

Systemy ratingowe w sporcie

Jan LASEK*

* Absolwent Wydziału Matematyki
i Informatyki UW

Trwa rywalizacja drużyn w Mistrzostwach Świata w piłce nożnej. Z tej okazji chcielibyśmy przybliżyć Czytelnikom tematykę budowy tzw. systemów ratingowych w sporcie. Są to mechanizmy przyporządkowywania drużynom liczb (*ratingów*), które mają odzwierciedlać ich potencjał. Pełnią one ważne funkcje. Przede wszystkim dostarczają porównawczej miary siły drużyn, co w sytuacjach gdy mamy dużą liczbę uczestników oraz stosunkowo małą liczbę meczów umożliwia ich uporządkowanie. W ten sposób m.in. ratingi stosowane są do parowania rywali na podobnym poziomie, rozstawiania drużyn w turniejach i ich rundach kwalifikacyjnych, a nawet do przyznawania zawodnikom pozwolenia na grę w danej lidze. Systemy ratingowe znajdują również zastosowanie poza sportem. Często stosowane są przy ocenie złożonych algorytmów rywalizujących w rozwiązywaniu danego zagadnienia, porównywanych parami przez pewną wyrocznię, na przykład grono ekspertów (zob. *Arena Leaderboard* dla modeli językowych).

Model Elo

Oryginalnie system Elo został stworzony dla szachistów. Tutaj przyjmujemy jednak narrację piłkarską, pisząc o drużynach.

Jednym z najbardziej znanych systemów ratingowych jest model zaproponowany przez amerykańskiego profesora fizyki węgierskiego pochodzenia Arpada Elo (1903–1992), który był również utytułowanym szachistą. Przyjmijmy, że mamy zadanie polegające na wyznaczeniu ratingów $r_1, r_2, \dots, r_d$ dla d drużyn na podstawie wyników n meczów między nimi. Niech $r_i^{(k)}$ będzie ratingiem i -tej drużyny po k -tym meczu ($k \leq n$). Zaczynamy, przyjmując pewne wartości startowe, np. $r_i^{(0)} = 0$ dla wszystkich i . Powiedzmy, że w k -tym meczu grały drużyny i oraz j ($i < j$), i zakończył się on wynikiem $o_{ij}^{(k)} \in \{0, 0.5, 1\}$, gdzie wartość 1 odpowiada wygranej drużyny i , wartość 0.5 – remisowi, a wartość 0 – wygranej drużyny j . Nie wykluczamy przy tym wielokrotnego rozgrywania meczów między dwiema tymi samymi drużynami. W najczęściej przytaczanym wariancie modelu Elo ratingi drużyn i oraz j aktualizowane są w następujący sposób:

$$\begin{aligned} r_i^{(k)} &= r_i^{(k-1)} + \kappa \cdot (o_{ij}^{(k)} - p_{ij}^{(k)}), \\ r_j^{(k)} &= r_j^{(k-1)} - \kappa \cdot (o_{ij}^{(k)} - p_{ij}^{(k)}), \end{aligned}$$

przy czym $p_{ij}^{(k)}$ odpowiada prawdopodobieństwu wygranej pierwszej drużyny (pod warunkiem braku remisu) i określone jest za pomocą funkcji

$$p_{ij}^{(k)} = \frac{1}{1 + \exp(r_j^{(k-1)} - r_i^{(k-1)})} = \sigma(r_i^{(k-1)} - r_j^{(k-1)}),$$

gdzie $\sigma(t) = \frac{1}{1 + \exp(-t)}$ to tzw. *funkcja logistyczna*. Parametr $\kappa > 0$ określa rozmiar kroku adaptacyjnego. Ratingi drużyn niebiorących udziału w k -tym meczu pozostają oczywiście bez zmian.

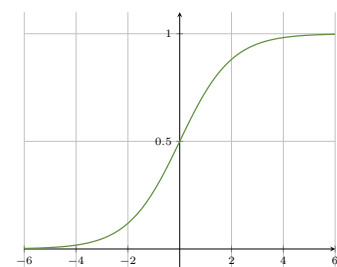
Zwróćmy uwagę na czytelną interpretację wzoru aktualizującego r_i wyżej:

jeśli z perspektywy drużyny i faktyczny wynik $o_{ij}^{(k)}$ jest powyżej oczekiwań $p_{ij}^{(k)}$, jej rating zostaje podwyższony (i odwrotnie w przypadku porażki).

