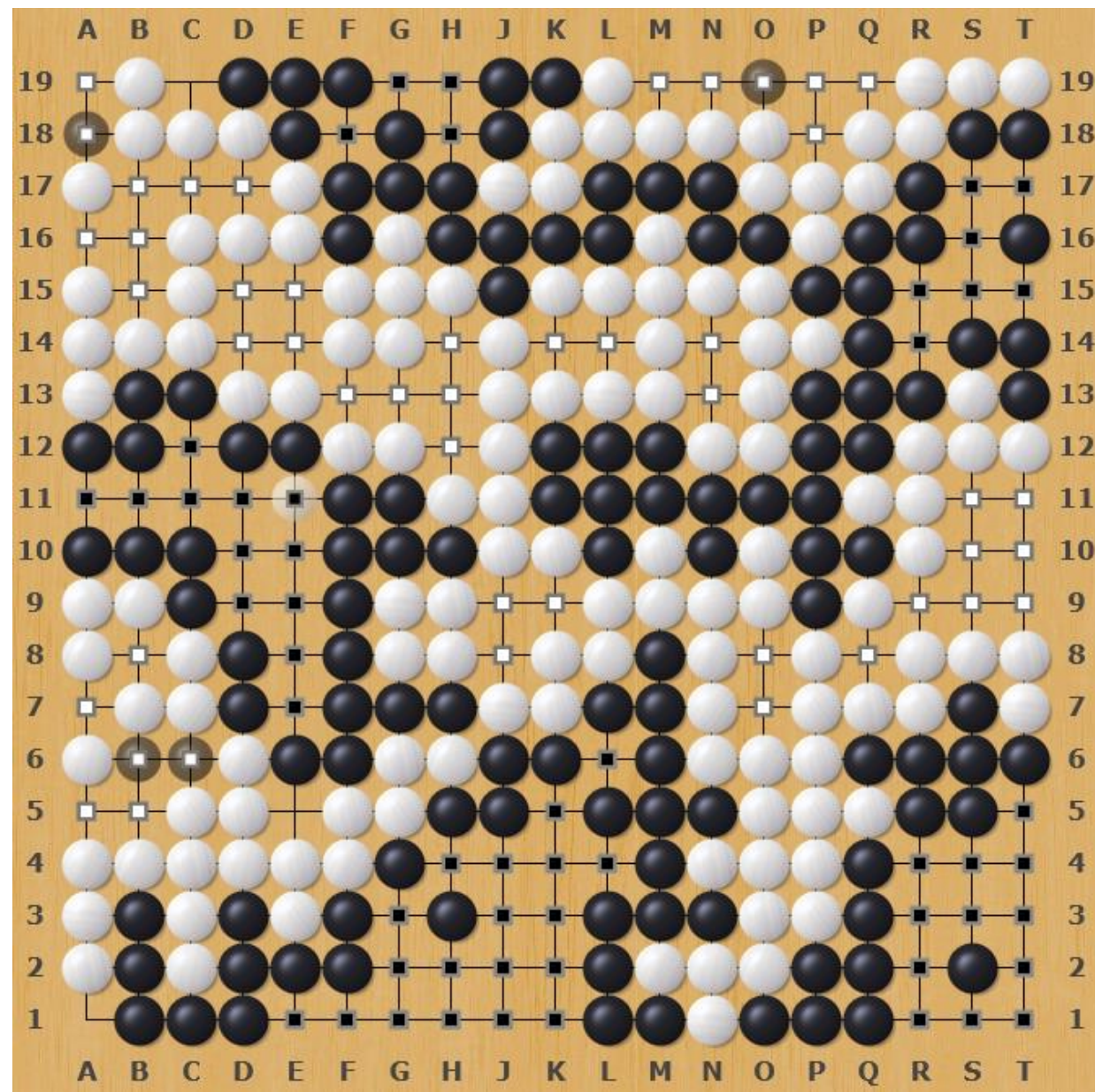


Выявление нечестной игры через анализ временных рядов

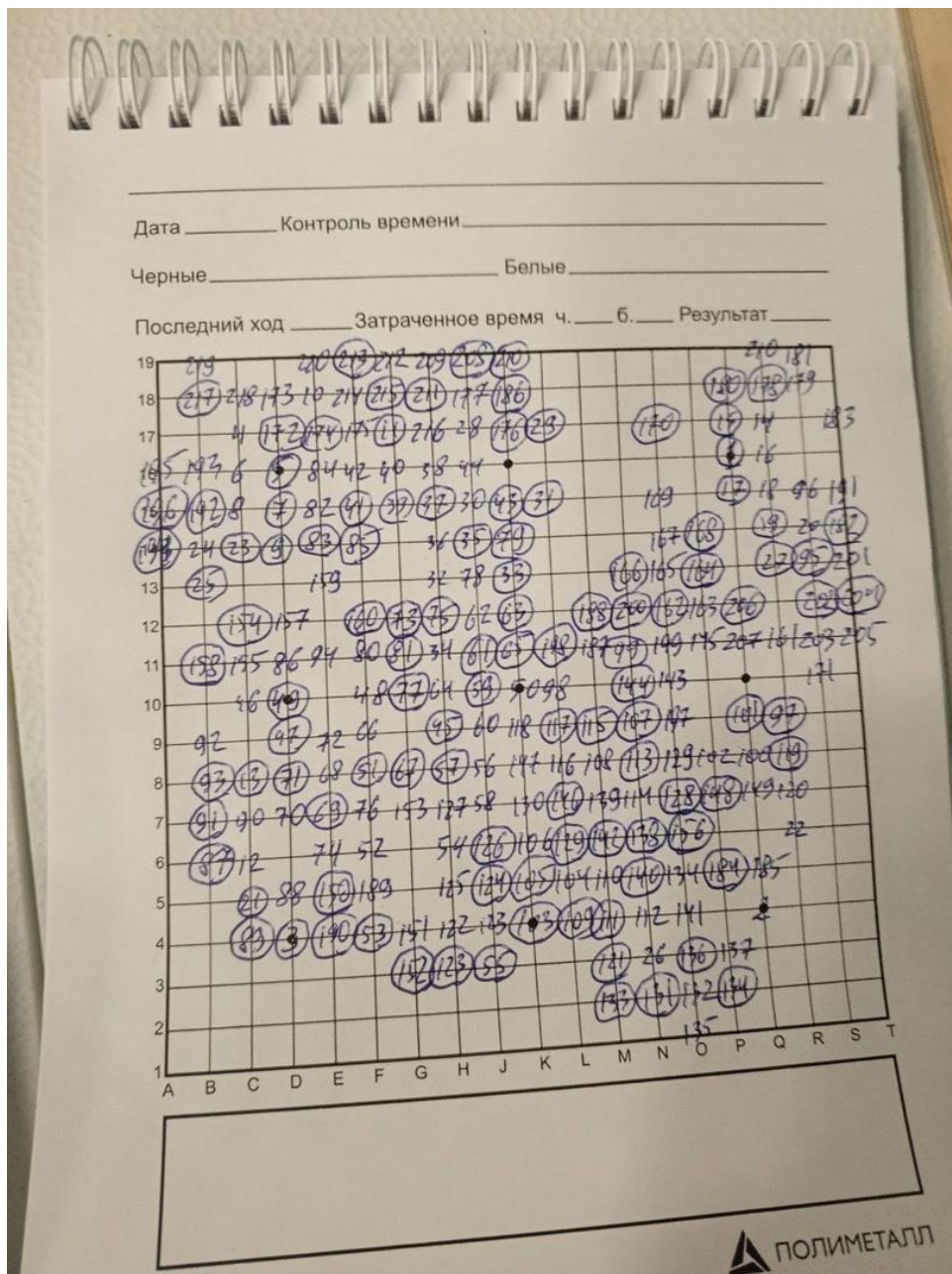
на примере игры го

Предметная область



Постановка задачи

Есть игроки играющие в го. У каждого игрока есть некоторая история – предыдущие партии этого игрока. На основании этого опыта и новых данных (новой партии), требуется определить, сыграл игрок сам (игрок честный) или ему помогал более сильный игрок (игрок обманщик).



