

10/09/2011

Gedoe over ratings

Door: Luc Scheepmaker

N_t

$$R_n := R_1 + \sum_{i=1}^k (W - W_i) \cdot k$$

Met ingang van dit seizoen verandert de manier waarop de interne rating wordt berekend. "Oh, wordt die dan berekend?" zal de reactie van de meeste mensen zijn. Ja, en er zijn mensen die zich daar heel druk om maken....

Achtergrond

Vroeger – toen De Raadsheer nog een grote vereniging was – speelden we intern in groepen. Het hele seizoen speelde je tegen een beperkt aantal tegenstanders van ongeveer gelijke sterkte. Als je naar de club kwam om een partijje te schaken, kreeg je meestal een tegenstander die 'je kon hebben': de tegenstander was niet veel sterker of zwakker dan jij. Maar dat was vroeger...

Sinds een aantal jaren (vele jaren eigenlijk) daalt het ledental en is het spelen in groepen geen optie meer. Zeven jaar geleden zijn we overgestapt op het Keizersysteem: we spelen in één grote groep. Gevolg daarvan is dat je vaker tegen iemand wordt ingedeeld 'van wie je zou moeten winnen' of juist 'tegen wie je geen kans maakt'. Niet ideaal, maar er wordt tenminste nog geschaakt.

Een ander nadeel is dat het berekenen van een interne rating hierdoor een stuk problematischer is geworden. Een ratingsysteem werkt pas goed als er veel wordt gespeeld en vooral veel tegen spelers van vergelijkbare sterkte.

Het oude systeem

Het interne ratingsysteem dat tot nu toe werd gebruikt is een erfenis van onze toenmalig wedstrijdleider, Wim van Vugt, en ook van de toenmalige speelwijze in groepen. Een vrij eenvoudige formule bepaalt de TPR van een speler in de interne competitie:

$$TPR = G - (400 * \text{LOG}(N/S - 1))$$

G = gemiddelde van de interne startrating(s) tegenstander(s)

N = aantal partijen

S = score

(er worden nog wat correcties toegepast als je score 0% of 100% is en als het verschil tussen G en jouw startrating groter is dan 400)

De clubrating wordt vervolgens berekend op de volgende manier:

$$\text{Clubrating} = (\text{Startrating} * 20 + \text{KNSB-rating} * 10 + \text{TPR} * N) / (20 + 10 + N)$$

(de factor 10 valt weg als iemand geen KNSB-rating heeft)

Dit systeem is een goede benadering van de interne rating van een speler, vooral als je deze ratings niet iedere week maar – zeg – drie keer per jaar publiceert.

Het probleem

Bovenstaande formule voor de berekening van de TPR heeft een zwak punt als je de TPR week-op-week berekent: Wanneer je tegen een speler wordt ingedeeld met een veel lagere rating – en wint – dan kan je TPR (en daarmee je interne rating) dalen! Verklaring is dat de invloed op de factor G groter is dan de factor $(400 * \text{LOG}(N/S - 1))$. Het omgekeerde is echter ook mogelijk: dat je rating stijgt als je verliest van een veel sterkere speler!

Hoe moet ik het in Jip-en-Janneke-taal uitleggen? Het is net zoets als een 2% loonsverhoging krijgen terwijl de inflatie 3% is. Uiteindelijk ben je slechter af.

Hoe kan ik het uitleggen? Op woensdagavond word je aan de bar aangestoten door een hevig verontwaardigde Dick Beijne die week in, week uit, zijn interne rating op de website controleert. Hem is onrecht aangedaan! Na een winstpartij is zijn interne rating gedaald. Met een smekende blik in zijn ogen vraagt hij "Kun je me niet een paar puntjes erbij geven?". Ik zie het helemaal voor me: iedere woensdagavond ga ik in een hoekje van de zaal zitten. Achter een klassiek notenhouten bureau houd ik spreekuur, als een Don Corleone, en geef ratingpunten weg aan mensen die met de hoed in de hand om gerechtigheid komen smeken.

Nee, dan maar een ander ratingsysteem.

Het nieuwe systeem

Het nieuwe systeem kijkt per partij naar de verwachte score. Deze wordt bepaald door het ratingverschil tussen twee spelers, een aantal voorbeelden:

Ratingverschil	Verwachte score
-400	0,08
-200	0,24
-100	0,36
-50	0,43
0	0,50
50	0,57
100	0,64
200	0,76
400	0,92

Uitleg: als je tegenstander 200 ratingpunten sterker is dan jij, dan heb je 24% kans om te winnen. Of beter gezegd: als jullie vier partijen tegen elkaar spelen word je geacht ongeveer 1 uit 4 te scoren.

