

Review der VAEP Werte

July 12, 2021

Ich habe die Werte für das Beispiel aus dem public notebook auf github noch mal genauso nachgerechnet und mit den Werten von uns verglichen. Daraus habe ich folgende Erkenntnisse gezogen:

- selbst mit der selben Berechnungsgrundlage weichen die Werte voneinander ab, jedoch teilweise nicht mehr so stark
- zur Berechnung des VAEPs werden nur die Spiele eines Wettbewerbs verwendet, nicht wettbewerbsübergreifend. Es werden also nicht zum trainieren des Modells die Spiele aus allen Wettbewerben genommen sondern jeder Wettbewerb bekommt sein eigenes Modell
- es gibt zum Teil unterschiede wie events in actions umgewandelt werden (siehe Beispiel Bilder)

Es ist also nicht so, dass unser Modell falsch ist, da die socceraction Leute es genauso wie wir gemacht haben. Meiner Meinung nach könnte eine mögliche Ursachen für die Diskrepanz der Werte sein:

- Datengrundlage des Modells: Da wir das Problem hatten, dass sich nicht alle Methoden aus python nach R übersetzen ließen, musste ich quasi weniger selektieren. Dadurch hatten wir in der Regel mehr Actions, als wenn wir das in Python gemacht hätten. Damit man unsere action aber mit den VAEP Werten zusammengeführt werden können, musste ich einige Selektionsfaktoren aus dem Python Code löschen. Wir hatten also immer mehr Actions im Python Skript als im R Skript. Ein Beispiel war zum Beispiel, wie Dribblings ausgewählt wurden: Für dribblings wurde in der socceraction Implementierung eingeführt, dass diese nur eine bestimmte Länge und nur eine Dauer von 10s berücksichtigen. Das hatte ich soweit ich mich erinnern kann entfernt. Grundlegend hatte ich aber nichts an der Logik, wie events in actions umgewandelt werden verändert. Meines Erachtens könnte also die größere Zahl der Actions in unserem Modell dazu führen, dass wir andere VAEP Werte bekommen, da bin ich mir bei der Begründung jedoch unsicher ob das so richtig ist: Ich gehe davon aus, dass durch die höhere Zahl an Actions eventuell "unnötige" Actions in das Modell einfließen und so ein Bias entsteht, jedoch finde ich, dass die Selektion z.B. bei den Dribblings, wie ich beschrieben habe, unbegründet

und sehr subjektiv ist, da ich der meinung bin, dass die Restriktion aus unter 10s nicht unbedingt sinnvoll ist.

Marouane Fellaini-Bakkioui	100.5882353	29.2658228	101.470588	35.294118	pass	success	head
Michy Batshuayi Tunga	101.4705882	34.4303797	100.588235	33.569620	dribble	success	foot
Michy Batshuayi Tunga	100.5882353	33.5696203	1.000000	1.000000	bad_touch	fail	foot
Michy Batshuayi Tunga	104.1176471	33.5696203	102.352941	33.569620	dribble	success	foot
Michy Batshuayi Tunga	102.3529412	33.5696203	1.000000	1.000000	shot	fail	foot

Figure 1: Action Sequenz der R Implementierung aus dem Beispiel des public notebooks aus github

type_name	bodypart_name	player_name	nickname
pass	head	Marouane Fellaini	Marouane Fella...
dribble	foot	Michy Batshuayi	Michy Batshuayi
bad_touch	foot	Michy Batshuayi	Michy Batshuayi
dribble	foot	Michy Batshuayi	Michy Batshuayi
dribble	foot	Michy Batshuayi	Michy Batshuayi

Figure 2: Selbe Sequenz nur aus der Python Implementierung wie vorher erwähnt stimmt der letzte Actiontype nicht überein

Pass	24	792	Belgium	1	Regular Play	792	Belgium	0.68000	0	["pass", ("1002e3d9-e23-4...", 3296.00000	Marouane Fellaini...
Pressure	24	792	Belgium	1	Regular Play	798	England	0.62000	0	["002e3d9-e23-4...", 3244.00000	Joan Stones
Ball Receipt*	24	792	Belgium	1	Regular Play	792	Belgium	nan	0	["1d22a6-828-48...", 3457.00000	Michy Batshuayi T...
Carry	24	792	Belgium	1	Regular Play	792	Belgium	0.38000	0	["carry", ("1002e3d9-e23-4...", 3457.00000	Michy Batshuayi T...
Miscontrol	24	792	Belgium	1	Regular Play	792	Belgium	nan	0	["", 3457.00000	Michy Batshuayi T...
Ball Recovery	24	792	Belgium	1	Regular Play	798	England	nan	0	["ball_rec...", 3468.00000	Jordan Pickford
Ball Recovery	24	792	Belgium	1	Regular Play	792	Belgium	nan	0	["", 3457.00000	Michy Batshuayi T...
Carry	24	792	Belgium	1	Regular Play	792	Belgium	0.62000	0	["carry", ("1d479e86-53ea-4...", 3457.00000	Michy Batshuayi T...
Shot	24	792	Belgium	1	Regular Play	792	Belgium	0.98000	0	["shot", ("1985db09-1357-41...", 3457.00000	Michy Batshuayi T...

Figure 3: Betrachtet man die rohen Eventdaten sieht man, dass das letzte Event auch als Shot gekennzeichnet ist. Bei der Umwandlung in actions aber zum dribble wurde ..