Предмет для анализа

Партии бывают в бумажном виде или в виде файлов *.sgf

Мы анализировали записи партий в виде файлов *.sgf

Где взять записи партий?

1) Архив партий профессионалов (порядка 70 000 партий)

<https://homepages.cwi.nl/~aeb/go/games/games/>

2) Брать записи партий отдельных игроков на серверах для онлайн игры: OGS, KGS, FoxWeiqi, Pandanet

3) Обойти всех друзей и спросить у них

```

1  (;
2  EV[Google DeepMind Challenge Match]
3  RO[1]
4  PB[Lee Sedol]
5  BR[9p]
6  PW[AlphaGo]
7  TM[2h]
8  KM[7.5]
9  RE[W+R]
10 DT[2016-03-09]
11 PC[Four Seasons Hotel, Seoul, Korea]
12 RU[Chinese]
13 OT[3x60s byo-yomi]
14
15 ;B[qd];W[dd];B[pq];W[dp];B[fc];W[cf];B[ql];W[od];B[ld];W[qc]
16 ;B[rc];W[pc];B[re];W[of];B[pg];W[og];B[ph];W[id];B[lf];W[oh]
17 ;B[pi];W[lh];B[kh];W[ke];B[le];W[lg];B[kg];W[kf];B[ne];W[oe]
18 ;B[jc];W[ic];B[jd];W[ie];B[je];W[jf];B[if];W[jg];B[li];W[mi]
19 ;B[hf];W[ih];B[mb];W[gd];B[ki];W[mj];B[kk];W[ib];B[ob];W[ml]
20 ;B[lm];W[nc];B[nb];W[kb];B[lc];W[mm];B[ln];W[kl];B[ll];W[lk]
21 ;B[jj];W[jl];B[hj];W[hi];B[gj];W[gf];B[ii];W[jh];B[ij];W[mn]
22 ;B[lo];W[mo];B[lp];W[mp];B[lq];W[mq];B[im];W[qo];B[fq];W[gg]
23 ;B[cn];W[dn];B[dm];W[fp];B[gp];W[gq];B[fr];W[co];B[en];W[do]
24 ;B[ep];W[cm];B[dl];W[lr];B[kr];W[rb];B[jb];W[ja];B[mf];W[mh]
25 ;B[nd];W[qj];B[pj];W[qk];B[pl];W[pk];B[ok];W[rh];B[rl];W[qf]
26 ;B[ri];W[rf];B[pf];W[qe];B[qh];W[cc];B[bn];W[bm];B[bl];W[bo]
27 ;B[rg];W[mr];B[po];W[jr];B[kq];W[pn];B[oo];W[qp];B[on];W[pp]
28 ;B[op];W[qq];B[or];W[pr];B[oq];W[pd];B[qr];W[rr];B[ps];W[rs]
29 ;B[rn];W[ro];B[qn];W[so];B[cl];W[an];B[ks];W[om];B[ol];W[ci]
30 ;B[hh];W[hg];B[dr];W[dj];B[bq];W[cq];B[cr];W[bp];B[dq];W[br]
31 ;B[cp];W[ap];B[ek];W[fi];B[bj];W[bi];B[pb];W[qb];B[sf];W[rd]
32 ;B[ai];W[ah];B[aj];W[bh];B[gi];W[fj];B[fk];W[oc];B[mc];W[cj]
33 ;B[al];W[nm];B[pm];W[aq];B[gh];W[fh]
34 C[Time used: B: 1h32m, W: 1h55m]
35 )
36

```

В каждой партии:

- Игрок белыми (имя и уровень игры)
- Игрок чёрными (имя и уровень игры)
- Дата
- Результат
- Последовательность ходов
- Время, затраченное каждым игроком

Последовательность ходов

- Партия играется на доске 19x19
- У каждого хода есть порядковый номер, а также координаты пункта доски, в который был сделан ход.
- Партия длится от 0 ходов минимум до 361 хода (обычно максимум, но есть исключения).
- Получаем что партия описывается признаками:
8 признаков общих данных + $361 * 2$ признаков ходов.
- Получаем 730-мерный вектор.

Что делать?



НИКОЛАЙ ГАВРИЛОВИЧ

ЧЕРНЫШЕВСКИЙ

ЧТО ДЕЛАТЬ?

*Книга, изменившая мир.
Листовки, объединившие
поколения.*

РУССКАЯ КЛАССИКА

Понижаем размерность!

- Прогоняем файлы *.sgf через нейросеть KataGo. Как итог получаем оценку хода в очках.
- Теперь каждый ход описывается количеством очков, которые теряет игрок, когда делает этот ход

