Applications of Machine Learning in DOTA2: Literature Review and Practical Knowledge Sharing

Daniil Yashkov, Peter Romov, Kirill Neklyudov,

Aleksander Semenov and Daniil Kireev

ML for E-Sport

- Huge amount of data, collected automatically every day
- Data is clean
- It is a rapidly growing industry
- Over $150 million market

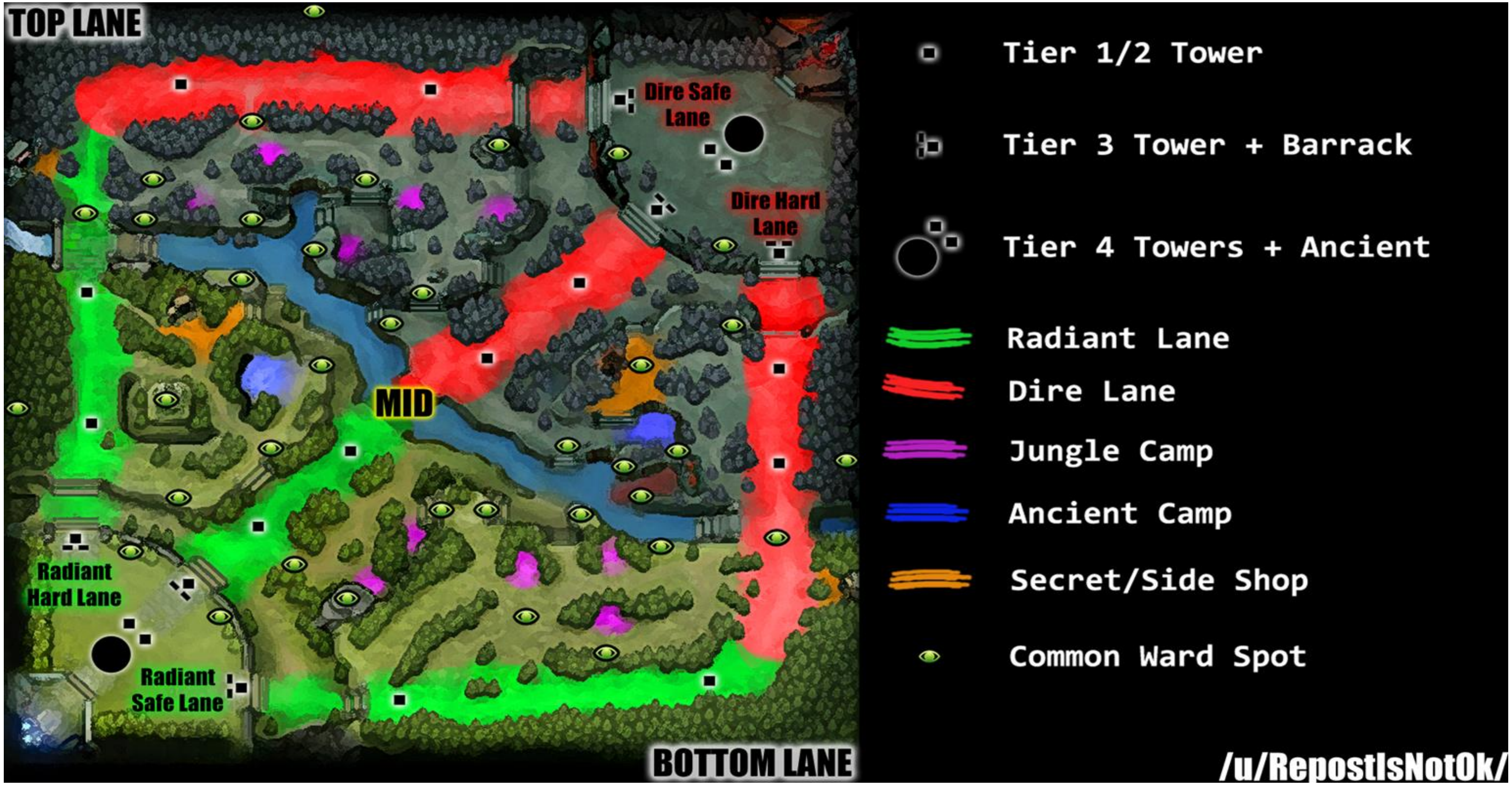
Multiplayer Online Battle Arena (MOBA): Dota 2

- 2 teams, each one formed of 5 players
- 1st stage – draft stage :

  players from every team choose their heroes
- 2nd stage – each team is aimed at destroying “Ancient” building of the enemy
- During the game each player improve their heroes, gaining gold, experience, killing enemies, buying items, etc.

