

پاسخ به سوال نقل و انتقالات

مجید فرزانه

هدف اصلی یافتن ۵ شوت زن برتر از بین تمام بازیکنان موجود در مجموعه داده است. مدلی داریم که ۸۳۸ امتیاز از ۱۰۰۰ امتیاز را تاکنون کسب کرده است. از این رو قادر است با دقت نسبتاً مناسبی احتمال گل شدن هر موقعیت را حساب نماید. در صورتی که این مدل را معیار مورداعتمادی در نظر بگیریم، می‌توانیم بررسی کنیم که هر بازیکن در هر شوتش چقدر مناسب عمل کرده است. به عنوان مثال اگر احتمال گل شدن یک شوت پایین باشد، ولی شوت گل شده باشد، می‌توان گفت بازیکن مهارت بالایی داشته است. بالعکس اگر احتمال گل شدن یک موقعیت بالا باشد، ولی بازیکن، توپ را گل نکرده باشد، می‌توان گفت که مهارت شوت‌زنی بازیکن پایین است.

برای اطمینان بالاتر از روش بالا، تمام شوت‌هایی را که هر بازیکن در مجموعه داده‌ی train داشته بررسی می‌کند. بر اساس رابطه‌ی زیر یک امتیاز به هر بازیکن اختصاص می‌دهیم:

$$Score = \frac{1}{N_{shoots}} \left(100 \times \frac{N_G}{\sum_{i \in G} pr_i} - N_M \sum_{j \in M} pr_j \right) \quad (1)$$

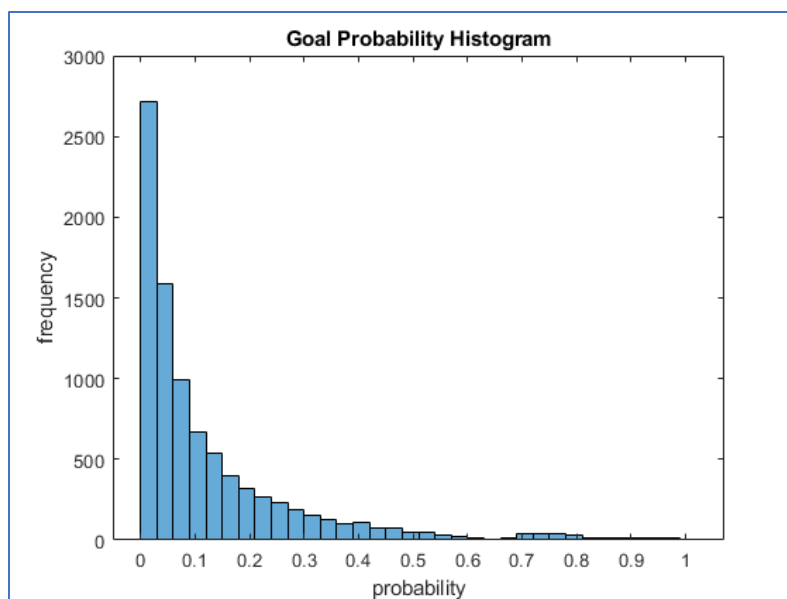
که pr_i احتمال گل شدن شوت i -ام از مجموعه شوت‌های گل شده‌ی (G) بازیکن است. همچنین pr_j احتمال گل شدن شوت j -ام از مجموعه شوت‌هایی است که گل نشده‌اند (M). مقادیر N_G و N_M به ترتیب تعداد شوت‌های گل شده و گل نشده از بازیکن هستند. بدیهی است که:

$$N_{shoots} = N_G + N_M$$

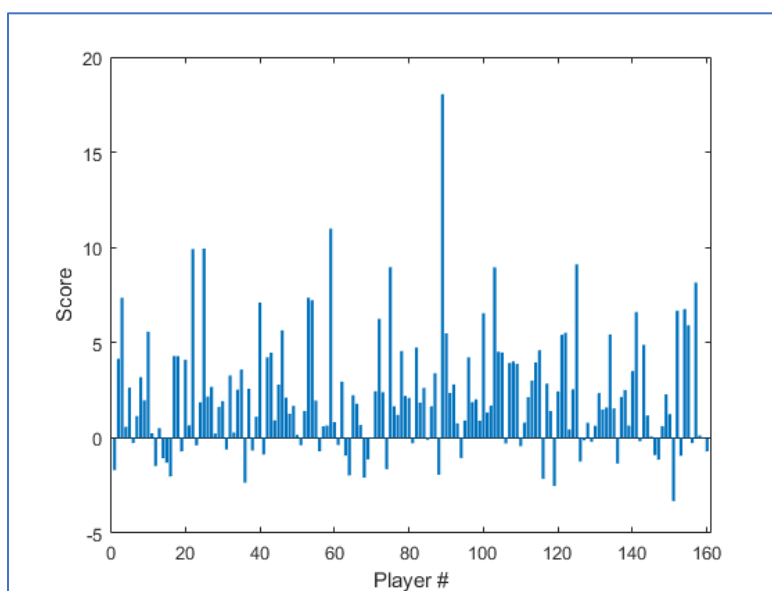
باید به تعداد شوت‌های گل شده به بازیکن امتیاز بدهیم. اما هرچه احتمال گل شدن شوت‌ها کمتر باشد باید امتیاز بزرگتری به بازیکن بدهیم. از این رو مجموع احتمالات تمام شوت‌های گل شده را در مخرج قرار داده‌ایم. همچنین باید به ازای تمام شوت‌های گل نشده از بازیکن امتیاز کسر شود ولی هرچه احتمال گل شدن شوت بزرگتر باشد باید امتیاز بیشتری از بازیکن کسر کنیم. از این رو مجموع احتمالات شوت‌های گل نشده در تعداد شوت‌های گل نشده ضرب شده است. در نهایت مقدار به دست آمده بر تعداد کل شوت‌های بازیکن تقسیم می‌شود. زیرا در مجموعه داده تعداد شوت‌های بازیکن‌های مختلف متفاوت است. برای جلوگیری از منفی شدن امتیازات یک عدد ۱۰۰ نیز در قسمت اول رابطه ضرب شده است.

با استفاده از کد سوال اول مقادیر N_G ، N_M و N_{shoots} برای هر یک از ۱۶۰ بازیکن قابل محاسبه است. حال کافی است مجموعه‌ی G و M به ازای هر بازیکن به دست آید. این مجموعه‌ها به کمک مدل یادگیری ماشین استفاده شده در سوال دوم محاسبه می‌شوند.

شکل ۱ نمودار فراوانی احتمال گل شدن شوت‌ها را بر اساس مدل آموزش داده شده نمایش می‌دهد. شکل ۲ مقدار امتیاز به دست آمده برای هر بازیکن را بر اساس رابطه‌ی (۱) نمایش می‌دهد. حال می‌توان بهترین شوت‌زن‌ها را بر اساس بیشترین امتیازهای به دست آمده مطابق با جدول ۱ نمایش داد.



شکل ۱: هیستوگرام احتمال گل شدن شوت‌ها بر اساس مدل یادگیری ماشین



شکل ۲- امتیاز بازیکن‌ها بر اساس رابطه‌ی ۱

اولویت	نام بازیکن	نرخ گل به شوت	امتیاز
۵	p_67	0.0417	9.1146
۴	p_118	0.0377	9.9124
۳	p_120	0.1522	9.9370
۲	p_151	0.0370	10.9902
۱	p_34	0.0789	18.0511

جدول ۱: انتخاب‌های نهایی