Оценка ходов

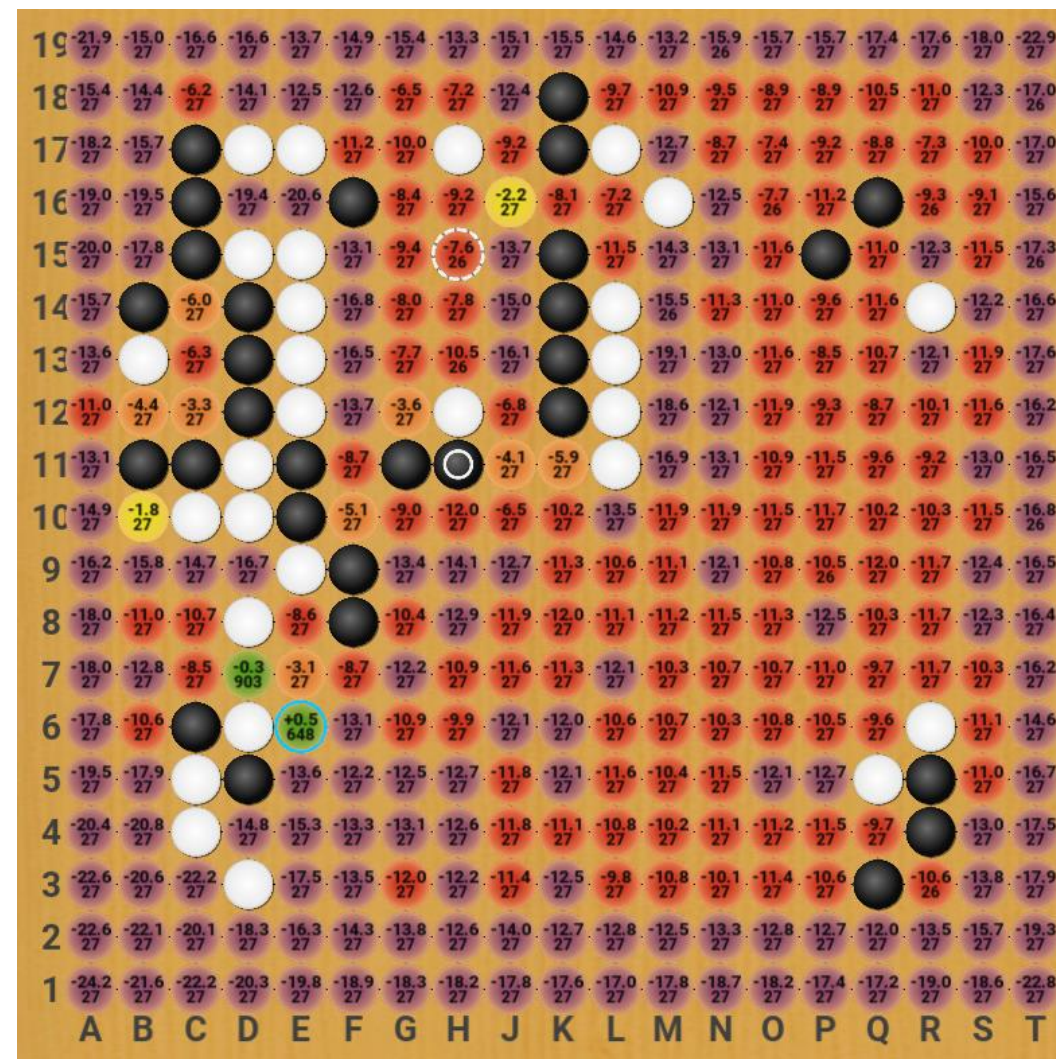



```

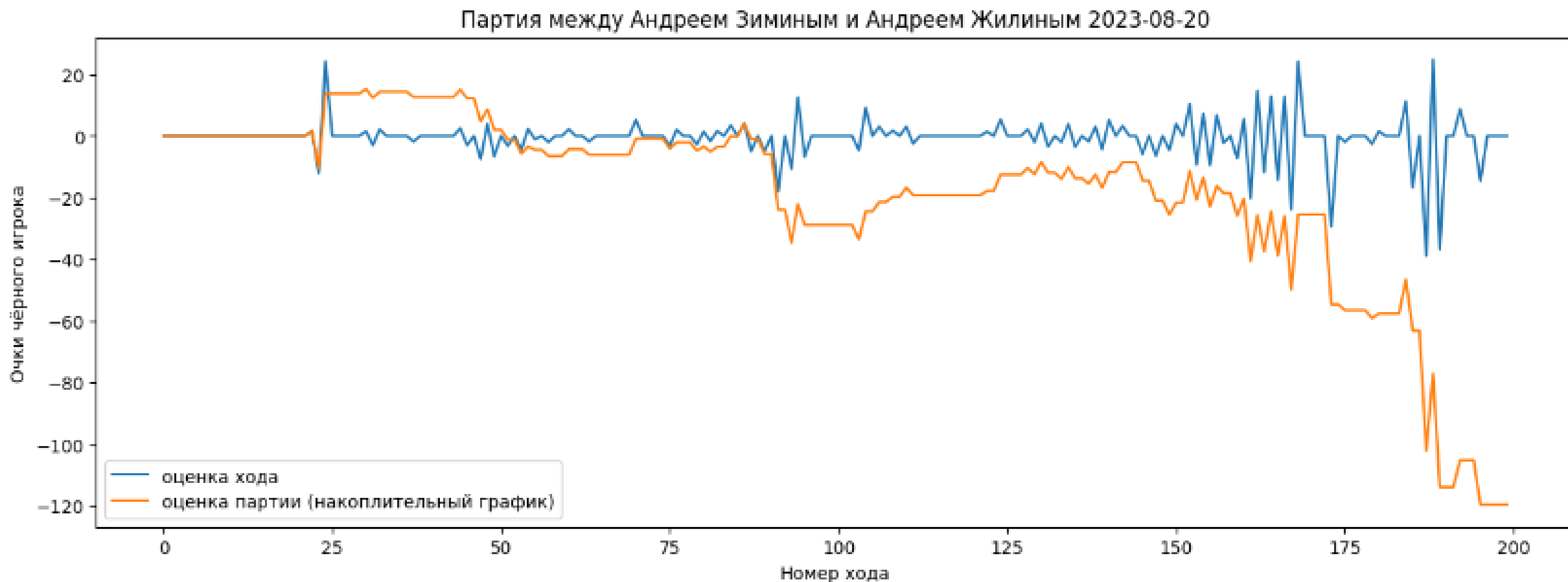
1  (;GM[1]FF[4]SZ[19]
2  AP[ai-sensei.com]
3  CA[UTF-8]ST[2]
4  KM[6.5]RU[japanese]
5  DT[1928-02-29]
6  PW[Go Seigen]
7  PB[Liu Changhua]
8  RE[B+1]
9  C[Mistakes:
10
11  Player   Move    Loss   Score
12
13  White      6     1.5   B+1.8 points
14  White      8     1.1   B+2.3 points
15  Black      9     1.4   B+0.9 points
16  White     12     1.4   B+2.4 points
17  Black     13     1.9   B+0.5 points
18  White     14     1.9   B+2.4 points
19  Black     15     2.1   B+0.3 points
20  White     16     2.1   B+2.4 points
21  White     18     1.2   B+2.8 points
22  Black     19     1.8   B+1.0 points
23  Black     39     2.2   W+0.8 points
24  White     40     1.7   B+0.9 points
25  Black     41     1.9   W+1 points
26  Black     43     1.4   W+2.1 points
27  White     46     1.4   W+0.3 points
28  Black     53     1.0   W+0.9 points
29  White     64     1.3   W+1.3 points
30  Black     69     2.6   W+3.3 points
31  White     70     1.2   W+2.1 points
32  White     74     1.4   W+0.3 points
33  White     76     1.2   B+0.2 points
34  Black     79     2.7   W+1.8 points
35  White     80     1.1   W+0.7 points
36  Black     81     1.2   W+1.9 points
37  White     82     2.3   B+0.4 points

```

1 ход \approx 3 секунды

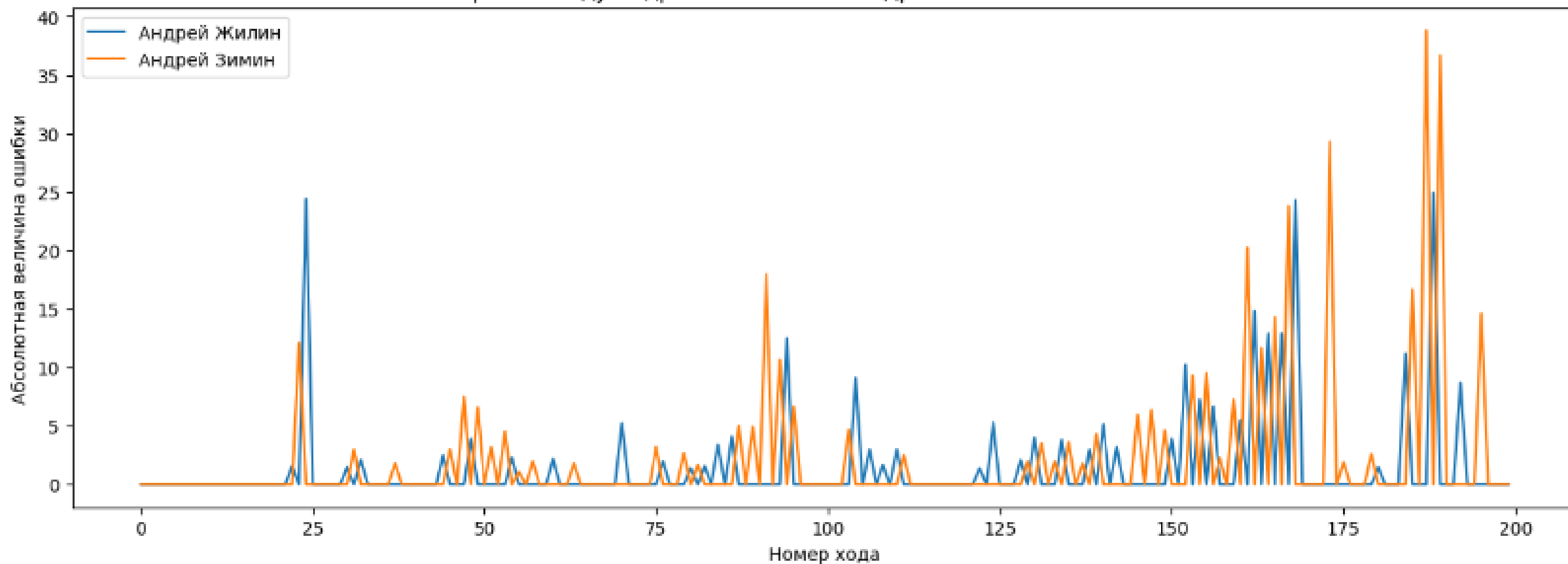


Понижаем размерность!



Понижаем размерность!