System więc w przejrzysty sposób dokonuje autokorekty dotychczasowych estymat na podstawie wyników nowych gier, proporcjonalnie do różnicy między rzeczywistym wynikiem a przewidywaniami.

Związek z algorytmem najszybszego spadku

Ciekawą obserwacją jest fakt, że model Elo może być wyprowadzony jako pewien wariant algorytmu najszybszego spadku (ang. *steepest descent* lub *gradient descent*). W największym skrócie przypomnijmy, że algorytm ten iteracyjnie dostosowuje wartości danego parametru x zgodnie z regułą $x := x - \gamma \cdot \frac{\partial L}{\partial x}$ dla pewnej „funkcji straty” L podlegającej minimalizacji, a $\gamma > 0$ jest tzw. parametrem uczenia (*learning rate*). Jaką funkcję wybrać w przypadku problemu konstrukcji ratingów?



Wykres funkcji logistycznej $\sigma(t) = \frac{1}{1 + e^{-t}}$.
Odnotujmy symetrię tego wykresu:
spełniona jest tożsamość $\sigma(t) + \sigma(-t) = 1$.

Obrazowym przedstawieniem algorytmu *gradient descent* jest próba wejścia na szczyt po omacku poprzez wykonywanie po jednym kroku w najbardziej stromym kierunku.

Tak określona strata dla przewidywania zmiennej przyjmującej skończenie wiele wartości (np. wyniku meczu) jest powszechna w uczeniu maszynowym i występuje tam pod nazwą *straty logarytmicznej*. Ma ona też klarowną interpretację probabilistyczną jako logarytm tzw. *funkcji wiarygodności*, czyli prawdopodobieństwa uzyskania zaobserwowanego ciągu wyników w zależności od nieznanymi parametrów.



To popularny i stosowany od dawna model, zob. np. w: Maher M.J., „Modelling association football scores” (1982).

W kącie (Ω_{06}^1) można więcej przeczytać o rozkładzie Poissona i w szczególności dlaczego możemy się spodziewać, że liczba strzelonych bramek jest opisana tym rozkładem.

Założmy, że ratingi $r_1, \dots, r_d$ mają tę własność, że kiedy spotykają się drużyny i oraz j ($i < j$), to szansa na wygranie pierwszej z nich to $p_{ij} = \sigma(r_i - r_j)$. Zauważmy przy tym od razu, że wówczas szansa na wygranie drugiej drużyny to $1 - p_{ij} = \sigma(r_j - r_i)$. Niech $\mathcal{M}$ będzie zbiorem trójek (i, j, k) , gdzie $k = 1, 2, \dots, n$, zaś $i < j$ są numerami drużyn grających w k -tym meczu. Rozważmy następującą funkcję wektora ratingów $\mathbf{r} = (r_1, r_2, \dots, r_d)$:

$$L(\mathbf{r}) = - \sum_{(i,j,k) \in \mathcal{M}} (o_{ij}^{(k)} \log p_{ij} + (1 - o_{ij}^{(k)}) \log(1 - p_{ij})).$$

Spełnia ona intuicyjne wymagania stawiane wobec funkcji straty dla tego problemu – jest tym większa, im bliżej wartości prawdopodobieństwa p_{ij} są częstotliwości wygranych drużyny i nad drużyną j . Oznaczmy pojedynczy składnik powyższej funkcji kosztu przez $l_{ij}^{(k)}$. Można nietrudno sprawdzić, że funkcja logistyczna σ spełnia $\sigma'(t) = \sigma(t)(1 - \sigma(t))$, a zatem $(\log \sigma(t))' = 1 - \sigma(t)$. Wynika stąd, że

$$\begin{aligned} \frac{\partial l_{ij}^{(k)}}{\partial r_i} &= - (o_{ij}^{(k)} \frac{\partial \log \sigma(r_i - r_j)}{\partial r_i} + (1 - o_{ij}^{(k)}) \frac{\partial \log \sigma(r_j - r_i)}{\partial r_i}) \\ &= - o_{ij}^{(k)} (1 - p_{ij}) + (1 - o_{ij}^{(k)}) p_{ij} \\ &= p_{ij} - o_{ij}^{(k)}. \end{aligned}$$

Okazuje się zatem, że stosując kolejne iteracje algorytmu najszybszego spadku dla pojedynczych meczów zgodnie z kolejnością ich występowania, otrzymujemy regułę równoważną „dynamicznej” aktualizacji ratingów w systemie Elo.

