FROM MACHINE LEARNING TO OPTIMIZATION IN SPORTS ANALYTICS

Oliver Schulte
School of Computing Science
Simon Fraser University
Vancouver, Canada
SPORTS ANALYTICS

- Sport
- Tracking
- Evaluation
- Prediction
OVERVIEW

Success Probabilities

Action Values

Player Ranking

Learning Success Probabilities

Optimization

Defining Success Probabilities

A fundamental problem in sports analytics

Success Probabilities ↓ Action Values

Action Values ↓ Player Ranking

Classifier, Reinforcement Learning

Finding optimal players, tactics, wagers
SUCCESS PROBABILITIES
WHAT IS SUCCESS?

• An outcome (binary event) that a team wants to bring about.
• Can be defined according to the interest of the analyst/coach/player.

Success Notion

- Winning the Game
- Scoring within a fixed time/number of steps
- Scoring Within a Possession
- Scoring the Next Goal
- Zone Entry
- Penalty for Opponent
- Scoring
SUCCESS PROBABILITY TICKER

- Assigns to each time $t$ in the match an estimated probability of future success
- Example: Win probability in NHL (Pettigrew 2015)

9th Annual MIT Sloan Sports Analytics Conference
SCORE IN POSSESSION

http://www.lukebornn.com/sloan_epv_curve.mp4
NEXT GOAL PROBABILITIES

- Y-axis: the chance of scoring the next goal

FROM SUCCESS PROBABILITIES TO ACTION VALUES
ACTION VALUES

• Success probabilities can be used to evaluate players and actions
• impact(action) = [success probability after action – success probability before action]
• “We assert that most questions that coaches, players, and fans have about basketball, particularly those that involve the offense, can be phrased and answered in terms of EPV [i.e. expected future success].” Cervone, Bornn et al. 2014.
• The impact of an action $a_t$ performed at time $t>0$ is the difference in successive success probabilities:

$$\text{impact}(a_t) = p_t - p_{t-1}$$

* = THOR baseline
FROM ACTION IMPACT TO PLAYER RANKINGS
work (Pettigrew, 2015; Cervone et al., 2014). We compare to those of the league-average player, similar to previous.

alent to comparing the actions taken by a specific player son impact score for the player. This procedure is equiv-
game, and then over a single season, to compute a net sea-
sum the impact scores of a player's actions over a single
action to the player as they perform the action. Next, we
To calculate player valuations, we apply the impact of an
players performing the action, yielding player valuations.

As players perform actions on behalf of their team, it is

7.2 PLAYER V ALUATIONS

Figure 3: 2013-2014 Player Goal Impact Vs. Season Points

A possible explanation is referees are reluctant to penalize
teams with higher leads (Schuckers and Brozowski, 2012).
the observation that there are more penalties called against

If a trailing team.

the range of action val-

Player performance =
total impact of all actions performed
Jason Spezza: high goal impact, low +/-.
• plays very well on poor team (Ottawa Senators 2013).
• Requested transfer for 2014-2015 season.
• 2015-16 NHL season
• Johnny Gaudreau and Mark Scheifele drew salaries below what their GIM rank would suggest.
• Later they received a $5M+ contract for the 2016-17 season.