Партия между Андреем Зиминым и Андреем Жилиным 2023-08-20

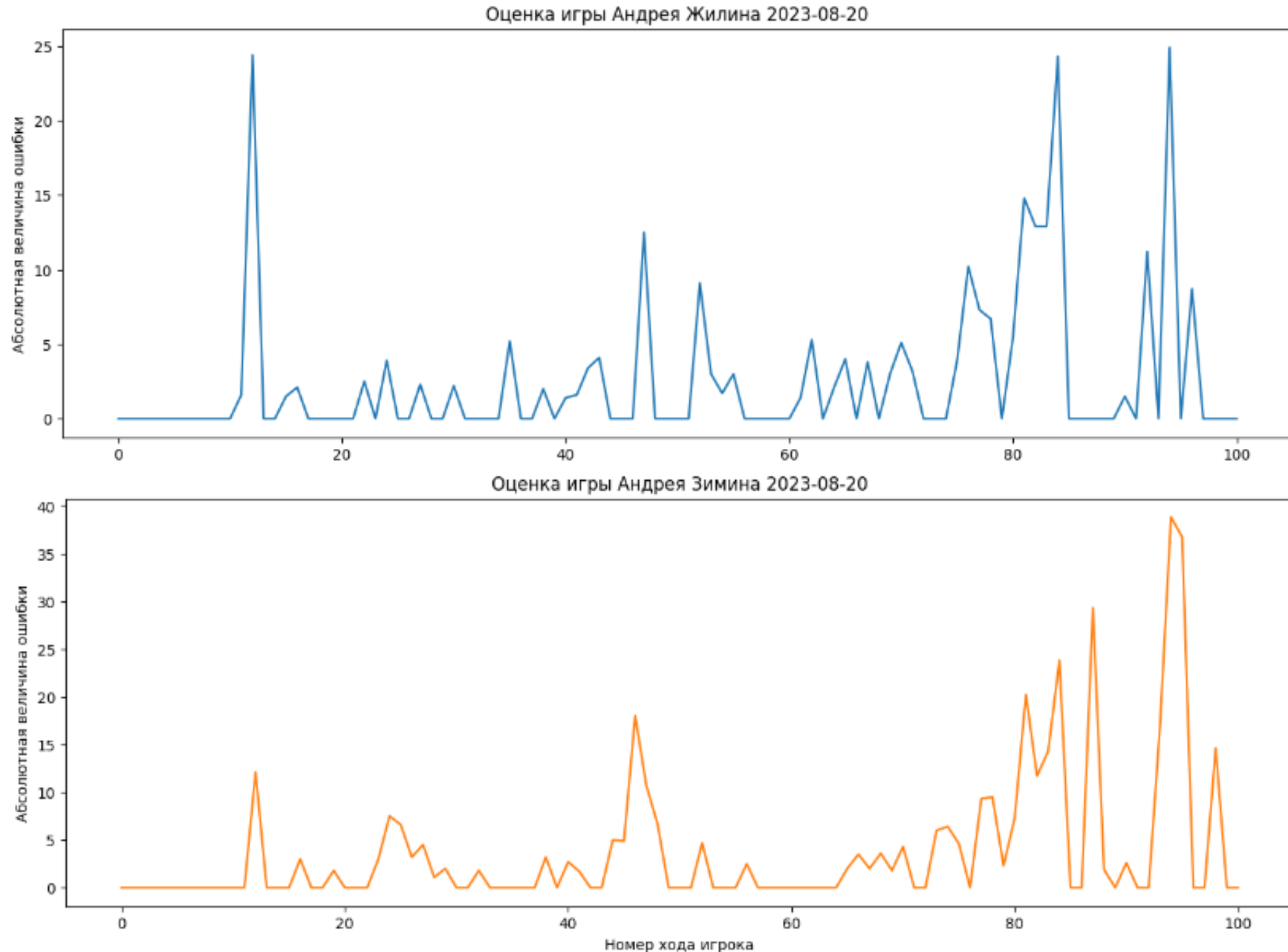


Наивный Андрей

Для упрощения будем считать, что величина ошибки ходов одного игрока не зависит от величины ошибки ходов другого игрока.

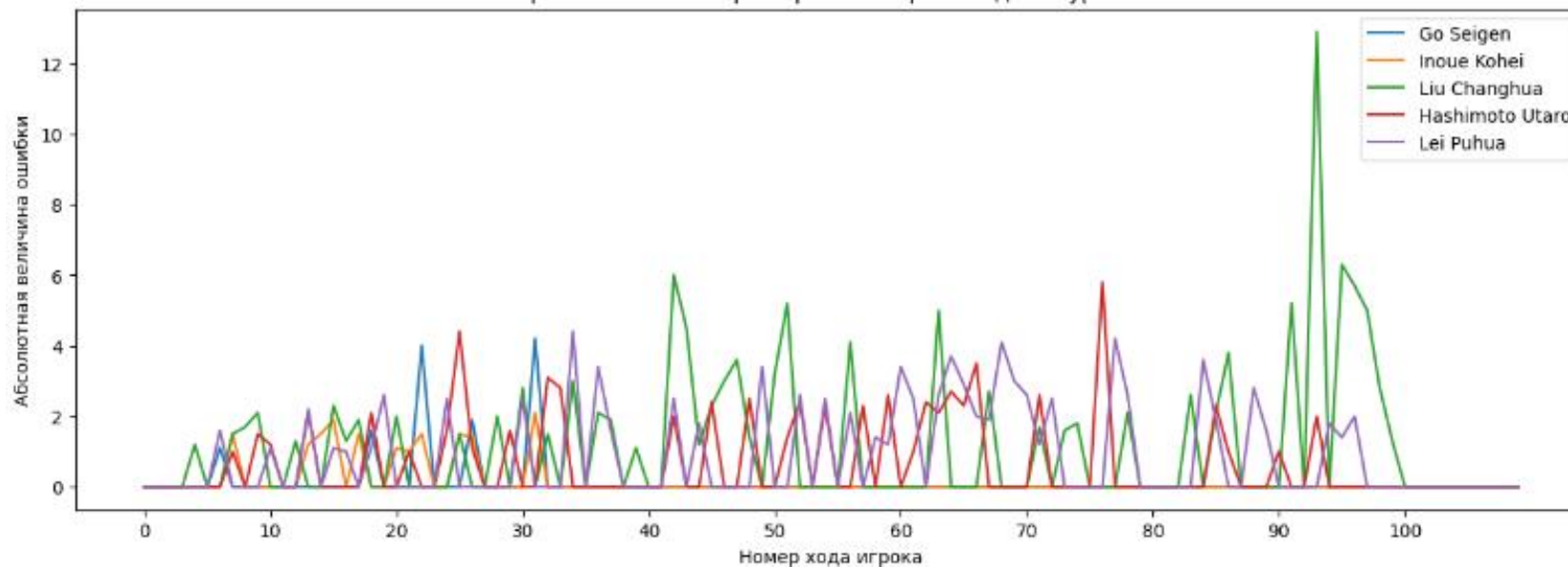
Теперь каждую партию запись в нашем датасете характеризуют:

