

## WINNEN MAG GEEN MINPUNTEN OPLEVEREN

*Een nieuw ratingsysteem voor het dubbelspel waarin elke wedstrijd telt, ook als spelers duidelijk in niveau verschillen. Dat is wat Helga Wiersma van de tennisbond KNLTB van de wiskundigen van de Studiegroep Wiskunde met de Industrie vraagt.*

We zien soms gekke dingen bij de rating van spelers die dubbelen, vertelt Helga Wiersma van de tennisbond KNLTB de wiskundigen bij de start van de studiegroep. “Zoals minpunten bij een gewonnen wedstrijd.” Dan wint een duo een tenniswedstrijd, en gaat de rating van één van de spelers toch achteruit. Het is een onbedoeld bijeffect van het ratingsysteem voor dubbels van de KNLTB, waarbij het resultaat van de wedstrijd afhangt van de speelsterkte van de tegenstander. “We lossen dat nu op door de wedstrijd niet mee te laten tellen, maar dat is voor spelers natuurlijk ook niet leuk. Dan levert winst niets op voor je rating”, vertelt Wiersma, Manager Wedstrijdtennis bij de KNLTB.

Voor tennissers op niveau gaat tennissen niet alleen om een lekker potje ballen, maar ook om het verbeteren van hun speelsterkte - die weer bepaalt welke tegenstanders ze krijgen. Spelers hebben een individuele rating voor zowel het enkelspel als het dubbelspel. In het huidige ratingsysteem kent de KNLTB bij dubbelwedstrijden geen score toe wanneer het niveauverschil tussen de spelers groot is, wat geregeld zo is. “We willen graag toe naar een ratingsysteem waarbij elke gewonnen wedstrijd loont”, benadrukt Wiersma.

Het huidige ratingsysteem DDS (dynamisch speelsterkte systeem) is ontworpen voor het enkelspel, en daarna uitgebreid naar de dubbels. Het ratingsysteem is een stroomschema, waar twee uitkomsten mogelijk zijn: de wedstrijd telt mee, of geeft geen resultaat. De rating is het gemiddelde over de wedstrijden van een jaar, waarbij ook de rating van afgelopen jaar wordt meegenomen wanneer de speler minder

dan zes wedstrijden speelt. “Voor iemand die veel speelt, hebben de wedstrijden aan het eind van het seizoen weinig invloed meer op hun rating, omdat we over zoveel wedstrijden middelen”, legt Wiersma uit.



Foto: KNLTB

De KNLTB heeft zelf al gekeken naar andere ratingsystemen. Zo was er bijvoorbeeld de rating van de Amerikaanse tennisbond, waarbij zelfs de setstanden ertoe doen. “Dat willen wij niet, omdat je rating dan achteruit kan gaan als je wél gewonnen hebt, maar met een klein verschil in setstanden”, vertelt Wiersma. “We vonden nog geen systeem dat helemaal aan onze wensen voldeed.” Wiersma heeft daarom het wensenlijstje van de KNLTB opgesteld en dat aan de wiskundigen voorgelegd.

“Ik tennis niet, maar ratings heb ik altijd interessant gevonden”, zo motiveert deelnemer Len Spek zijn keuze voor het tennisvraagstuk. Het was Speks eerste studiegroep, hij had er van andere promovendi goede verhalen over gehoord. De wiskundige werkt bij UTwente

aan een promotieonderzoek naar het simuleren van netwerken van hersencellen. In 2021 is de Studiegroep Wiskunde met de Industrie aan zijn universiteit. Het leek ook zijn begeleider daarom een goed idee dat Spek dit jaar mee zou doen. “Ik was onder de indruk hoeveel progressie je in een paar dagen maakt wanneer je als groep gefocust aan één probleem werkt”, vertelt de wiskundige naderhand. Dat is ook wat Wiersma na afloop zegt: “Mooi vond ik dat iedereen bereid was mee te gaan in de denkrichting van een ander. Daarbij stelden de wiskundigen elkaar steeds vragen, hoe pakt dit dan uit, wat doen we dan daarmee? Vragend oplossend, zo zou ik het proces omschrijven.”