<table>
<thead>
<tr>
<th>Name</th>
<th>GIM</th>
<th>Assists</th>
<th>Goals</th>
<th>Points</th>
<th>Team</th>
<th>Salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taylor Hall</td>
<td>96.40</td>
<td>39</td>
<td>26</td>
<td>65</td>
<td>EDM</td>
<td>$6,000,000</td>
</tr>
<tr>
<td>Joe Pavelski</td>
<td>94.56</td>
<td>40</td>
<td>38</td>
<td>78</td>
<td>SJS</td>
<td>$6,000,000</td>
</tr>
<tr>
<td>Johnny Gaudreau</td>
<td>94.51</td>
<td>48</td>
<td>30</td>
<td>78</td>
<td>CGY</td>
<td>$925,000</td>
</tr>
<tr>
<td>Anze Kopitar</td>
<td>94.10</td>
<td>49</td>
<td>25</td>
<td>74</td>
<td>LAK</td>
<td>$7,700,000</td>
</tr>
<tr>
<td>Erik Karlsson</td>
<td>92.41</td>
<td>66</td>
<td>16</td>
<td>82</td>
<td>OTT</td>
<td>$7,000,000</td>
</tr>
<tr>
<td>Patrice Bergeron</td>
<td>92.06</td>
<td>36</td>
<td>32</td>
<td>68</td>
<td>BOS</td>
<td>$8,750,000</td>
</tr>
<tr>
<td>Mark Scheifele</td>
<td>90.67</td>
<td>32</td>
<td>29</td>
<td>61</td>
<td>WPG</td>
<td>$832,500</td>
</tr>
<tr>
<td>Sidney Crosby</td>
<td>90.21</td>
<td>49</td>
<td>36</td>
<td>85</td>
<td>PIT</td>
<td>$12,000,000</td>
</tr>
<tr>
<td>Claude Giroux</td>
<td>89.64</td>
<td>45</td>
<td>22</td>
<td>67</td>
<td>PHI</td>
<td>$9,000,000</td>
</tr>
<tr>
<td>Dustin Byfuglien</td>
<td>89.46</td>
<td>34</td>
<td>19</td>
<td>53</td>
<td>WPG</td>
<td>$6,000,000</td>
</tr>
<tr>
<td>Jamie Benn</td>
<td>88.38</td>
<td>48</td>
<td>41</td>
<td>89</td>
<td>DAL</td>
<td>$5,750,000</td>
</tr>
<tr>
<td>Patrick Kane</td>
<td>87.81</td>
<td>60</td>
<td>46</td>
<td>106</td>
<td>CBJ</td>
<td>$13,800,000</td>
</tr>
<tr>
<td>Mark Stone</td>
<td>86.42</td>
<td>38</td>
<td>23</td>
<td>61</td>
<td>OTT</td>
<td>$2,250,000</td>
</tr>
<tr>
<td>Blake Wheeler</td>
<td>85.83</td>
<td>52</td>
<td>26</td>
<td>78</td>
<td>WPG</td>
<td>$5,800,000</td>
</tr>
<tr>
<td>Tyler Toffoli</td>
<td>83.25</td>
<td>27</td>
<td>31</td>
<td>58</td>
<td>DAL</td>
<td>$2,600,000</td>
</tr>
<tr>
<td>Charlie Coyle</td>
<td>81.50</td>
<td>21</td>
<td>21</td>
<td>42</td>
<td>MIN</td>
<td>$1,900,000</td>
</tr>
<tr>
<td>Tyson Barrie</td>
<td>81.46</td>
<td>36</td>
<td>13</td>
<td>49</td>
<td>COL</td>
<td>$3,200,000</td>
</tr>
<tr>
<td>Jonathan Toews</td>
<td>80.92</td>
<td>30</td>
<td>28</td>
<td>58</td>
<td>CHI</td>
<td>$13,800,000</td>
</tr>
<tr>
<td>Sean Monahan</td>
<td>80.92</td>
<td>36</td>
<td>27</td>
<td>63</td>
<td>CGY</td>
<td>$925,000</td>
</tr>
<tr>
<td>Vladimir Tarasenko</td>
<td>80.68</td>
<td>34</td>
<td>40</td>
<td>74</td>
<td>STL</td>
<td>$8,000,000</td>
</tr>
</tbody>
</table>
CORRELATIONS WITH STANDARD STATS

- GIM: our ranking (goal impact metric)
- Takeaway: high correlation with standard stats
- e.g. 0.93 with points