  All this data is logging and collecting.

Multiplayer Online Battle Arena (MOBA): Dota 2



Data analysis in Dota 2

- Win prediction :
    - at the start of the game
    - after draft stage
    - real-time
- Actions/strategies recommendations for players
- Player ranking
- Smart camera for commentators
- …

Draft stage win prediction

- Input data

	Normal Skill	High Skill	Very High Skill	Total
Captains Mode	33,037	5,599	8,840	47,476
Random Draft	86,472	15,560	39,407	141,439
Ranked All Pick	2,937,087	917,001	1,028,855	4,882,943
Total	3,056,596	938,160	1,077,102	5,071,858

- Match = 5 heroes for each team out of 113
- Target: win or lose? Whose pick is better?

Big variety of matches

What is different in matches:

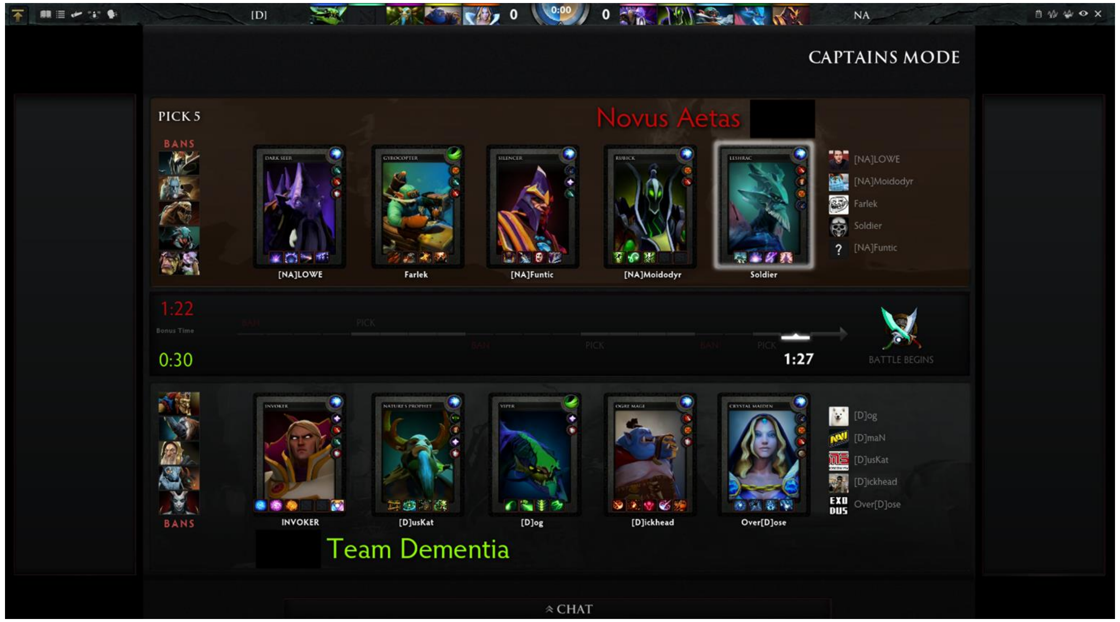
1. Players and their strategies
2. Picked heroes.

Total amount of combinations

$$C_{113}^{5} \times C_{108}^{5} \approx 1.5 \cdot 10^{16}$$

Matches played since 2013 $\approx 2.2 \cdot 10^{9}$

Each player choose one hero



113 heroes pool

Algorithms

- Features – 113 “hero” features for each team.

$$f_i = 1, \text{ if } i^{th} \text{ hero is picked by this team}$$

- Algortihms:
  - Xgboost
  - Factorization machines
  - Logistic regression

2nd order factorization model

$$P(w|\text{teams}) = \sigma\left(\beta_0 + \boldsymbol{\beta}^T\boldsymbol{x} + \sum_i \sum_{j>i} x_i x_j \boldsymbol{v}_i^T \boldsymbol{v}_j\right)$$