Vooraf had Spek verwacht dat er tal van ratingsystemen voor dubbels in de literatuur te vinden zouden zijn, maar dat viel tegen. “We vonden wat informatie over ratings bij online computerspellen, die ook in teams worden gespeeld. Die ratings voldeden niet aan onze eisen.” Dus ontkwamen de wiskundigen er niet aan om toch zelf wat te bedenken. Daarvoor begonnen ze bij een oud, veelgebruikt en gerespecteerd ratingsysteem: de eloring, ontwikkeld door natuurkundige en schaker Árpád Élő in de jaren 60. Het systeem is de standaard bij schaken en dammen. Dit systeem rekent voor elke wedstrijd een winstverwachting uit, gebaseerd op het krachtverschil tussen de spelers. Wint de sterkere speler, dan is dat in lijn met de verwachting: de rating van de winnaar stijgt dan slechts een beetje. Wint de zwakkere speler echter tegen de verwachting in, dan heeft dat een groter effect. Spek: “Het mooie is dat je rating altijd stijgt als je wint, en daalt als je verliest, al is het soms maar een beetje.”

Het elosysteem is later verfijnd door statisticus Mark Glickman. Hij nam in zijn glickorating de onzekerheid over de speelsterkte van een speler mee. De speelsterkte van een echte wedstrijdvreter is immers met meer zekerheid bekend dan die van iemand die maar af en toe een wedstrijd speelt, of die een tijdje gestopt is en weer begint. “Als iemand met een onzekere rating speelt tegen iemand met een zekere rating, heeft die wedstrijd meer invloed op de rating van de onzekere speler dan op die

van de welbekende speler”, legt Spek het mooie van het glickosysteem uit. Het meenemen van onzekerheid is iets wat de KNLTB graag wil. In dit systeem is het niet erg als de bond de speelsterkte van een herintredende speler verkeerd inschat: de glickorating trekt dat snel weer recht, zonder grote schade voor de tegenstanders.

Het was de wiskundigen snel duidelijk dat zowel het elo- als het glickosysteem geschikt zijn voor de KNLTB. Nu moest de stap van enkel- naar de dubbelspel worden gemaakt, want beide systemen zijn ontworpen voor één tegen één. “We hebben flink gediscussieerd over hoe dat moest”, vertelt Spek. Moest je uitgaan van twee teams met een bepaalde sterkte die tegen elkaar spelen - als ware het een enkelspel tussen teams? Of was het beter om de wedstrijden te beschouwen als vier potjes van één iemand met drie anderen, waarvan er twee tegenwerken en een helpt? Spek: “Uiteindelijk zijn we toch maar gewoon uitgegaan van twee teams tegen elkaar, ook omdat dat beter uit te leggen is.” Want ook uitlegbaarheid was een eis van de KNLTB, om gemopper van spelers te voorkomen.

$$\left. \begin{array}{l} R_A \\ R_B \end{array} \right\} R_{AB} \implies R'_{AB} \left\{ \begin{array}{l} R'_A \\ R'_B \end{array} \right.$$

*Bij een dubbelspel wordt de persoonlijke rating van de spelers in één team gecombineerd tot één gezamenlijke rating. Na de wedstrijd wordt de nieuwe teamrating weer opgesplitst in twee individuele ratings.*

De aanpak werd als volgt: twee spelers vormen samen een team met één rating, gebaseerd op de gecombineerde rating van de beide spelers. Vervolgens wordt er met het elo- of glickosysteem een winstverwachting bepaald voor de wedstrijd, als ware het gewoon een wedstrijd van één tegen één. “Na de wedstrijd rolt er een nieuwe rating

voor de teams uit, zowel voor de winnaars als voor de verliezers. Die splits je uit naar de spelers”, vervolgt Spek. Bij deze aanpak zijn er twee belangrijke vragen. Hoe combineer je de ratings van de spelers vooraf en hoe verwerk je het resultaat van de wedstrijd in de individuele ratings? Spek: “Als de spelers even sterk zijn, is het makkelijk. Hun gezamenlijke rating is gelijk aan hun individuele ratings, en ze stijgen of dalen na afloop evenveel. Maar bij verschil in sterkte, wat is dan eerlijk? Neem je het gemiddelde van de individuele ratings? Of is vooral de sterkere speler bepalend voor de uitslag, of juist de zwakkere?”

