Prediction of the outcomes of Chess games with the method of local fields

Vladimir Nikulin*
Department of Mathematics, University of Queensland
Brisbane, Australia

1 Method of local fields

Definition 1 *Let us denote by $\mathcal{N}_\theta$ the list of opponents of the player θ during the current M and a few past months, where results are not necessary.*

In order to simplify notations, we shall consider one particular game, because any other game may be considered similarly.

In this case we are dealing with two fields $\mathcal{N}_w$ and $\mathcal{N}_b$ corresponding to the white and black colors, $n_w = \#\mathcal{N}_w \geq 1, n_b = \#\mathcal{N}_b \geq 1$ - the numbers of the players in the corresponding fields.

Let us denote by r_w and r_b *global* ratings of the players (the term "global" means traditional ratings, which are used in most of the Chess ranking systems including ELO or Chessmetrics).

As a next step, we shall introduce the most important and influential *local* ratings for both "white" and "black" players.

Definition 2 *We shall call q_w and q_b*

$$q_w = \frac{1}{n_b} \sum_{\theta \in \mathcal{N}_b} I\{r(\theta) < r_w\}; \;\; q_b = \frac{1}{n_w} \sum_{\theta \in \mathcal{N}_w} I\{r(\theta) < r_b\}, \tag{1}$$

as local ratings for the white and black players, where I is an indicator function.

Remark 3 *Note that the value of $r(\theta) = r(\theta, M)$ in (1) depends on the month M of the game.*

Remark 4 *Local ratings may be regarded as a secondary ratings after the global ratings and the scheduling of the games during the current a a few past months.*

*Email: vnikulin.uq@gmail.com

In addition, we shall compute less influential secondary statistics (or features), where we shall write $r \in \mathcal{N}$ instead of $\theta \in \mathcal{N}$ to simplify notations

$$q_w^{(1)} = \exp\,(\min_{r \in \mathcal{N}_b} r - r_w), \tag{2a}$$

$$q_w^{(2)} = \exp\,(\max_{r \in \mathcal{N}_b} r - r_w), \tag{2b}$$

$$q_w^{(3)} = \exp\,(\frac{1}{n_b} \sum_{r \in \mathcal{N}_b} r - r_w), \tag{2c}$$

$$q_w^{(4)} = \sqrt{\frac{1}{n_b} \sum_{r \in \mathcal{N}_b} (r - r_w)^2}, \tag{2d}$$

$$q_w^{(5)} = \sqrt{\frac{1}{n_b} \sum_{r \in \mathcal{N}_b} r^2 - \left(\frac{1}{n_b} \sum_{r \in \mathcal{N}_b} r\right)^2}, \tag{2e}$$

where the last one (2e) feature corresponds to the standard deviation. Note, also, that the features corresponding to the black color (with index "b" may be defined similarly).

The final solution was produced by the *gbm* (gradient boosting machine) function in R [1]. The data-matrix as an input of *gbm* includes two parts 1) local fields, which may be fully defined in the terms of the statistics (1) and (2a - 2e), see Section 4; and 2) traditional, which we shall describe below.

2 Some secondary (standard) features, which were used in the main model

The traditional database includes exactly 22 components:

1) Compute the number of games for the white player with the common opponents (common means that the player had at least one game with the black player);

2) compute weighted score for the white player with the common opponents,

where weighting was done according to the current month M:

$$weight = \left(\frac{M}{132}\right)^{\alpha}, \alpha = 3.4, M = 1, \ldots, 132. \tag{3}$$

3-4) The same as (1-2) for the black player/color;

5-7) numbers of losses, draws and wins in the past for white color;

8-10) numbers of losses, draws and wins in the past for black color;

11) r_w;

12) $r_w(M) - r_w(M-1)$, where $r_w(M) = r_w$ and $r_w(M-1)$ is the rating of the white player which was computer using past $M-2$ months;

13) $(exp(r_w(M)) + \Delta) \,/\, (\exp(r_w(M-1)) + \Delta)$, where $\Delta = 0.01$;

14) number of game during current month M;

15-18) The same as (11-14) for black color;

19) $H(M) = r_w(M) - r_b(M)$ - the approximation of the first derivative;

20) $2H(M) - H(M-1)$ - the approximation of the second derivative;

21) $3H(M) - 3H(M-1) + H(M-2)$ - the approximation of the third derivative;

22) $(exp(r_w(M) - r_w(M-1)) + \Delta) \,/\, (\exp(r_b(M) - r_b(M-1)) + \Delta)$.

