Predicting NBA Draft Prospect Value with Learning to Rank

Introduction

- Evaluating NBA prospects is tricky
- Historically, teams have misjudged player value from the top of the draft to the bottom
- Evaluation on an individual basis is insufficient–drafting is a *ranking* problem
- Learning to rank via LambdaMART could be applied to the problem of drafting NBA players

Data

Both college and NBA data were obtained from Sports Reference (called Basketball Reference for NBA stats).

- Height and weight
- College box score stats
- College advanced stats
- Draft year and pick position
- NBA advanced metrics

Feature Engineering

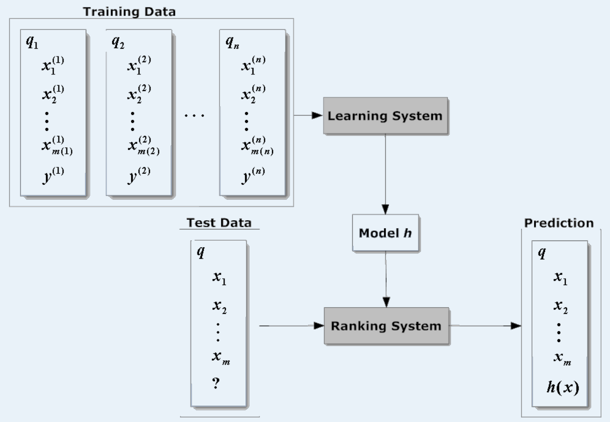
Examples of features that were generated in addition to those scraped from the online database:

Feature	Definition
`years`	Number of years in college
`bling`	Number of awards earned in college
`top_conf`	Player played for a top 5 conference[1]
`top_school`	Player played for a top 5 school[2]
`rsci_10`	Player was a top 10 high school recruit

[1] SEC, ACC, Big-10, Big-12, Pac-12
[2] Duke, Kentucky, Kansas, UNC, UCLA

Learning to Rank

LTR is a machine learning paradigm centered around ranking items based on some relevancy metric.



In terms of this project, q_i is the i^{th} draft from 1998 to 2018, x_{ij} the college stats (and other features) for the j^{th} player from q_i, and y_{ij} is the corresponding relevancy based on NBA VORP.

LambdaMART

LambdaMART is essentially XGBoost with a specialized loss function proportional to the Normalized Discounted Cumulative Gain (NDCG).

$$NDCG_k = \frac{DCG_k}{IDCG_k} \tag{1}$$

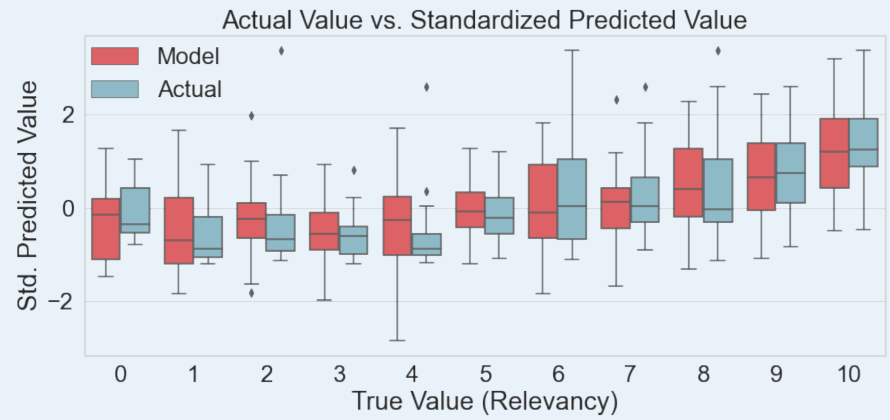
where the Discounted Cumulative Gain (DCG) at rank position k (for relevancy sorted by predicted value) is given by

$$DCG_k = \sum_{i=1}^{k} \frac{2^{rel_i} - 1}{\log_2(i+1)} \tag{2}$$

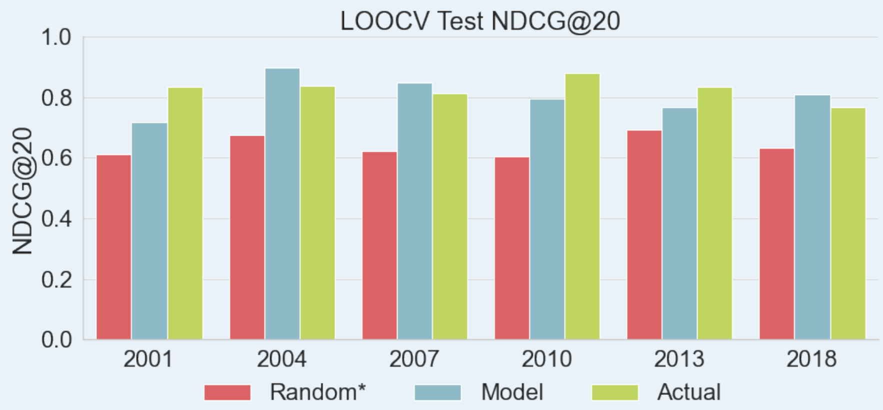
and IDCG_k is the maximum value of DCG_k reached at the optimal ordering.

Model Evaluation

We can compare the model rankings to the actual order in which players were drafted, thereby using NBA teams' decisions as a benchmark for model performance.



Grouping by true relevancy, we can see the distributions of the predicted scores are quite similar. The model appears more optimistic about worse players, though.



Again, the model performs similarly to NBA teams in evaluating players. It even outperforms the true draft order according to NDCG@20 in three of the six test drafts.

Jack McCarthy
jack.mccarthy@duke.edu
Duke University