Results

Skill / Method	Normal auc	Normal log_loss	High auc	High log_loss	Very High auc	Very High log_loss
libFM	0.706	0.898	0.670	0.933	0.660	0.940
XGBoost	0.701	0.903	0.664	0.937	0.654	0.944
XGBoost_roles	0.702	0.902	0.663	0.938	0.653	0.945
LogReg	0.687	0.916	0.656	0.943	0.643	0.952
LogReg_BoW	0.688	0.915	0.656	0.943	0.643	0.952
NaiveBayes	0.685	0.917	0.653	0.945	0.641	0.954
Dummy	0.500	0.996	0.500	0.999	0.500	0.999

- <u>Set of picked heroes explains at least</u>
  - **6%** of information(Shannon) for very high skill players
  - **10%** of information for normal skill

YASP dataset

- Timeseries of heroes features (points every 30s) such as:
  - Gold
  - Experience
  - Items (purchasing)
  - Abilities

- heroes trajectories (coordinates on map)
- Special buildings(such as tower) states (destroyed or not)

Task:

- Predict winner using first 5 minutes of match

Data:

≈120 000 preprocessed matches

- Final task for ML course as Kaggle In-class competition
- One of the most popular kaggle in-class contest:

  650 solo competitors (teams were not allowed)
- A lot of different ideas, special features
- Very good feedback

Knowledge • 602 teams

Dota 2: Win Probability Prediction

Thu 4 Feb 2016 — Sun 1 Jan 2017 (8 months to go)

Dashboard

Public Leaderboard - Dota 2: Win Probability Prediction

This leaderboard is calculated on all of the test data.

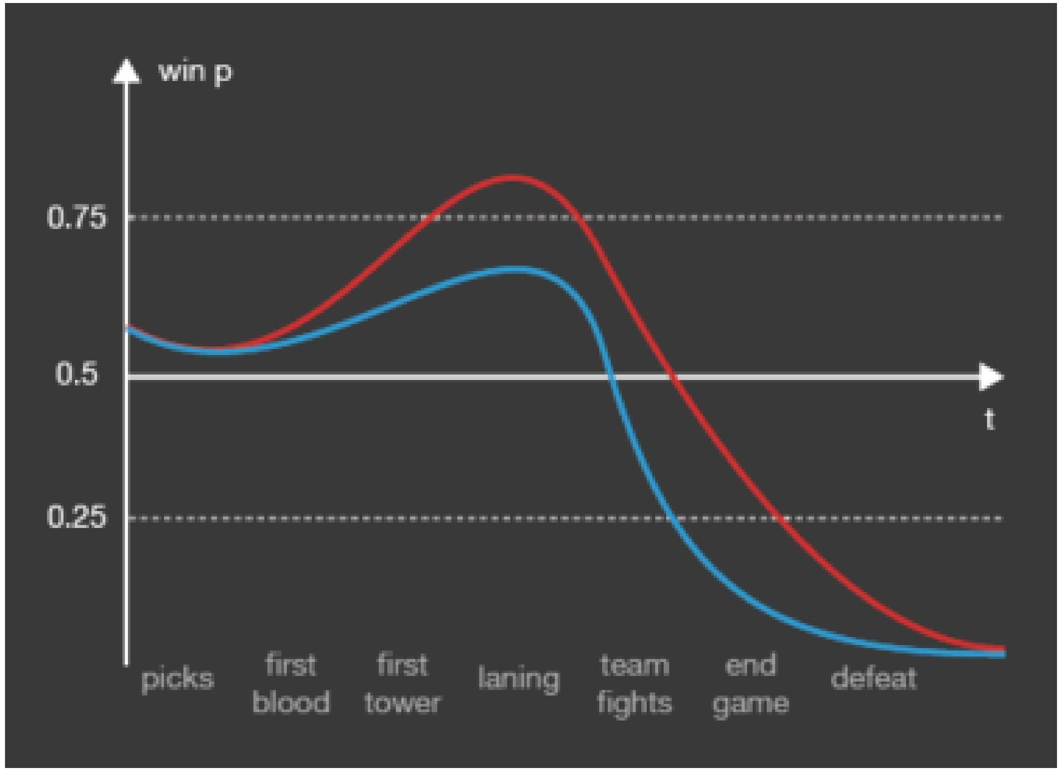
See someone using multiple accounts? Let us know.