- Дата
- Имя игрока
- 180 оценок ходов

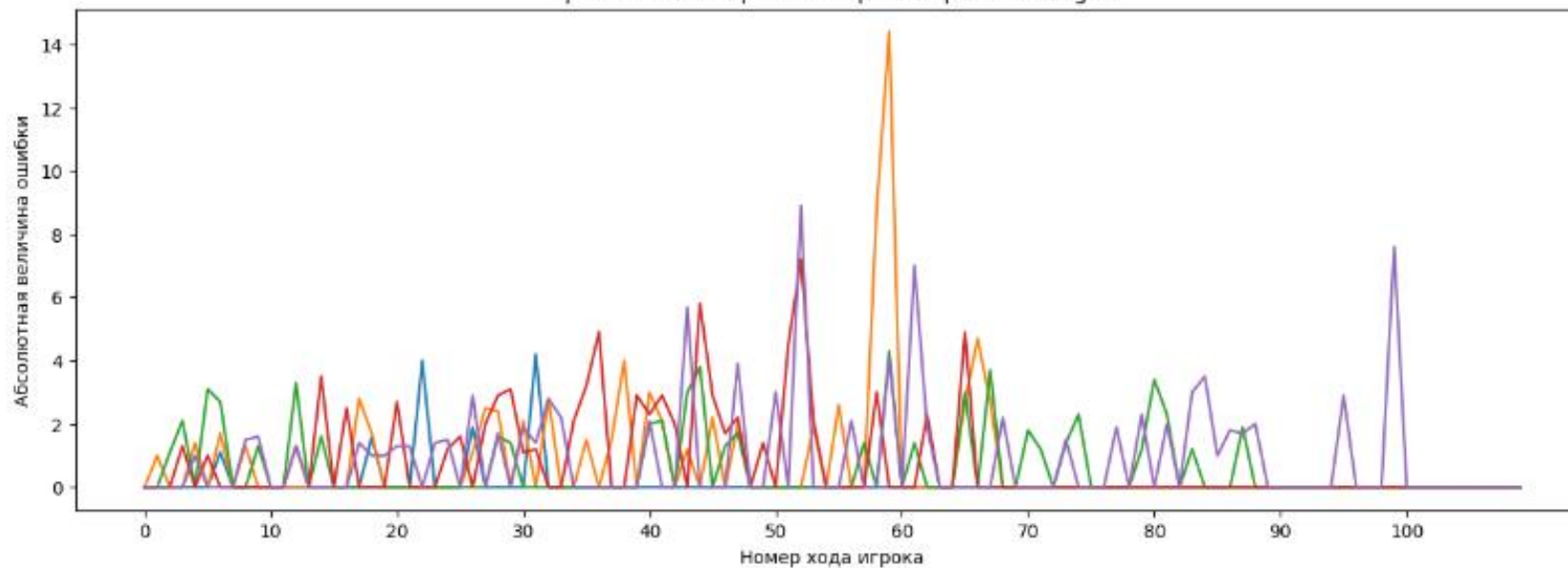


Смотрим на результаты

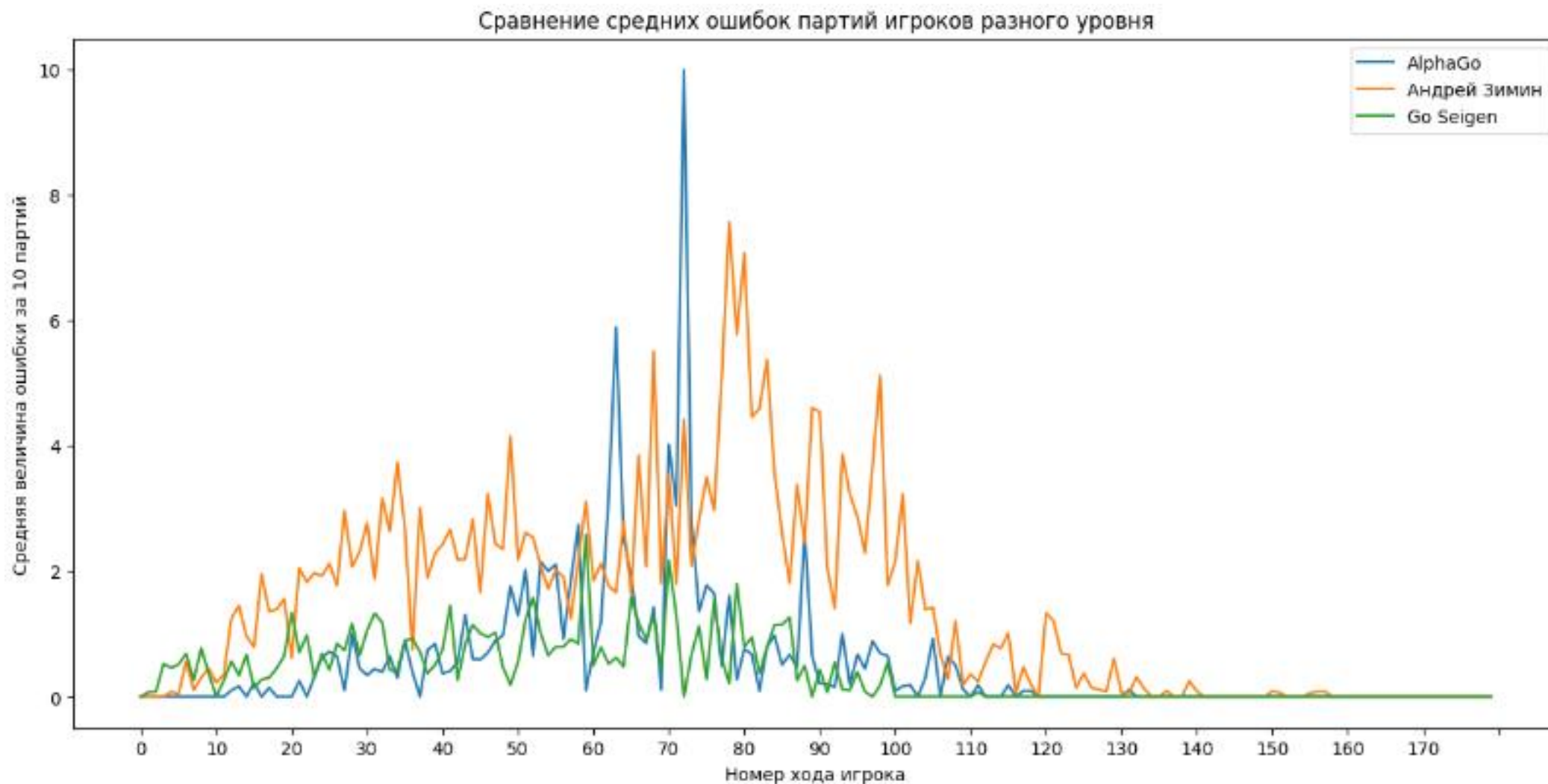
Сравнение пяти партий разных игроков одного уровня



