



computerschaak

DE SPEELSTERKTE VAN EEN SCHAAKPROGRAMMA 2

In het boek 'Chess Skill in Man and Machine', second edition, hebben J.H. Condon en Ken Thompson hun werk beschreven onder de titel: **Belle**. Het is een imponerend relaas van twee onderzoekers, dat besloten wordt met de verslaglegging van het Belle-toernooi dat we vorige maand in SN bespraken. Bij een bezoek aan Bell Laboratories vertelde Ken eens: "Nu heb ik zoveel gegevens; ik weet niet wat de juiste manier is om ze te presenteren, laat staan welke gevolgtrekkingen er gemaakt mogen worden".

Nieuwe presentaties

Bij de presentatie bestaan er twee problemen: (i) Welke rating moet een bepaalde Pn toegekend worden? (ii) Wat is de betekenis van de ontmoeting P9 - P3: 19½-½ (zie vorige aflevering)? In het boek besloten de schrijvers tenslotte de in hun ogen relevante gegevens als volgt te publiceren.

RATING	P4	P5	P6	P7	P8	P9
P4	1235	X	5	0,5	0	0
P5	1570	15	X	3,5	3	0,5
P6	1826	19,5	16,5	X	4	1,5
P7	2031	20	17	16	X	5
P8	2208	20	19,5	18,5	15	X
P9	2328	20	20	18,5	16	14,5

We merken op dat de resultaten van P3 verdwenen zijn en dat de ratings veranderd zijn t.o.v. de vorige aflevering. Evenals Condon en Thompson willen we de (belangrijkste) conclusies uit deze tabel aan de lezer overlaten. We geven hierbij echter wel de volgende aanvullende informatie.

- Deze ratings sluiten zo dicht mogelijk aan bij de ratings van Belle (gezet op de verschillende diepten), die verkregen zijn in menselijke toernooien.
- De programma's speelden zonder 'contempt'-faktor (= ongehoorzaamheidsregel). Dit betekent als P9 iets minder staat tegen P4, het programma P9 zal proberen remise te maken; met een contempt-faktor van één pion zou het zelfs bij achterstand van een pion proberen te winnen.
- De evaluatiefunctie is voor ieder programma Pn hetzelfde.

Overeenkomst in "denken"

De bewering dat iedere speler een rating heeft die overeenkomt met zijn inzicht, is duidelijk onjuist, want zoals bekend zijn er

heel wat meer factoren van belang bij het spelen van een partij dan inzicht alleen. Zo heeft ook niet ieder programma een rating die overeenkomt met zijn evaluatiefunctie. Wat is hierbij het geval?

Laten we als voorbeeld eens nemen de bekende ontmoeting Nuchess-Belle (WK Linz, 1980). Het is een aantrekkelijke partij die we zonder commentaar afdrukken.

1. ♠4 e5 2. e4 exf4 3. Pf3 d5 4. exd5 Pf6 5. Lb5 + e6 6. dxc6 Pxc6 7. d4 Ld6 8. 0-0 0-0 9. Pc3 Lg4 10. Lc4 Lb4 11. Pe2 Ld6 12. Pxf4 Lxf3 13. Txf3 Lc5 14. c3 Pe5 15. Le2 Pxf3 16. Lxf3 Ld6 17. Lxb7 Tb8 18. Df3 Te8 19. Lc6 Tel + 20. Kf2 Txc1 21. Txc1 Txb2 + 22. Kg1 Dc7 23. g3 Lxf4 24. gxf4 Txa2 25. Tb1 Dd6 26. Lb5 Db8 27. Ld3 De8 28. Kh1 a5 29. Lb5 Dd8 30. Lc4 Te2 31. Tb7 Da8 32. Lxf7 + Kh8 33. Ld5 De8 34. Tb1 Pxd5 35. Dxd5 Txc3 36. De5 Da8 + 37. d5 Tc8 38. Tg1 Db7 39. f5 Df7 40. Tb1 a4 41. Tb8 Txb8 42. Dxb8 + Dg8 43. De5 a3 44. f6 h6 45. fxg7 + Dxc7 46. De8 + Kh7 47. De4 + Kh8 48. De8 + Kh7 49. De4 + Kh8 50. De8 + remise.

Beide programma's speelden "in hun eigen gedachtenwereld" ongeveer dezelfde partij. De evaluatie van de stelling kwam voortdurend met elkaar overeen. Bovendien voorspelde Nuchess van de 43 zetten (7 zetten zaten in de openingsbibliotheek) er maar liefst 26 goed. Volgens deze tendens van het langs dezelfde lijnen "denken" is het niet verwonderlijk dat de partij in remise eindigde. Een mogelijke gedachte hierbij is: bij grootmeesters is dit niet anders, ofschoon zij, filosofisch gezien, een geheel ander spel spelen dan computers (brein-in-het-vat discussie).

Ratings en extrapolaties

De verklaring van Enklaar/Dekker (SN 9/1983, pp. 290-291) voor hun twijfel over de verhoging in speelsterkte bij toenemende zoekdiepte luidde dat het steeds ging om versies van hetzelfde programma. "Ze 'denken' dus op dezelfde manier, alleen de ene versie kijkt dieper dan de andere". Deze verklaring is plausibel en aanvechtbaar tegelijk. Het probleem van de extrapolatie ligt nl. veel dieper.

De evaluatiefunctie van de Belle Bn-programma's bevat, aldus Condon en Thompson, slechts die elementen waarover een 1600-speler beschikt. Het "weet" veel over koningsveiligheid en pionnenformaties, maar heel weinig over zwakke velden en samenwerking van stukken. Het is duidelijk dat P8 en P9 zeer veel voordeel zouden trekken uit een evaluatiefunctie die deze aspecten ook in aanmerking zou nemen,

terwijl zo'n uitgebreide evaluatiefunctie aan P5 en P6 niet besteed is.

Konklusies

Betekent dit dan niet dat P8 en P9 met zo'n uitgebreide evaluatiefunctie nog sterker gaan spelen dan uit de extrapolatie zou volgen? Neen, want zo'n uitgebreide evaluatiefunctie kost juist veel meer tijd en het aantal stellingen neemt al schrikbarend (exponentieel) toe op niveau 8 en 9. Maar zou de oorspronkelijke konklusie dan niet gehandhaafd kunnen worden over de stijging van de speelsterkte? Vermoedelijk ook niet want voor een nog dieper kijkend programma zal de gehanteerde eenvoudige evaluatiefunctie misschien wel eens steeds minder discriminerend blijken te kunnen werken. Kortom, we hebben te maken met een zeer interessant probleem, waaraan de computerschaakwereld voorlopig zijn tanden zal stukbijten, terwijl de gewone schaakwereld in spanning de resultaten afwacht.

JAAAP VAN DEN HERIK

Schakend Nederland - april 1984

Prof. dr. H.J. van den Herik: De speelsterkte van een schaakprogramma 2