<table>
<thead>
<tr>
<th>methods</th>
<th>Assist</th>
<th>Goal</th>
<th>GWG</th>
<th>OTG</th>
<th>SHG</th>
<th>PPG</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>+/-</td>
<td>0.236</td>
<td>0.204</td>
<td>0.217</td>
<td>0.16</td>
<td>0.095</td>
<td>0.099</td>
<td>0.118</td>
</tr>
<tr>
<td>GAR</td>
<td>0.527</td>
<td>0.633</td>
<td>0.552</td>
<td>0.324</td>
<td>0.191</td>
<td>0.583</td>
<td>0.549</td>
</tr>
<tr>
<td>WAR</td>
<td>0.516</td>
<td>0.652</td>
<td>0.551</td>
<td>0.332</td>
<td>0.192</td>
<td>0.564</td>
<td>0.532</td>
</tr>
<tr>
<td>EG</td>
<td>0.783</td>
<td>0.834</td>
<td>0.704</td>
<td>0.448</td>
<td>0.249</td>
<td>0.684</td>
<td>0.891</td>
</tr>
<tr>
<td>SI</td>
<td>0.869</td>
<td>0.745</td>
<td>0.631</td>
<td>0.411</td>
<td>0.27</td>
<td>0.591</td>
<td>0.898</td>
</tr>
<tr>
<td>GIM-T1</td>
<td>0.873</td>
<td>0.752</td>
<td>0.682</td>
<td>0.428</td>
<td>0.291</td>
<td>0.607</td>
<td>0.877</td>
</tr>
<tr>
<td>GIM</td>
<td>0.875</td>
<td>0.878</td>
<td>0.751</td>
<td>0.465</td>
<td>0.345</td>
<td>0.71</td>
<td>0.912</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>methods</th>
<th>Point</th>
<th>SHP</th>
<th>PPP</th>
<th>FOW</th>
<th>P/GP</th>
<th>TOI</th>
<th>PIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>+/-</td>
<td>0.237</td>
<td>0.159</td>
<td>0.089</td>
<td>-0.045</td>
<td>0.238</td>
<td>0.141</td>
<td>0.049</td>
</tr>
<tr>
<td>GAR</td>
<td>0.622</td>
<td>0.226</td>
<td>0.532</td>
<td>0.16</td>
<td>0.616</td>
<td>0.323</td>
<td>0.089</td>
</tr>
<tr>
<td>WAR</td>
<td>0.612</td>
<td>0.235</td>
<td>0.531</td>
<td>0.153</td>
<td>0.605</td>
<td>0.331</td>
<td>0.078</td>
</tr>
<tr>
<td>EG</td>
<td>0.854</td>
<td>0.287</td>
<td>0.729</td>
<td>0.28</td>
<td>0.702</td>
<td>0.722</td>
<td>0.354</td>
</tr>
<tr>
<td>SI</td>
<td>0.869</td>
<td>0.37</td>
<td>0.707</td>
<td>0.185</td>
<td>0.655</td>
<td>0.955</td>
<td>0.492</td>
</tr>
<tr>
<td>GIM-T1</td>
<td>0.902</td>
<td>0.384</td>
<td>0.736</td>
<td>0.288</td>
<td>0.738</td>
<td>0.777</td>
<td>0.347</td>
</tr>
<tr>
<td>GIM</td>
<td>0.93</td>
<td>0.399</td>
<td>0.774</td>
<td>0.295</td>
<td>0.749</td>
<td>0.835</td>
<td>0.405</td>
</tr>
</tbody>
</table>
LEARNING SUCCESS PROBABILITY MODELS

Try this at home
MACHINE LEARNING APPROACH

Deployment

Match Context at time t → Context Feature Vector at time t → Classifier → Success Probability at time t

Training

Observed Match Contexts → Observed Context Feature Vectors → Machine Learning → Classifier

EVENT DATA

• Illustrate approaches with event data
  • available from nhl.com
  • also pre-crawled
• Less work on tracking data
  • Sportlogiq.com is Montreal-based data provider
# Dataset Statistics 2015-16

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Teams</td>
<td>30</td>
</tr>
<tr>
<td>Number of Players</td>
<td>2,233</td>
</tr>
<tr>
<td>Number of Games</td>
<td>1,140</td>
</tr>
<tr>
<td>Number of Events</td>
<td>3.3M</td>
</tr>
</tbody>
</table>
SPECIFYING LOOK-AHEAD

- Suppose we start with event data
- For each time $t$, add a success column $Y_t$ depending on whether the team succeeded after time $t$.
  - E.g., did they score the next goal?
- Could also annotate whether they score the next goal in $k$ steps (Descroo et al. 2019) or fixed time interval (Shuckers and Curro 2013 THoR)