#	Δ1w	Team Name	Score	Entries	Last Submission UTC (Best – Last Submission)
1	—	Daniil Kireev	0.80649	18	Sun, 27 Mar 2016 18:58:20
2	—	zz	0.77029	8	Wed, 23 Mar 2016 15:09:57
3	↑1	AlexeyKozulin	0.77001	104	Sat, 09 Apr 2016 07:09:55
4	↓1	SDil	0.76999	12	Tue, 29 Mar 2016 18:22:43 (-40.5h)
5	—	rafanaskin	0.76850	42	Sun, 27 Mar 2016 20:53:39
6	↑4	Alexey S	0.76813	51	Thu, 07 Apr 2016 07:39:53 (-23.1h)
7	↓1	sagol	0.76735	120	Fri, 08 Apr 2016 17:01:23
8	↓1	Fanis Kalimullin	0.76652	34	Thu, 31 Mar 2016 12:54:06 (-41.1h)
9	↑8	Sourcerer	0.76617	24	Sat, 02 Apr 2016 16:00:55 (-0.1h)
10	↓2	AlanEdgarovichBasiev	0.76599	62	Sat, 02 Apr 2016 19:17:10 (-2.5d)
11	↓2	Александр Берсенёв	0.76568	61	Sun, 27 Mar 2016 06:52:21
12	↓1	Lim Exp	0.76540	12	Sun, 27 Mar 2016 14:03:01

Winner's solution

1. Use Logistic Regression instead of more complex models (e.g. Random Forest, GBDT)
2. Find good informative features
   - Statistics for each team
   - One-hot encoded picked heroes in the teams
   - First time team used some items (bottle, courier, ward)
   - Often combinations of heroes in the team: pairs and triples (need to be accurately selected, easy to overfit)
   - Aggregated hero characteristics

Realtime win prediction



$$\text{Score} = \log_2 P(\text{winner}) + 1$$

$$\text{Score-Realtime} = \int_{t=t_{\text{start}}}^{t_{\text{finish}}} Score(t)dt$$

https://github.com/romovpa/dotascience-hackathon

Dota Science Challenge

Matches Leaderboard Examples

Predictions

Match #48: Team Secret vs. Team Liquid

Finished (winner: Team Secret)

Real-time Predictions

#	Team	Probabilities	Last Probability	Last Score	Real-time Score
1	hack__mipt_rak		1.0 / 0.0	1.0 / -100	2552.3483625
2	hack__gapes		0.99 / 0.01	0.985500430305 / -5.64385618977	2095.01485171
3	hack__kotiki		1.0 / 0.0	1.0 / -100	2048.68598576
4	hack__solo_ptz_abuze		0.897998983634 / 0.102001016366	0.84478571722 / -2.29334456721	2022.392221
5	hack__zdes_mojet_byt_vasha_reklama		1.0 / 0.0	1.0 / -100	1907.46525961
6	hack__fontain_guards		0.893333333333 / 0.106666666667	0.837270499962 / -2.2288186905	1845.68484529
7	hack__krasnyi_podorojnik		0.9999 / 0.0001	0.999855723282 / -12.2877123795	1794.21518583
8	hack__random		0.995969272756 / 0.00403072724371	0.994173138645 / -6.95474412426	1771.68941885
9	hack__wild_beasts		0.999 / 0.001	0.99855658313 / -8.96578428466	1693.59462411
10	baseline		0.993298803956 / 0.00670119604437	0.990299679401 / -6.22136567069	1644.54186313
11	hack__pudge		0.9 / 0.1	0.847996906555 / -2.32192809489	1409.34013586
12	hack__Liquid_Secret		0.659774449984 / 0.340225550016	0.400044815184 / -0.555436607334	1384.84652669
13	hack__fake_team		0.92121345869 / 0.0787865413065	0.881607393804 / -2.66590698711	1344.67734635
14	hack__32		1.0 / 0.0	1.0 / -100	1305.199309
15	hack__EC.Dota2		0.999 / 0.001	0.99855658313 / -8.96578428466	1131.31437049
16	hack__molodoy_trezini		0.970037540331 / 0.029962459668	0.95611248561 / -4.06070012785	1096.51302341
17	dota2ruhub_poll		0.6602 / 0.3398	0.400975043673 / -0.557242242365	1024.24652247

Hackathon:

- Realtime leaderboard during Shanghai Major
- 35 teams competed
- Usage of external data

External data:

- odds parsed from websites
- Additional data from steam API
- Parsed replays

Summary

- Large dataset of Dota2 matches
- Game outcome prediction using drafts stage

  auc = 0.66 – 0.7 (depending on skill)
- Kaggle In-class contest: win prediciton having first 5 minutes

  auc = 0.8
- Dota Science hackathon – realtime win prediction

  baseline quality practically doubled