Z uwagi na swoją elegancję oraz skuteczność w estymacji model Elo został z powodzeniem zaaplikowany w wielu sportach. W piłce nożnej na poziomie rozgrywek międzynarodowych jest on podstawą oficjalnego rankingu FIFA (zarówno dla reprezentacji męskich, jak i kobiecych) oraz w alternatywnej implementacji dostępnej na EloRatings.net. Oczywiście w zastosowaniach systemy te różnią się pewnymi szczegółami od bazowego wariantu opisanego wyżej, m.in. ważeniem meczów rangą rozgrywek (np. przez rozróżnienie meczów na towarzyskie i pucharowe) lub modyfikacjami wartości wyniku $o_{ij}^{(k)}$ na podstawie strzelonych bramek (czyli np. wynik 4 : 0 jest opisywany pewną wartością bliżej 1 niż wynik 2 : 1 o mniejszym marginesie zwycięstwa) czy też modyfikacją współczynnika κ jako rosnącej funkcji bezwzględnej różnicy bramek.

Idąc tropem metody najszybszego spadku, można otrzymać całą gamę alternatywnych systemów ratingowych przy zastosowaniu innych funkcji modelujących wyniki spotkań. Poniżej prezentujemy jeden taki system na bazie modelu liczby goli szczególnie popularnego w piłce nożnej.

Model ratingowy marginesu zwycięstwa

Z wyżej przedstawionego wyprowadzenia wzoru na ratingi Elo wynika, że nie są one dostosowane do przewidywania liczby goli strzelonych przez poszczególne drużyny w danym meczu. Przedstawimy teraz wyprowadzony w podobnym duchu alternatywny sposób oceny drużyn, lepiej skrojony pod takie zastosowanie. Jest on oparty na tzw. *rozkładzie Poissona*, tzn. rozkładzie prawdopodobieństwa na zbiorze liczb całkowitych nieujemnych, który liczbie $k \in \mathbb{Z}_0^+$ przypisuje prawdopodobieństwo $\frac{\mu^k}{k!} e^{-\mu}$, gdzie $\mu > 0$ jest ustalonym parametrem (będącym jednocześnie wartością oczekiwaną zmiennej wylosowanej z tego rozkładu). Zakładamy, że zmienne $G_i^{(k)}$ oraz $G_j^{(k)}$ opisujące liczbę goli drużyn w meczu k mają rozkład Poissona oraz że są niezależne. Parametry tych rozkładów to μ_{ij} oraz μ_{ji} , odpowiednio, gdzie

$$\mu_{ij} = \exp(c + r_i - r_j), \quad \mu_{ji} = \exp(c + r_j - r_i),$$

przy czym r_i, r_j są ratingami drużyn i, j , odpowiednio, zaś c jest pewną stałą wspólną dla wszystkich meczów. Prawdopodobieństwo wyniku $x : y$ wynosi wtedy

$$\mathbb{P}(G_i^{(k)} = x, G_j^{(k)} = y) = \frac{\mu_{ij}^x}{x!} \exp(-\mu_{ij}) \cdot \frac{\mu_{ji}^y}{y!} \exp(-\mu_{ji}).$$

Następnie przyjmijmy następującą funkcję kosztu:

$$L(\mathbf{r}) = - \sum_{(i,j,k) \in \mathcal{M}} (\log \mathbb{P}(G_i^{(k)} = g_i^{(k)}) + \log \mathbb{P}(G_j^{(k)} = g_j^{(k)})),$$

gdzie $g_i^{(k)}$ oznacza liczbę bramek drużyny i w danym meczu k przeciwko drużynie j . Ponownie, stosując metodę najszybszego spadku dla pojedynczych meczów w kolejności ich występowania w czasie, otrzymujemy równania aktualizacji dla parametrów

$$\begin{aligned} r_i^{(k)} &= r_i^{(k-1)} + \gamma \cdot [(g_i^{(k)} - g_j^{(k)}) - (\mu_{ij}^{(k)} - \mu_{ji}^{(k)})], \\ r_j^{(k)} &= r_j^{(k-1)} - \gamma \cdot [(g_i^{(k)} - g_j^{(k)}) - (\mu_{ij}^{(k)} - \mu_{ji}^{(k)})], \end{aligned}$$

przyjmując

$$\begin{aligned} \mu_{ij}^{(k)} &= \exp(c + r_i^{(k-1)} - r_j^{(k-1)}), \\ \mu_{ji}^{(k)} &= \exp(c + r_i^{(k-1)} - r_j^{(k-1)}). \end{aligned}$$