M. Schuckers and J. Curro, “Total Hockey Rating (THoR): A comprehensive statistical rating of National Hockey League forwards and defensemen based upon all on-ice events,” 2013. 7th Annual MIT Sloan Sports Analytics Conference
<table>
<thead>
<tr>
<th>game</th>
<th>player</th>
<th>Period</th>
<th>team</th>
<th>x</th>
<th>y</th>
<th>Manpower</th>
<th>Action</th>
<th>Y = next goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>849</td>
<td>402</td>
<td>1</td>
<td>15</td>
<td>-9.5</td>
<td>1.5</td>
<td>Even</td>
<td>Recovery</td>
<td>0</td>
</tr>
<tr>
<td>849</td>
<td>402</td>
<td>1</td>
<td>15</td>
<td>-24.5</td>
<td>-17</td>
<td>Even</td>
<td>Carry</td>
<td>0</td>
</tr>
<tr>
<td>849</td>
<td>417</td>
<td>1</td>
<td>16</td>
<td>-75.5</td>
<td>-21.5</td>
<td>Even</td>
<td>Check</td>
<td>1</td>
</tr>
<tr>
<td>849</td>
<td>402</td>
<td>1</td>
<td>15</td>
<td>-79</td>
<td>-19.5</td>
<td>Even</td>
<td>Pass</td>
<td>0</td>
</tr>
<tr>
<td>849</td>
<td>413</td>
<td>1</td>
<td>16</td>
<td>-92</td>
<td>-32.5</td>
<td>Even</td>
<td>Turnover</td>
<td>1</td>
</tr>
<tr>
<td>849</td>
<td>413</td>
<td>1</td>
<td>16</td>
<td>-92</td>
<td>-32.5</td>
<td>Even</td>
<td>Pass</td>
<td>1</td>
</tr>
<tr>
<td>849</td>
<td>389</td>
<td>1</td>
<td>16</td>
<td>-98</td>
<td>0</td>
<td>Even</td>
<td>Goal</td>
<td>1</td>
</tr>
</tbody>
</table>
• Context can be represented as a feature vector
  • E.g. score differential, manpower differential
• What do to about the previous match history?
• Simple Approach:
  • Fix a sliding window size $k$ (common values are 3, 4, 10).
  • Use previous $k$ events context for current event
  • [Statsbomb approach](#)
### EXAMPLE: SLIDING WINDOW

**k=2**

<table>
<thead>
<tr>
<th>Manpower</th>
<th>Action</th>
<th>Y=next goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Even</td>
<td>Turnover</td>
<td>1</td>
</tr>
<tr>
<td>Even</td>
<td>Pass</td>
<td>1</td>
</tr>
<tr>
<td>Even</td>
<td><strong>Goal</strong></td>
<td>1</td>
</tr>
</tbody>
</table>

**Table: Sliding Window Example**

<table>
<thead>
<tr>
<th>MP(-2)</th>
<th>Action(-2)</th>
<th>MP(-1)</th>
<th>Action(-1)</th>
<th>Manpower(0)</th>
<th>Action(0)</th>
<th>Y=next goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>Even</td>
<td>Turnover</td>
<td>1</td>
</tr>
<tr>
<td>*</td>
<td>*</td>
<td>Even</td>
<td>Turnover</td>
<td>Even</td>
<td>Pass</td>
<td>1</td>
</tr>
<tr>
<td>Even</td>
<td>Turnover</td>
<td>Even</td>
<td>Pass</td>
<td>Even</td>
<td><strong>Goal</strong></td>
<td>1</td>
</tr>
</tbody>
</table>
• Given a list of pairs (context vector, success target), can simply run any classifier (R/Weka/Scikit-Learn)
  • E.g. logistic regression, gradient boosted decision trees
• Excerpt from logistic model tree: success probability of home team
• Alternative Approaches to Classification:
  • Recurrent Neural Networks (handles sequences)
  • Reinforcement Learning (value function learning)
    • My own work, RealAnalytics

CP 2023
LEARNING SUCCESS PROBABILITIES: OVERVIEW

1. estimate transition probabilities
2. apply dynamic programming

- Hockey
  Schulte et al. 2017
  Ljung et al. 2019

- NFL
  Chan and Puterman 2019

- continuous data
  space + time
  discretize

- raw data
  classifier or reinforcement learning
  Classifier

- sliding window
  Decroos et al. 2019
  recurrent NN

- neural net
  temporal difference learning
  model-free

- Hockey
  Liu et al. 2018

- Soccer
  Fernandez et al. 2019
OPTIMIZATION PROBLEMS
OPTIMIZATION MAKES SPORTS ANALYTICS ACTIONABLE

Sports Data → Machine Learning Model (represents domain knowledge) → Optimization → Selection/Decision
FANTASY LEAGUE/HOCKEY POOL

- See Tauhid Zaman’s talk.
- You “draft” a team of \( m \) players (the lineup).
- Your players earn points through the season (goals + assists in real games)
- At the end of the season the “manager” with the most points wins the pool.
- Variations:
  - Players take turns drafting
  - add constraints on selections, e.g. at least 1 goalie, at most 4 forwards.
  - Lemmer (2013) applies integer programming with constraints
  - Estimated 35M people play fantasy sports in North America (Becker and Sun 2016)

REFERENCES

- https://www.officepools.com/
- https://www.sloansportsconference.com/event/fantasy-sports-analytics MIT Sloan Sports Analytics has talks and panels on fantasy play
• The drafting problem but for real teams and players
• Commercial Software: https://octothorpesoftware.com
• Optimal Lineups for a specific opponent seems to be a new problem
  • e.g. optimal lineup in Cricket for England against India vs. against Australia
• Perera et al. study optimal lineups against average opponents

