Outcome Forecasting in Sports

Ondřej Hubáček

Motivation & Challenges

Motivation

- exploiting betting markets
- performance optimization

Challenges

- no available datasets
- difficulties with establishing the state-of-the-art
- the best models are not published
- gap between science and practice
- citation graph not connected

Sports

- individual vs team
- most of the popular sports are team sports
- more detailed statistics are gathered in team sports
- team sports events are more common
- team sports provide more betting opportunities
- individual sports suffer more from performance variance

$\Longrightarrow$ team sports are more suitable for applying ML

Task Characteristics

- the actual results are stochastic in nature
- we are usually interested in probabilities of the outcomes
- it looks like there is a glass ceiling about 75 % accuracy
- lot of space for feature engineering
- the features are more important than the selected ML algorithm
- relational character of the data

Types of Data

- results
  - \+ always available
  - – not enough information
- box-score statistics
  - \+ usually available
  - – information aggregated without context, not always objective
- play-by-play data
  - \+ provide better context
  - – rarely available
- player-tracking-data
  - \+ almost complete description of the game
  - – not available for free, only for top leagues

Bradley-Terry model [1]

- probability, that team i beats team j is given by

$$P(T_i \succ T_j | \pi_i, \pi_j) = \frac{e^{\pi_i - \pi_j}}{1 + e^{\pi_i - \pi_j}}$$

- the team's strength π_i is given by

$$\pi_i = \sum_k \beta_k (x_{ik} - x_{jk}) + U$$

Elo Rating[4]

- player's skill conforms to normal dist. with fixed variance β^2
- outcome is a function of the two players' skill ratings s_1 and s_2

$$P(p_1 > p_2 | s_1, s_2) = \Phi(\frac{s_1 - s_2}{\sqrt{2}\beta})$$

- Φ denotes the cumulative density of $\mathcal{N}(0, 1)$
- after the game, the skill ratings s_1 and s_2 are updated such that the observed game outcome becomes more likely

Elo in practice

- Let r_i represent the initial Elo rating of player i
- $R_i = 10^{\frac{r_i}{400}}$
- expectation of game outcome $E_i = \frac{R_i}{R_i + R_j}$
- new rating $r_i' = r_i + K \cdot (S_i - E_i)$

$$S_i = \begin{cases} 1, & \text{if player } i \text{ won} \\ 0.5, & \text{if player } i \text{ tied} \\ 0, & \text{if player } i \text{ lost} \end{cases}$$

— does not differentiate white/black pieces („home/away“)

Glicko-2 rating[6]

- implemented on chess servers, Counter Strike: GO, ...
- each player has rating r and a rating deviation RD
- Glicko-2 introduces rating volatility σ
- volatility: degree of expected fluctuation in a player's rating
- RD increases with time since last game (affected by σ)
- 10-15 games long burn in period

TrueSkill^{TM}[7]

- developed by Microsoft, presented at NIPS
- builds on Glicko
- can asses individual skills from team results
- applicable for games with multiple teams
- applies Bayes rule

$$p(s|r, A) = \frac{P(r|s, A)p(s)}{P(r|A)}$$

- posterior distr. is approximated and used as prior for next game

Pi-ratings[3]

- state-of-the-art ranking system for soccer
- separate rating for home/away matches
- updating home team's home rating:

$$R'_{\alpha H} = R_{\alpha H} + \psi_H(e) \times \lambda$$

- updating home team's away rating:

$$R'_{\alpha A} = R_{\alpha A} + (R'_{\alpha H} - R_{\alpha H}) \times \gamma$$

- large wins are diminished

$$\psi(e) = c \times \log 10(1 + e)$$

Utilizing Boxscores

- the main challenge is how to aggregate the information
- calculation seasonal averages or sliding averages is common
- few features allows sampling multivariate distribution
- most of the papers consist of applying off-the-shelf learners
- ANNs and SVMs generally perform best
- opportunities for RNN and CNN

Modeling basketball play-by-play data[8]