De verwachte score (W_e) wordt afgezet tegen de werkelijke score (W):

$$R_t = (W - W_e) \cdot K$$

De K is de verversingsfactor; naarmate je meer partijen gespeeld hebt, en de interne rating dus betrouwbaarder is, is de K lager. In het nieuwe systeem ligt de K tussen de 25 en de 72. Een voorbeeld:

Rating wit : 1600
Rating zwart : 1700
K-factor wit : 25
K-factor zwart : 35
Uitslag : remise

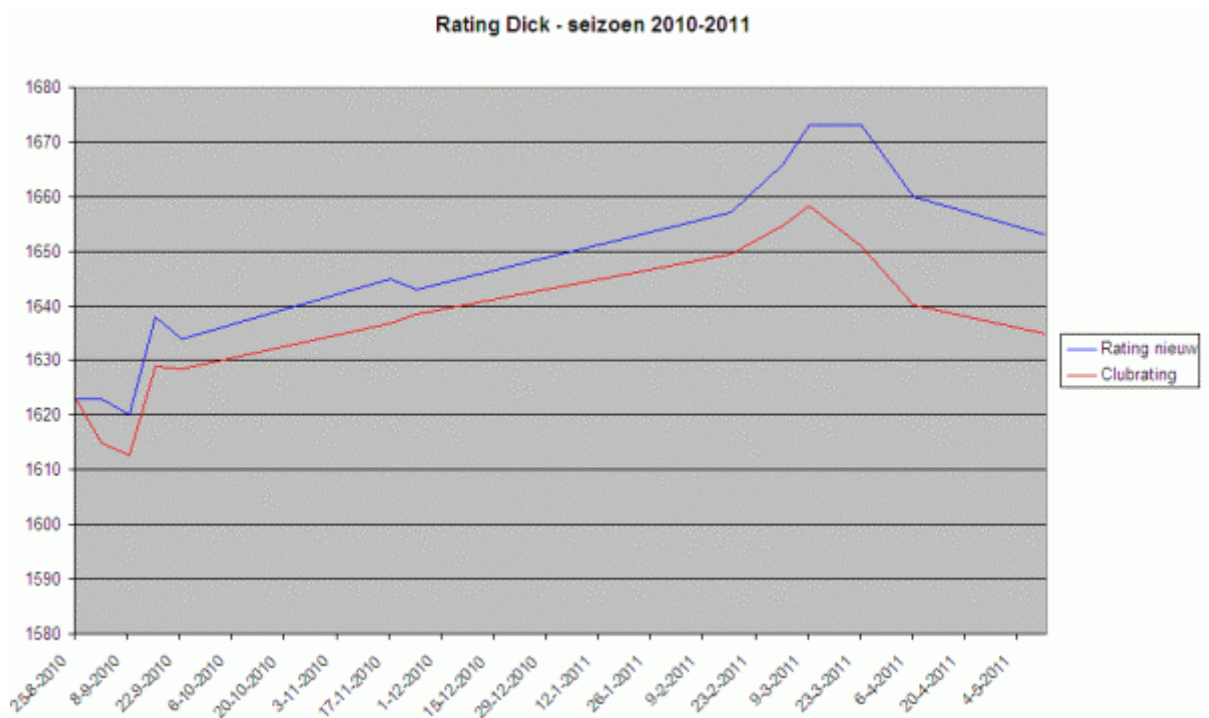
Nieuwe rating wit: $1600 + (0,5 - 0,36) \cdot 25 = 1603,50 \approx 1604$
Nieuwe rating zwart: $1700 + (0,5 - 0,64) \cdot 35 = 1695,10 \approx 1695$

Voordeel van dit nieuwe systeem is dat je rating niet kan dalen als je wint. Als je verwachte score dicht bij de 1 ligt kan het wel zijn dat je rating na een winstpartij gelijk blijft.

Nadeel van het nieuwe systeem is dat het niet op gemiddelden gebaseerd is. Je rating zal week-op-week heftiger omhoog en omlaag gaan. Een simulatie over het vorige seizoen leert echter dat de verschillen met het oude systeem niet groter zijn dan 40 punten in de plus of in de min.

Wat betekent dit nu voor Dick Beijne?

Nou, goed nieuws voor Dick. Hij werd vorig jaar twee keer gestraft voor een winstpartij, één keer was qua indeling wat eerlijker geweest. Hier de vergelijking in grafiekvorm:



De twee horizontale lijnstukken in de blauwe grafieklijn (aan het begin en in maart) laten zien dat in het nieuwe systeem een winstpartij tegen iemand die een veel lagere rating heeft, niet tot een lagere interne rating leidt. Het resultaat is ook weer niet schokkend als je beseft dat een rating er makkelijk 20 á 30 punten naast kan zitten.

In de grafiek kan je ook duidelijk zien dat, zoals eerder opgemerkt, de nieuwe rating wat heftiger beweegt dan de oude.

Wat niet verandert

De ratingberekening in een beperkte groep houdt altijd het risico in van ratinginflatie: het totaal van de ratings stijgt terwijl de spelers als groep niet sterker zijn geworden. Daarnaast is het niet reëel de interne rating van een speler al te zeer te laten afwijken van zijn/haar externe rating. Daarom vinden er jaarlijks twee correcties plaats:

Herijking collectief

Jaarlijks wordt gekeken of het totaal van de interne ratings van leden met een KNSB-rating afwijkt van het totaal van hun KNSB-ratings. Daar komt een correctie uit voort. Dit jaar was deze correctie overigens zeer gering, n.l. -3 punten. Dit betekent dat alle interne ratings met 3 punten zijn verlaagd.

Herijking individueel

Jaarlijks vindt er een correctie plaats waarbij de afwijking tussen de interne en externe rating van een speler wordt verkleind:

$$\text{Rating gecorrigeerd} = (20 \cdot \text{interne rating} + 10 \cdot \text{KNSB-rating}) / 30$$

Voor spelers waarbij de interne rating erg afwijkt van de externe rating zal aan het begin van het seizoen de interne rating veel lager of hoger zijn dan de interne rating aan eind van het voorgaande seizoen.

Tot slot

Geen enkel systeem is ideaal. Ik hoop dat ik met deze aanpassing voldoende tegemoet ben gekomen aan de geuite kritiek.