Podobnie jak w modelu Elo, otrzymujemy bardzo intuicyjną interpretację równania dla aktualizacji parametrów: jeśli obserwowana różnica bramek między drużynami $g_i^{(k)} - g_j^{(k)}$ różni się od oczekiwanej $\mu_{ij}^{(k)} - \mu_{ji}^{(k)}$, rating danej drużyny ulega korekcie w odpowiednim kierunku. Im większa różnica między przewidywaniami a faktycznym wynikiem, tym większa korekta ratingów rywalizujących drużyn.

W ten sposób dochodzimy do nowego systemu ratingowego opartego na różnicy goli. Wydaje się to bardziej naturalnym sposobem opisu danych piłkarskich (lub ogólniej w sportach, w których liczymy punkty) niż funkcja logistyczna stosowana do danych binarnych. Dzięki temu, że model ten bezpośrednio określa prawdopodobieństwo dowolnego wyniku, można go zastosować do symulacji przebiegu rozgrywek. Takie narzędzia są często używane m.in. w badaniach operacyjnych w sporcie, przy projektowaniu systemów rozgrywek (*tournament design*).

Możliwe rozszerzenia

Przytoczone modele można rozszerzać o szczegóły takie jak już wspomniane współczynniki istotności meczów. W praktyce szczególnie ważnymi aspektami modelowania są uwzględnienie przewagi gospodarza oraz regularyzacja funkcji kosztu. Krótko omawiamy te dwa zagadnienia niżej.

Modelowanie przewagi gospodarza w najprostszym ujęciu polega na zwiększeniu o pewną wartość ratingu drużyny grającej „u siebie”. Jest to dość ciekawe zjawisko, na które składa się zespół czynników (głównie psychologicznych) skutkujących tym, że drużyny grające na swoim stadionie statystycznie wygrywają częściej niż goście. Zjawisko to ma pewnego rodzaju odpowiednik w szachach jako przewaga pierwszego ruchu białych.

Warto również rozszerzyć funkcję straty o składnik tzw. regularyzacji parametrów postaci $\lambda \cdot \sum_i |r_i|^p$ (zwykle $p = 1$ lub $p = 2$, $\lambda > 0$). Umożliwia to uściślenie modelu przez tzw. identyfikację parametrów – w przeciwnym razie funkcja kosztu oparta jedynie na różnicy ratingów nie zmienia wartości po przesunięciu wektora ratingów o dowolną stałą wielkość. Co istotniejsze, regularyzacja umożliwia często uzyskanie dokładniejszych oszacowań prawdopodobieństw wyników meczów, kosztem nieznacznej komplikacji równań aktualizujących ratingi.

Elegancja modelu Elo oraz jakość generowanych przez niego ratingów przyczyniły się do popularności oraz licznych adaptacji tego systemu. Przytoczone tu modele nie uwzględniają szerokiej listy czynników, jak składy drużyn i absencje kluczowych graczy. Tym niemniej interpretowalność i przejrzystość przedstawionych metod jest cenna z punktu widzenia zastosowań. Warto również zaznaczyć, że te proste systemy stosunkowo dobrze wypadają w porównaniu z bardziej złożonymi modelami w przewidywaniu wyników meczów w piłce nożnej – sporcie o znaczącym wpływie czynnika losowego, często decydującym o wyniku rywalizacji. Komu będzie sprzyjać to szczęście w trwających finałach Mistrzostw Świata? Przekonamy się niebawem 🍀

Podobny system został opisany w pracy: Kovalchik S., „Extension of the Elo rating system to margin of victory” (2020), jako heurystyka bez głębszych podstaw teoretycznych. Ten i inne modele szczegółowo omawiamy w pracy: Lasek J. i Gągolewski M., „Interpretable sports team rating models based on the gradient descent algorithm” (2021).