- game as a Markov process $\{X_i, i \in N\}$ with state space φ
- state vector $< Evt, Qtr, Time.PtsDiff, a, h >$
- simulations generated using a random walk over state space
- transition probabilities conditional on a game context
- particularly useful for in-play betting

Common metrics

Brier score[2]

$$BS = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{R}(p_{ij} - o_{ij})^2$$

— does not consider the outcomes to be ordinal

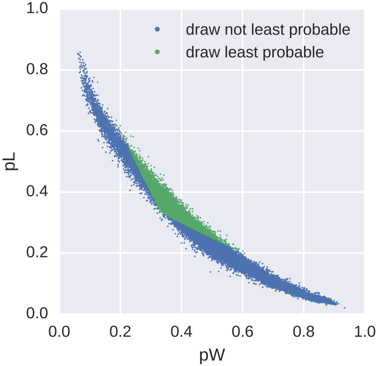
Ranked probability score[5]

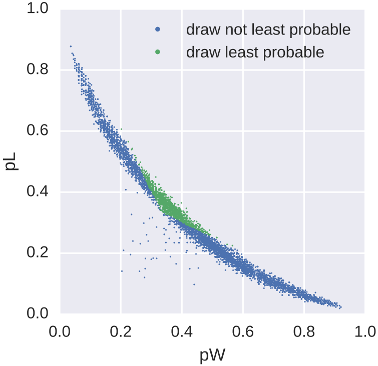
$$RPS = \frac{1}{R-1}\sum_{i=1}^{R}\sum_{j=1}^{i}(p_j - o_j)^2$$

+ does consider the outcomes to be ordinal

— does consider the outcomes to be ordinal

Ordinality of outcomes





Exploting betting markets using ML

- focus on profiting from betting market
- core idea: accuracy $\not\approx$ profit
- from gathering the data to evaluating betting strategies
- application of ANNs

Aggregating player-level statistics using convolution

- player-level statistics provide more information
- concatenating player statistics leads to large feature vector
- default team-level stats provide sum/average of players' stats
- convolution allows learning the aggregation function

Soccer Prediction Challenge

- over 200 000 matches from leagues all around the world
- RPS as evaluation metric
- data: League, Season, Date, Home/Away, Home/Away Score
- lot of feature engineering
- gradient boosted trees (xgboost)

Ranking teams using PageRank

- PageRank was originally used for ranking websites
- simulates a random surfer
- our use case: each league can be represented as a graph
- teams $\longrightarrow$ vertices, matches $\longrightarrow$ matches
- weight of the edge equals to number of expected points

Future work

- March Machine Learning Mania @kaggle
- utilize other types of data (play-by-play, pesstatsdatabase, ...)
- Dota 2 drafter
- RNNs/CNNs
- ideas from recommender systems, graph algorithms, ...

Bibliography I

[1] Bradley, R. A., and Terry, M. E.
Rank analysis of incomplete block designs: I. the method of paired comparisons.
Biometrika 39, 3/4 (1952), 324–345.

[2] Brier, G. W.
Verification of forecasts expressed in terms of probability.
Monthey Weather Review 78, 1 (1950), 1–3.

[3] Constantinou, A. C., and Fenton, N. E.
Determining the level of ability of football teams by dynamic ratings based on the relative discrepancies in scores between adversaries.
Journal of Quantitative Analysis in Sports 9, 1 (2013), 37–50.

Bibliography II

[4] Elo, A. E.
The rating of chessplayers, past and present, vol. 3.
Batsford London, 1978.

[5] Epstein, E. S.
A scoring system for probability forecasts of ranked categories.
Journal of Applied Meteorology 8, 6 (1969), 985–987.

[6] Glickman, M. E.
The glicko-2 system for rating players in head-to-head competition, jul. 2000.
Retrieved from the Internet on Oct 23 (2003).

Bibliography III

[7] Herbrich, R., Minka, T., and Graepel, T.
Trueskill™: a bayesian skill rating system.
In *Advances in neural information processing systems* (2007), pp. 569–576.

[8] Vračar, P., Štrumbelj, E., and Kononenko, I.
Modeling basketball play-by-play data.
Expert Systems with Applications 44 (2016), 58–66.