DEVELOPING ANALYTICS FOR DEVELOPMENT LEAGUES

Joshua Weissbock

Ottawa Analytics
7 February 2015

Outline

- Introduction to CHLStats.com
- Technologies
- Features
- Data Sources
- Issues
- Projects
- Future Statistics
  - Goalie Statistics
- Conclusion

Introduction to CHLStats.com

- Background in Machine Learning

- Blogging on CanucksArmy.com

- Interest in Draft Analytics and Prospect Prediction

- Developed CHLStats.com to further access to statistics at

  the major junior level

  - "Extra Skater" of Prospect Stats

Technologies

- Language: Python 2.7
- Framework: Flask 0.10
- Scraping: BeautifulSoup 4 with lxml
- Database: sqlite3
- Server side: wsgi on nginx
- Host: PythonAnywhere

Features – Team Statistics

OHL 2014-2015 Statistics

<< 2013-2014

Team	Est. Fenwick Close	GP	SF	SA	GF	GA	SOW	SOL	Sh%	Sv%	PDO
Oshawa Generals	58.022%	49	1812	1312	204	115	2	2	11.16	91.374	102.534
North Bay Battalion	56.91%	47	1602	1211	159	127	1	4	9.869	89.809	99.678
Erie Otters	56.578%	47	1664	1329	227	145	4	2	13.434	89.224	102.658
Sault Ste. Marie Greyhounds	55.954%	48	1853	1492	231	154	2	2	12.372	89.799	102.171
Kitchener Rangers	55.65%	47	1685	1388	152	158	3	6	8.859	89.001	97.86

- Est. Fenwick Close (B. Wendorf's 2p Sf%)
- Team Shooting + Save Percentage
- PDO

Features – Player Statistics

- eTOI (Tulsky)
- ePoints / 60
- PPG
- Team Point %
- Team Goal %
- ES GF%
- Relative ES GF%

- Age (in fractional years)
- Age Adjusted Points/Game
- NHLe
- IPP
- QMJHL:
  - SOG, Sh%, SOG/G, Dangerous Shots, DS%, FO%

Data Sources

- Parsing Game Sheets daily; No "RTSS" data

Credit: M. Pfeffer

Issues

- Poorly written HTML (harder to parse)
- Future changes to HTML will likely break tools
- Post-game changes to game data, difficult to adapt
- Slow websites (WHL.ca)
- Loading data with AJAX
- Untrustworthy data (e.g. height, scorer bias)
- Leagues will change game IDs (WHL)
- Partial Data (WHL shots)
- On ice data states varies between arena
- Player names and spelling is not always consistent
- Reordering names, player names with spaces
  - e.g. Vande Sompel, Mitchell. "Mitchel Vande Sompel"
  - e.g. Fournier, Jordan Ty. "Jordan Ty Fournier"

Future Statistics

- Statistics currently available to be implemented:
  - Player QoC, QoT
  - Vollman's Usage Charts Variants
  - Hockey-Reference's Goals Created
  - WAR
  - ES Scoring
  - Common Line Mates / Pairing Partners
  - Biological Metrics (Height / Weight)
- Team Level
  - Penalty Differentials
  - Pythag / Expected Win %
  - Strength of Schedule
- Goalie Data: Vollman's QS%, GSAA

Projects

- Machine Learning Game Predictions

  - Higher upper bound (~70%) vs NHL (64%)

- Drafting Inefficiencies

  - "Sham Sharron"

  - PPG vs CSS Ranks

- Height + PPG Success Binning

  - Credit (@MoneyPuck_)

- Era/Age Adjusting Points/Game

  - Credit: Rhys Jessop (@Thats_Offside)

Summary

- Analytics are evolving in other fields
- Code applicable to most North American Leagues
- Ideas applicable to most leagues
- Should use to identify players to further scout for drafting

Questions

CHLStats.com

@joshweissbock