TACTICS: FINDING OPTIMAL ACTIONS

- Basketball: go for 3 points (risky) or 2 points (safer)
- Basketball: doubling (e.g. put 2 defenders on LeBron James) Wang et al. 2018
- Hockey: when to pull the goalie? Beaudoin and Swartz 2010
- Often involves counterfactual questions: What if we tried a tactic that has never been tried before?
  - Related to off-line reinforcement learning

FINDING OPTIMAL BETS

- Single bet scenario. Given:
  - Bankroll (e.g. $100)
  - Win probability $p$ (from ML model)
  - Odds $b$ from bookmaker (e.g. $b=2$ if 2-1 return on winning bet)
- Output: fraction $f$ of bankroll to be wagered
- *Kelly Criterion:* The maximum expected payoff is achieved by $f^* = p - (1-p)/b$

- Example
  - $b = 2, p=1/2$ (even strength)
  - $F^* = 1/2 - (1/2/2) = 1/2 - 1/4 = 1/4$
  - With $100$ bankroll, you should bet $25$
EXTENSIONS

- Add constraints, e.g.
  - Maximum on bets
  - No risk of ruin (losing entire bankroll)
- Place bets with different bookmakers
- Bets on different matches and at different times (betting lines move) Insley et al. 2004
  - Related to portfolio management Lien et al. 2023

CONCLUSION

• Learning to predict success probabilities is a fundamental task for sports analytics
• Powerful approach to action values and player ranking
• Different machine learning models can be used
  • Classification, recurrent neural networks, reinforcement learning
• Machine learning provides the domain knowledge
• Optimization makes the domain knowledge actionable
  • Optimizing players, tactics, wagers
THANK YOU!
BACKUPS
REINFORCEMENT LEARNING

See My Aggregate Intellect Talk
STATE TRANSITION PROBABILITIES

- Step 1: Estimate the probabilities of getting one from match state to the other
- **Basketball Demo**
- In our discrete NHL models, we estimated state transition probabilities for 1.3M states
- Step 2: Estimate the chances of reaching a success state using *dynamic programming*
MULTI-STEP TRANSITION PROBABILITIES

To compute $P_{l+1}(s_1|s_0)$: the probability of reaching state $s_1$ from $s_0$

Answer: $0.1 \times 0.2 + 0.3 \times 0.5 + 0.6 \times 0.3$
DYNAMIC PROGRAMMING

- **Input**: State Transition Probabilities
- **Output**: Probability of Future Success for *every* match state
- For lookahead $L = 1,...$
  - Compute probability of success in $L+1$ steps using
    - 1-step state transitions and $L$ step success probabilities from previous lookahead
  - Terminate at convergence or at fixed bound
- For the NHL, our computation converged at $L = 13$
- [Xthreat Visualization](#)

CP 2023
MONTE CARLO VS. TEMPORAL DIFFERENCE

- Target = final outcome
- “Monte Carlo Learning”
- E.g. if a possession ends in a goal, then outcome = target$_t$ = 1
- Standard with sports analysts
  - Leverages supervised methods (e.g. classifiers)
  - Ignores temporal dependencies and dynamics

- Target$_t$ = $v_{t+1}$
- “Temporal Difference Learning”
- Connects predictions at different times and for different actions
- Standard method in RL

University of Waterloo 2022
## TOY EXAMPLE

<table>
<thead>
<tr>
<th>Current Match State</th>
<th>Action</th>
<th>Current Score (Goal)</th>
<th>Estimated Next Goal Chance</th>
<th>TD-target</th>
<th>TD-error</th>
<th>MC-target</th>
<th>MC-error</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1-0, even strength, DZ]</td>
<td>Carry</td>
<td>0</td>
<td>55%</td>
<td>62%</td>
<td>(55%-62%)²</td>
<td>1</td>
<td>(55%-100%)²</td>
</tr>
<tr>
<td>[1-0, powerplay, OZ, …]</td>
<td>Pass</td>
<td>0</td>
<td>62%</td>
<td>75%</td>
<td>(62%-75%)²</td>
<td>1</td>
<td>(62%-100%)²</td>
</tr>
<tr>
<td>[1-0, powerplay, OZ, …]</td>
<td>Shot</td>
<td>0</td>
<td>75%</td>
<td>1</td>
<td>(75%-100%)²</td>
<td>1</td>
<td>(75%-100%)²</td>
</tr>
<tr>
<td>[2-0, ES]</td>
<td>Face-off</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>