Сравнение пяти разных партий игрока Go Seigen



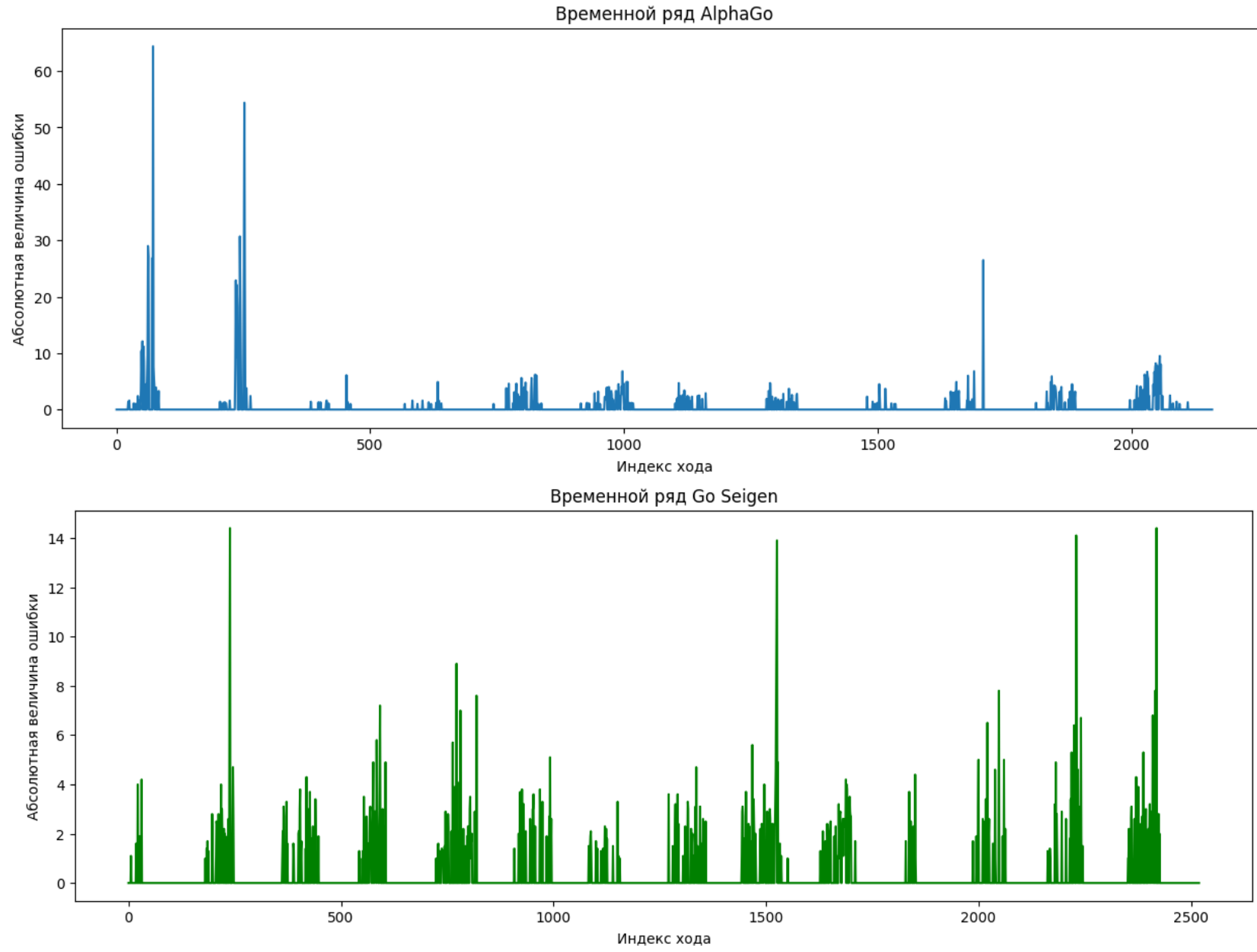
Сравнение игроков разного уровня



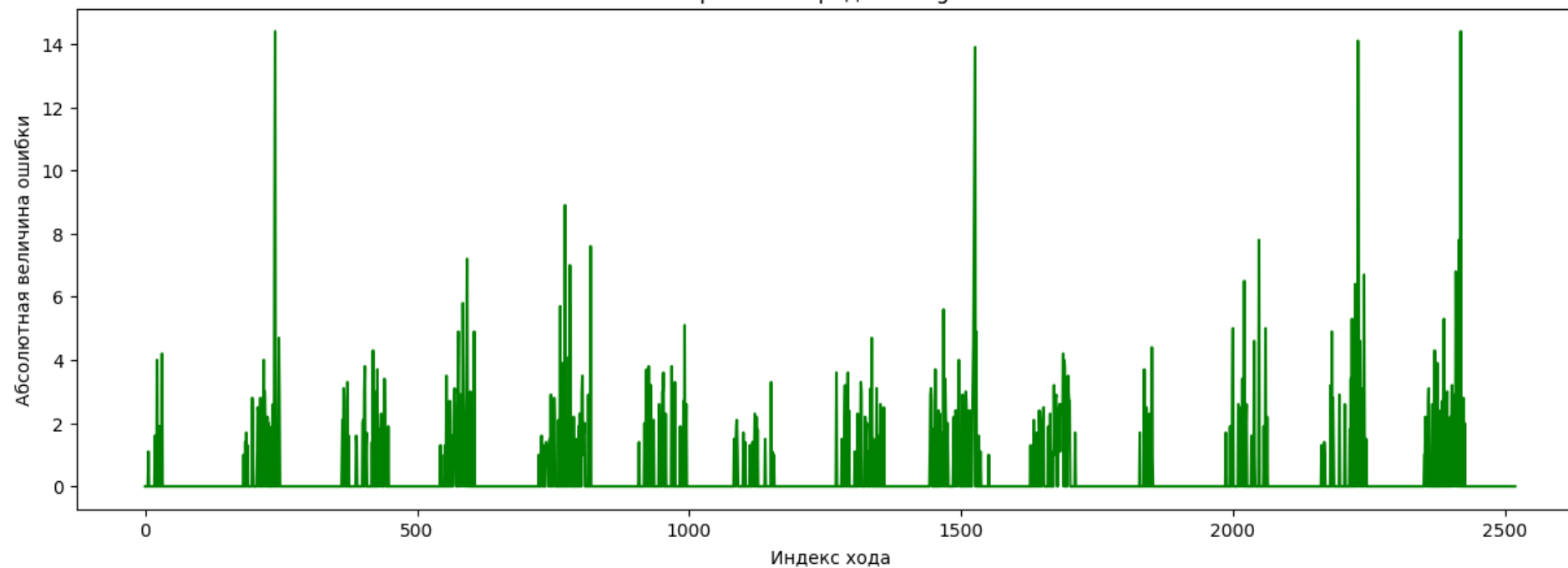
Ряды рядов

- Каждая отдельная партия является временным рядом, где значение – ход (или величина ошибки), а время – номер хода.
- Совокупность партий одного игрока является временным рядом, где значение – отдельная партия, а время – дата проведения партии