Remark 5 *With the standard database alone it was possible to achieve the public score of 0.2547 in the terms of the binomial deviance.*

3 Computation of two matrices of ratings

The first matrix A_1 (for training) with estimated ratings for the months from 120 to 131, and the second matrix A_2 (for testing) with estimated ratings for the months from 121 to 132. Note that any information from the 132th month was fully excluded from the computation of the matrix A_1.

In order to estimate the required ratings we used the Outis-like system [2] with 21 global iterations, where initial ratings were computed by the Chessmetrics with only one global iteration.

Remark 6 *We used a few modifications of the Outis sytem, which are not very significant. For example, we used the binomial deviance as a target function instead of the squared error. Also, we used different weight structure for the global iteration, and additional monthly weighting (3) with $\alpha = 0.7626$. In total, we applied 7 parameters for the Outis model and 9 parameters for the Chessmetrics. All the parameters were re-optimised using specially designed software written in C.*

4 Dramatic improvement with 3 sets of local fields features

The first question we have to address here is how many past months to select, see Definition 1, in order to define the local field. Based on our CV experiments and on some public scores on the Leaderboard, we decided to use three local fields with 1) M -current month; 2) M and $M-1$; and 3) M, $M-1$ and $M-2$.

Any particular local field may be considered similarly and includes exactly 18 features as described below:

1) n_w;

2) q_w;

3-6) $q_w^{(i)}, i = 3, \ldots, 6$;

7-12) the same as (1-6) for black color;

13) $n_w/(n_b + n_w)$;

14) $q_w/(q_b + q_w)$;

15-18) $(q_w^{(i)} + \Delta)/(q_b^{(i)} + q_w^{(i)} + 2\Delta), i = 1, \ldots, 4$.

In total, we have got a database with 76 features, where 22 are standard and 54 $(3 \cdot 18)$ corresponding to the local fields. With this database we were able produce the score below 0.249 (in public).

5 Learning from the future scheduling

We can exploit the following hypothesis: "most likely, a player had been successful during the current month if he/she will have more intense schedule, or more stronger opposition in the following month compared to the current month." We shall describe here the most basic strategy, which we did not explore deeper. Note that in order to implement the main hypothesis it is not necessary to have the outcomes of the games, and we can apply 1) the monthly sums of the games $x(M)$ and 2) the monthly averages of opponents ratings $y(M)$.

Further, we can create the following features:

1) $\exp(\beta(x_w(M+1) - x_w(M)))$, where $\beta = 0.1$;

2) $\exp(\beta(x_b(M+1) - x_b(M)))$;

3) $\exp(\beta(x_w(M+1) - x_w(M) - x_b(M+1) + x_b(M)))$;

4) $\exp(y_w(M+1) - y_w(M))$;

5) $\exp(y_b(M+1) - y_b(M))$;

6) $\exp(y_w(M+1) - y_w(M) - y_b(M+1) + y_b(M))$.

Remark 7 *We used above 6 features in our final submission and achieved improvement of about 0.0004 in private.*

Remark 8 *Finally, it is interesting to note that a huge amount of the spurious games may affect severely not only the models which are learning from the future, but, also, those models, which are learning from the past. We had noticed this fact immediately during the final experiments against the follow-up data. The data for the follow-up experiment included considerably larger amount of the spurious games in order to prevent learning from the future. However, our model, which uses the information from the current and a few past months was affected severely as well.*

References

[1] J. H. Friedman. Greedy Function Approximation: A Gradient Boosting Machine, The Annals of Statistics, **29**(5), pp. 1189-1232

[2] Y. Sismanis. How I won the "Chess ratings - Elo vs the Rest of the World" Competition, *arXiv:1012.4571v1*, 2010.