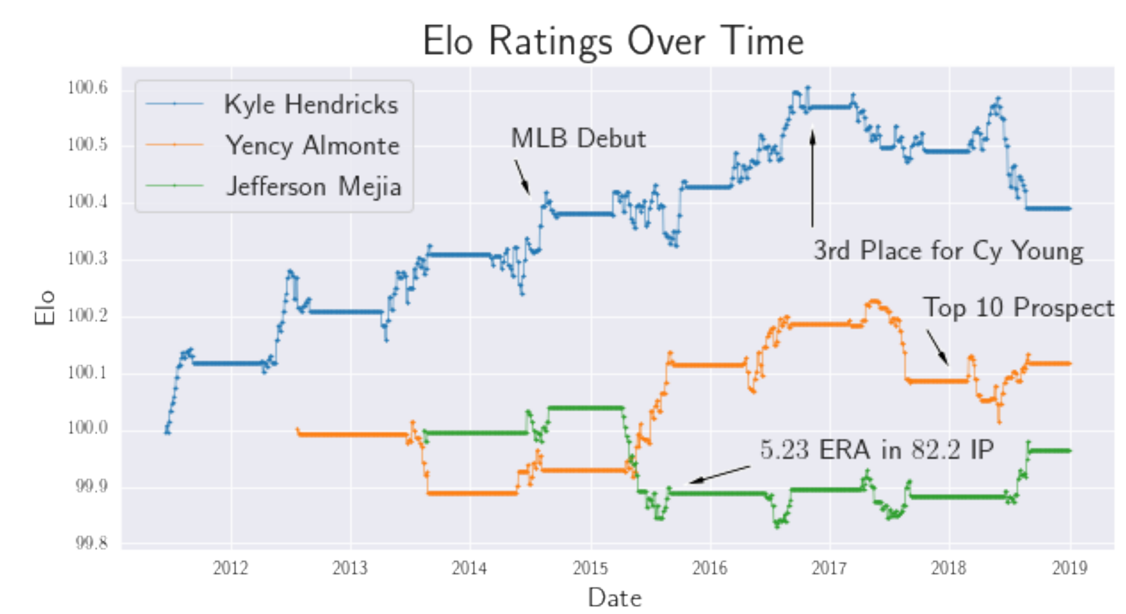
Samples are weighted by inverse of class frequency to combat class imbalance. Models are compared on AUPRC metric.

Results

Elo Ratings

Table 2: Hyperparameters and Loss of Optimal Model

K	λ	β	O	log-loss
0.01	0.01	0.9	0	0.623341



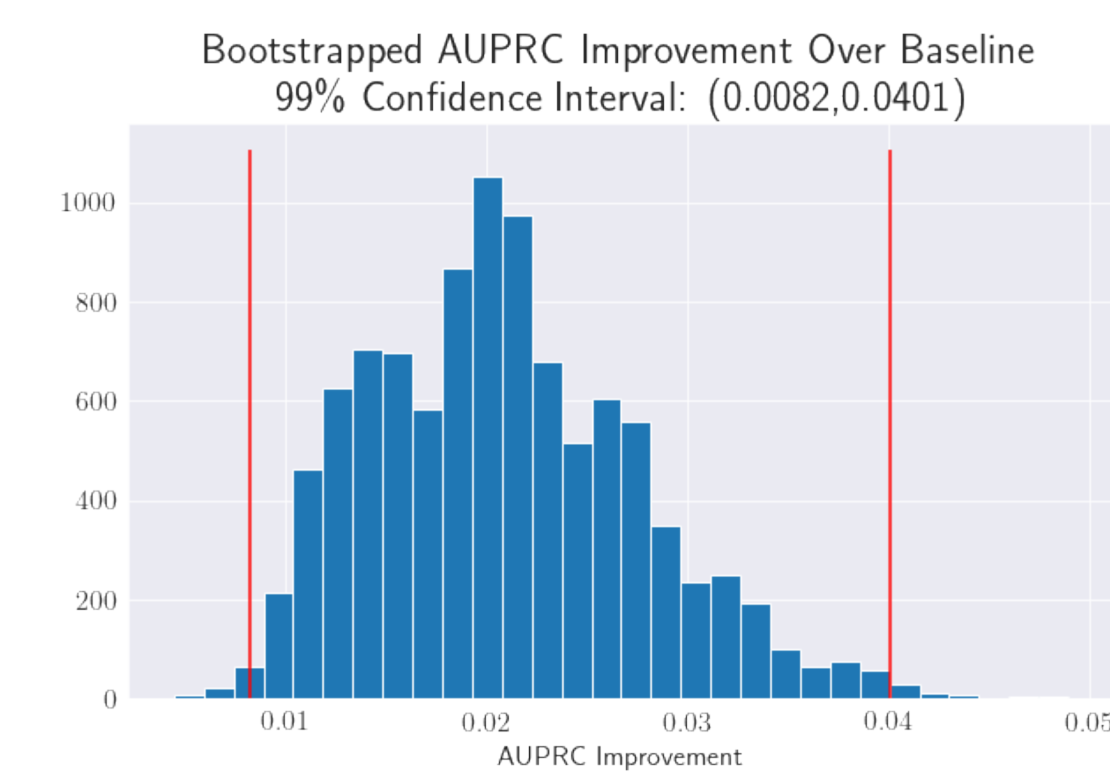
Sample elo ratings over time for three professional pitchers. Increases/decreases broadly correlate with increases/decreases in performance.

Classifier Perfomance

Table 3: AUPRC Scores for Logistic Regression Classifiers

Model	Dev AUPRC	Test AUPRC
h_0 (baseline)	0.3941	0.4136
h	0.4141	0.4234

We observed modest improvements over baseline model. Boostrapping suggests this improvement is significant.



Discussion

- Cursory examination of players' elo ratings suggests that fluctuations in elo rating correlate with other indicators of success, such as promotion to more advanced levels and career milestones.
- Relative scales of elo ratings suggests that searched parameters were not on the proper scale.
- Inclusion of elo in our classifier improved AUPRC significantly, verified by bootstrap. However, the size of this improvement was small.

Areas For Improvement

- Smarter/more efficient hyper-parameter search.
- Weighting of outcomes; home runs and extra-base hits could be weighted to cause more drastic changes in elo rating.
- Pre-processing of elo rating; extract more complex signals instead of just raw value.

References

1. Baseball America. *Baseball America Home Page* 2019. https://www.baseballamerica.com/rankings/ (2019).
2. FanGraphs. *FanGraphs Prospect Coverage* 2019. https://www.fangraphs.com/prospects/.
3. Wikipedia. *Elo rating system* 2019. https://en.wikipedia.org/wiki/Elo_rating_system (2019).
4. Silver, N. *2019 MLB Predictions* 2019. https://fivethirtyeight.com/features/how-our-mlb-predictions-work/.
5. Sismanis, Y. How I won the "Chess Ratings - Elo vs the Rest of the World" Competition, 1–8 (2010).
6. BaseballReference. *2016 Minor League Encyclopedia* 2019. https://www.baseball-reference.com/register/league.cgi?group=Minors&year=2016 (2019).
7. Bureau, C. *Baseball Data Bank* 2019. http://chadwick-bureau.com/open-data/.