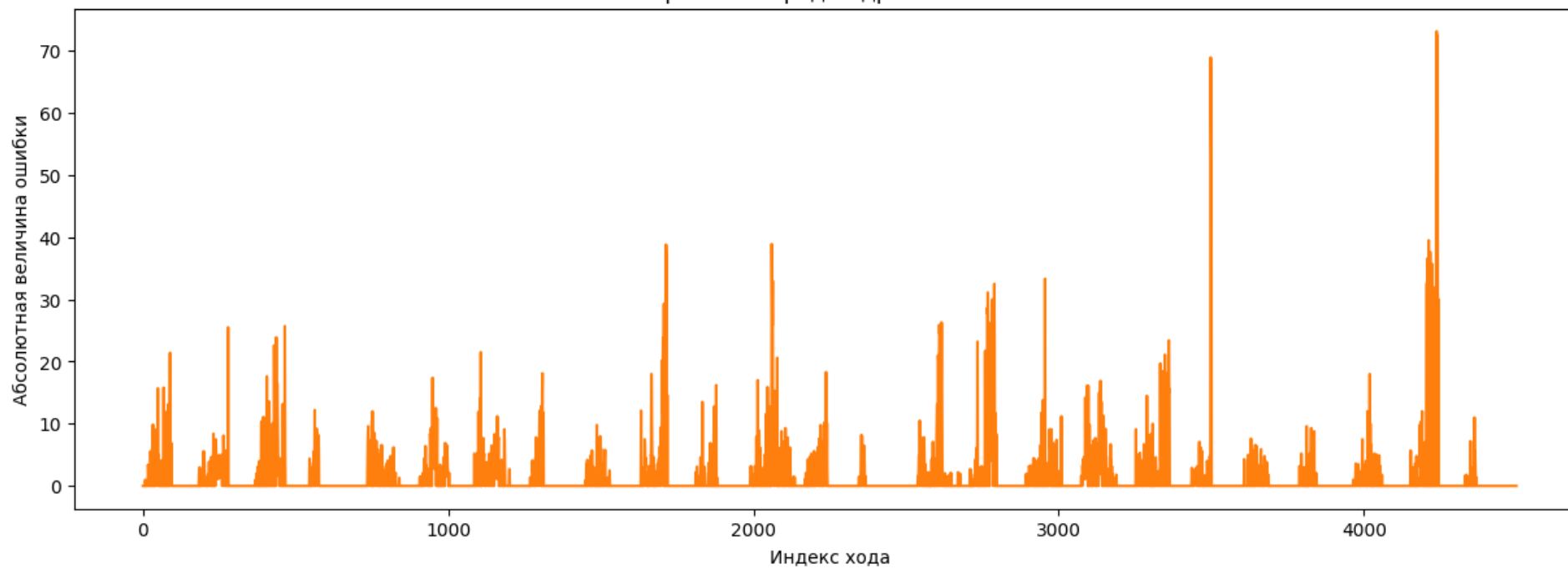
Вытягиваем партии в один временной ряд



Временной ряд Go Seigen



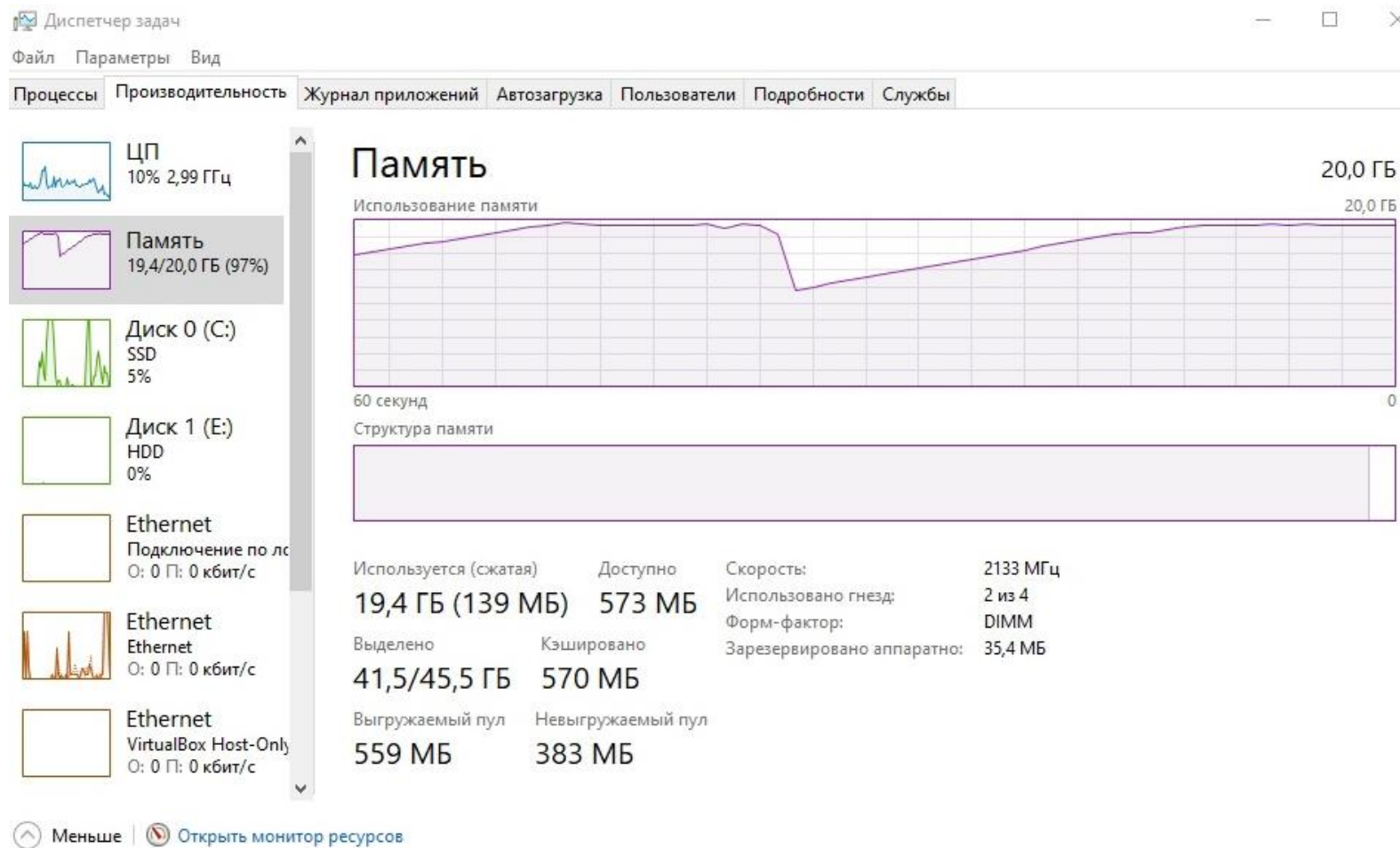
Временной ряд Андрея Зимина



Что осталось?

Тут налицо явно выраженная сезонность и отсутствие тренда, этому запускаем SARIMA и всё просто

Не всё так просто



То что получилось

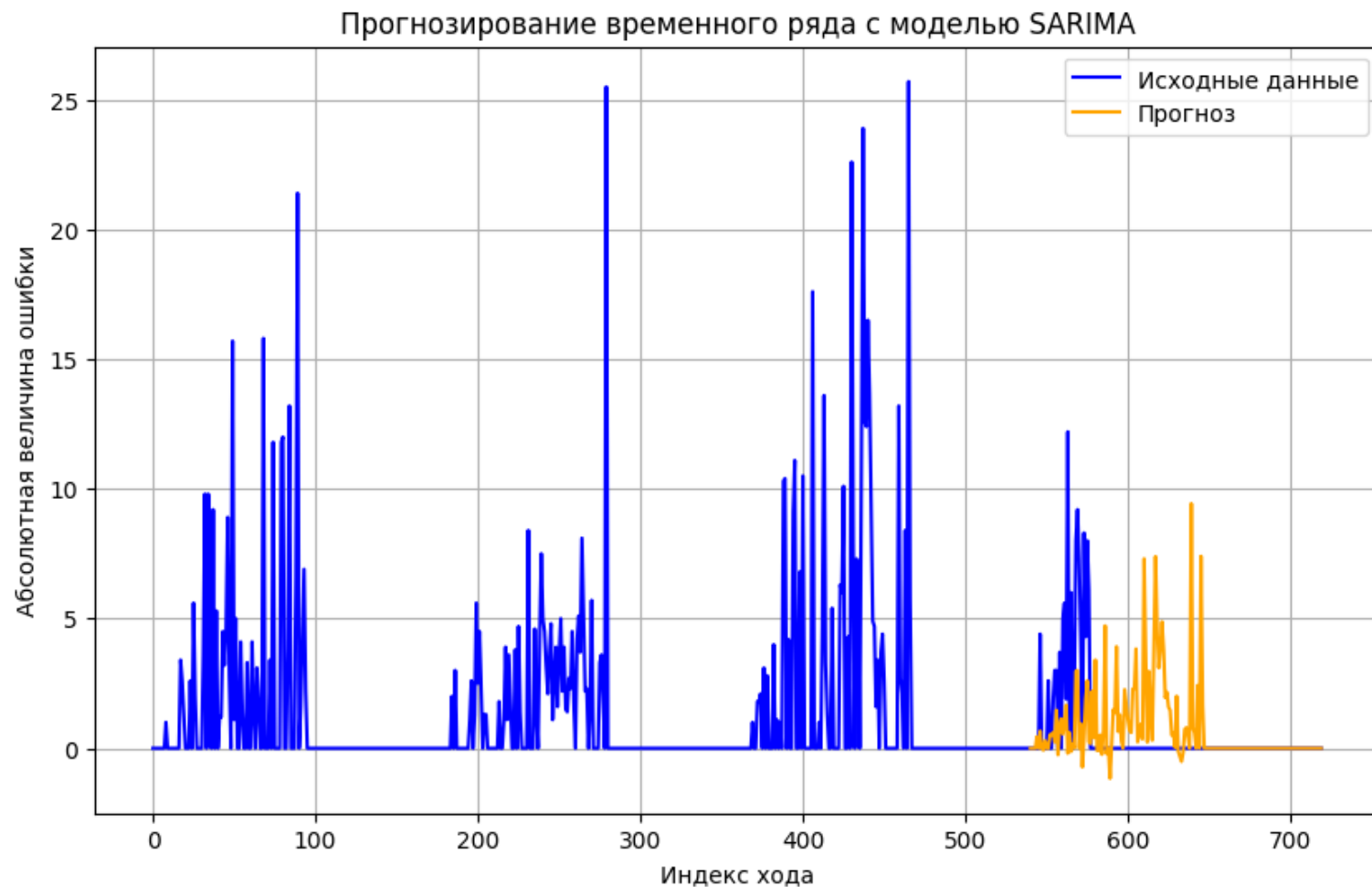
Гиперпараметры:

$p, d, q = 0, 0, 0$

$P, D, Q = 2, 0, 2$

Время работы:

≈ 30 минут



Играл игрок сам или нет?