Om daar inzicht in te krijgen, gebruikten de wiskundigen een dataset met uitslagen van de KNLTB. Spek: “Het gemiddelde van twee ratings blijkt een behoorlijk goede voorspeller van de uitslag, ook als het niveauverschil tussen twee spelers in één team groot is.” Hij geeft hierbij wel aan dat dit deels het gevolg kan zijn van het huidige ratingssysteem, dat beide spelers bij winst ongeveer gelijk beloont. “Dat zou je nog kunnen onderzoeken.”

De wiskundigen keken ook of bij gemengd dubbel de man of de vrouw meer bepalend is voor de uitslag, maar ook geslacht lijkt niet uit te maken. De wiskundigen raden de KNLTB dan ook voor de rating van een team gewoon de individuele ratings te middelen. Zij gaven de tennisbond ook de formules om de nieuwe gezamenlijke rating weer netjes te splitsen naar de spelers. “Het mooie was dat we hier een exacte oplossing voor konden geven”, vertelt Spek. Voor de fijnproevers: die formule is afgeleid uit een minimalisatie van het verlies van informatie over de wedstrijd.

De wiskundigen presenteerden uiteindelijk twee systemen: een dubbelspelersversie van het elosysteem en het complexere glickosysteem. “In het glickosysteem is de formule voor het splitsen van het resultaat niet zo eenvoudig uit te leggen aan de spelers”, zegt Spek. “Glicko heeft zeker onze voorkeur, omdat je er onzekerheid in speelsterkte in

meeneemt”, zegt Wiersma na afloop. “Het principe van het systeem krijgen we wel uitgelegd zonder die formules.”

Het is niet zo dat het ratingsysteem klaar is voor implementatie. Spek: “De bond moet nog wel keuzes maken.” Zo groeit de onzekerheid van een rating van een niet-spelende tennisser met de tijd, maar hoe snel dat gaat is een parameter die de bond kan kiezen. “Kies je een snelle groei van de onzekerheidsfactor, dan zullen ratings sterker fluctueren.” Spek legt uit dat je daar misschien niet voor alle type spelers dezelfde keuze wil maken. “Zo spelen professionele tennissers maar weinig KNLTB-wedstrijden, terwijl hun niveau vrij constant is. Daar wil je dat de onzekerheid langzaam oploopt. Bij jeugdspelers kan je meer onzekerheid inbouwen, omdat jeugd zich snel ontwikkelt”, zegt Spek. Hij adviseert om hier keuzes te maken op basis van speeldata. Wiersma benadrukt dat de veranderingen in de rating moet kloppen met het gevoel: “Spelers willen een rating die meebeweegt met de wedstrijden, maar niet alle kanten opschiet.”

De KNLTB verwacht de resultaten te implementeren, maar niet van vandaag op morgen. Wiersma wil graag meer data-analyse doen voordat er beslist wordt hoe je twee ratings tot één combineert: “Het kan zijn dat de sterkere speler toch iets meer invloed heeft. Daar zou ik graag nog eens goed naar kijken voordat we besluiten of we de ratings middelen, of toch een iets andere verhouding gebruiken.”

Wiersma verwacht dat ook de rating van de singles verandert. “Het lijkt me wat gek om met twee systemen naast elkaar te werken.” Wiersma overlegt met de studiegroep wat er nog moet gebeuren voor de implementatie kan beginnen. Het uitwerken van de details wordt misschien een afstudeerproject. “Ik heb bewondering voor wat de groep in een week heeft bereikt”, zegt Wiersma. “Het was een project waarvan we al een tijdje wisten dat we er wat mee wilden. Door deze week heeft het vleugels gekregen.”