- Строим временной ряд по его партиям
- Обучаем SARIMA
- Берём новую партию, которой не было в обучающей выборке
- Смотрим на разницу предсказанных и фактических значений.
- Построить 95 процентный коридор для минимального возможного значения ошибки. Если игрок не попал в такой коридор, то считаем что он жульничал

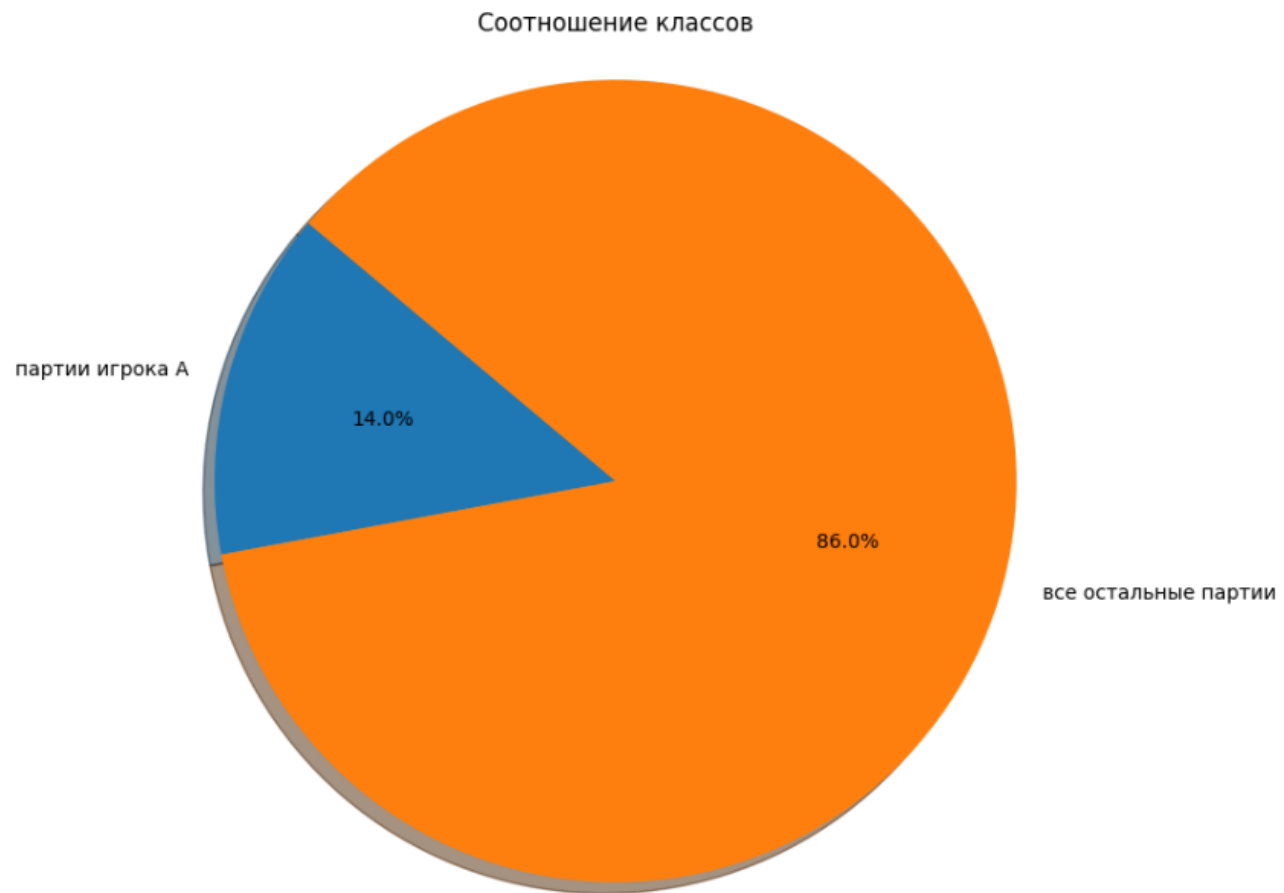
Или попробуем что-то другое

Что если порядок партий не важен?

Случайный лес

- Что если выбрать некоторого игрока A , как и в прошлый раз, но теперь отказаться от временного признака
- Каждую партию будем рассматривать как объект, где 180 признаков
- Target = партия не принадлежит игроку A

Относительно некоторого игрока



Метрики

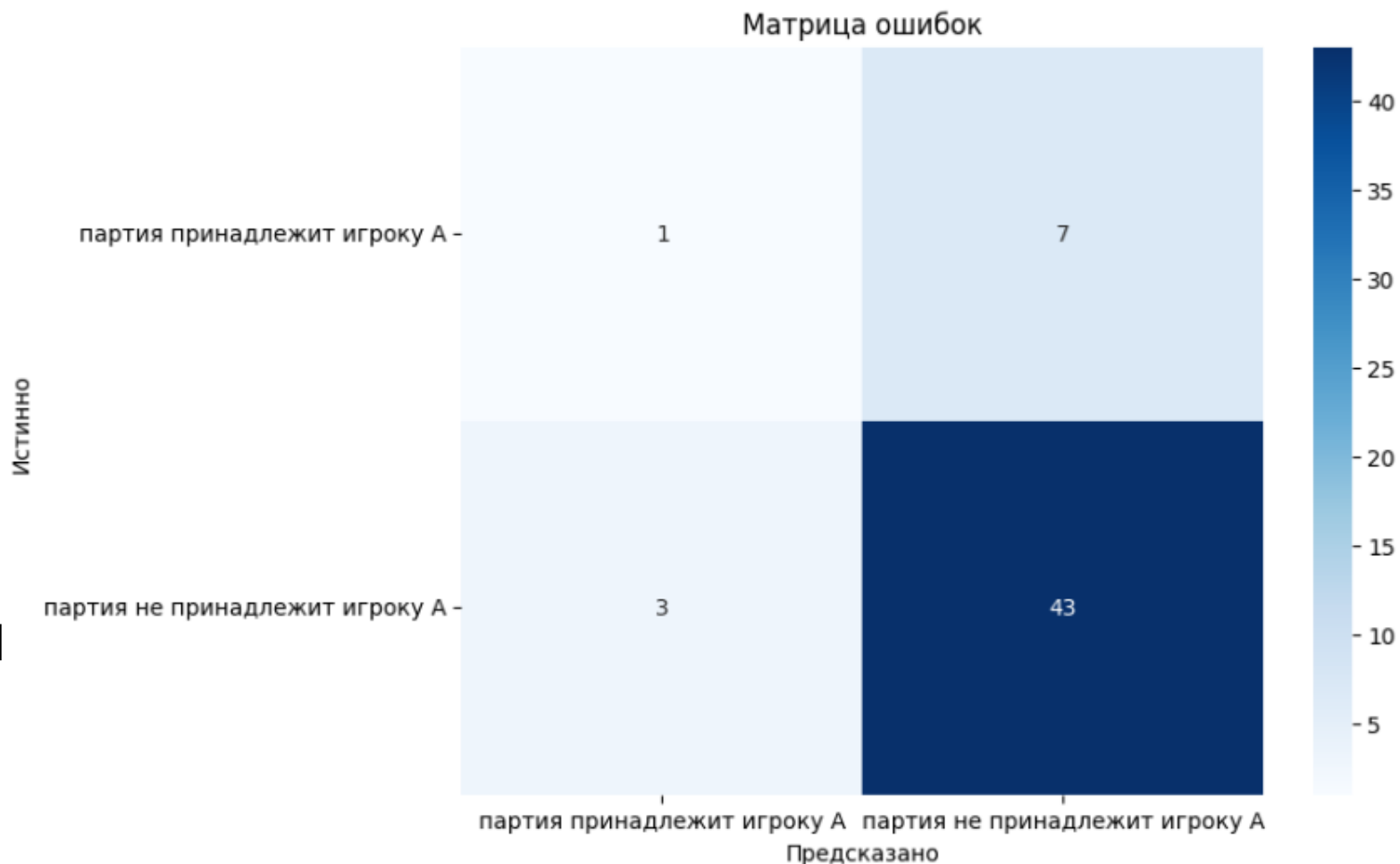
Accuracy: 0.81

F1 Score: 0.79

Precision: 0.77

Recall: 0.81

Наиболее важная
метрика для нас
Precision



Выводы

- Случайный лес сработал очень плохо, потому что он не учитывает временную составляющую. К тому же для случайного леса здесь слишком мало объектов.
- А временные ряды работали слишком долго чтобы можно было успеть настроить оптимальные веса и посчитать метрики.