

Spieglein, Spieglein an der Wand ...

Chrilly Donninger über Erfolgskontrollen in der Schachprogrammierung

„Wer ist die Schönste im ganzen Land? Sie sind's, Frau Königin, aber Schneewittchen über den sieben Bergen bei den sieben Zwergen ist tausendmal schöner als Sie“ (aus Gebrüder Grimm: Schneewittchen). Wie einfach wäre Schachprogrammierung, wenn es auch für Computerschach einen derartigen Spiegel gäbe. Spieglein, Spieglein an der Wand, was ist die beste Nimzo-Version im ganzen Land? Die vom 5. Mai 1996 – aber die Entwicklungsversion von Weihnachten 95 ist 14 Elo besser.

Die meisten Schachprogrammierer leiden keineswegs an einem Mangel an Ideen. Das eigentliche Problem besteht darin, wie man feststellt, ob die Idee etwas taugt. Bei 2400 Elo muß man schon eine relativ gute Idee haben, um nochmals 10 Elo zuzulegen. Als gelernter Statistiker weiß ich, daß man erst nach 1000 Spielen (!) mit 95 Prozent Wahrscheinlichkeit sagen kann, ob die Idee wirklich ein Fortschritt war. Diese Berechnung ist insofern noch optimistisch, als sie von einer einheitlichen Elozahl ausgeht.

Bei der Nimzo-Entwicklung teste ich eine neue Version, eine neue Idee, zuerst auf meinen 486/50 PCs mit der Einstellung 30 Sekunden pro Zug. Wenn ich nach ein paar Spielen ein gutes Gefühl habe, wird die Version an die übrigen Tester der Nimzo-Crew weitergegeben. Selbst bei sehr vielen Spielen sind die von den Testern anschließend gemeldeten Resultate oft widersprüchlich. Sie hängen ganz offensichtlich von der vom Tester verwendeten Hardware, der Zeiteinstellung und den ausgewählten Gegnern ab (manche Tester sind sogar vom Einfluß der Erdstrahlen auf das Spielergebnis überzeugt). Was gegen Genius gut ist, kann sich gegen Fritz als Rohrkrepieler erweisen.

So bin ich z.B. mit CSS-Tester Jürgen Faas einer Meinung, daß die Vorabversion, die ich ihm Ende 95 habe zukommen lassen, sowohl bei Teststellungen als auch im Spiel gegen Genius besser ist als die endgültige, im Handel erhältliche Version. Allerdings kam diese Vorabversion zumindest auf meinen Rechnern gegen Fritz regelmäßig unter die Räder. Nachdem ich gegenüber Frederic Friedel einmal großspurig erklärt hatte, „wenn man nicht einmal den Fritz putzt, sollte man mit der Schachprogrammierung aufhören“, konnte ich diese Version unmöglich veröffentlichen. Tatsächlich

macht die offizielle Version gegen Fritz eine ganze gute Figur. Mit Genius tut sie sich dagegen wesentlich schwerer als das Nimzo-Christkind.

Offensichtlich herrschen auch auf den 200 MHz Pentium-Pros des Nimzo-Chefmechanikers Wolfgang Zugrav andere Gesetze als auf meinem nun schon etwas mickrigen 486ern.

Besser zwei gute Ideen als eine?

Selbst wenn man über den ersten Berg ist und wieder ein paar Elo zugelegt hat, ist die Sache noch lange nicht gelaufen. Beim Ausprobieren der (über)nächsten guten Idee kann es unter Umständen wie beim Mensch-Ärgere-Dich-Nicht „zurück an den Start“ heißen. Zwei gute Ideen in Kombination sind nämlich manchmal ein Desaster.

Ein klassisches Beispiel dafür sind die vom Deep-Blue-Team erfundenen „Singular Extensions“. In der Suche werden alle Varianten erweitert, tiefer berechnet, in denen es nur einen guten, forcierten Zug gibt. Nach den ersten Berichten des Teams war das nicht die übliche gute 10 Elo-Idee, sondern ein 100 Elo-Genieblitz. Nach weiteren Untersuchungen pendelte sich der vom Deep-Blue-Team veröffentlichte Wert auf das ortsübliche Ideenniveau ein.

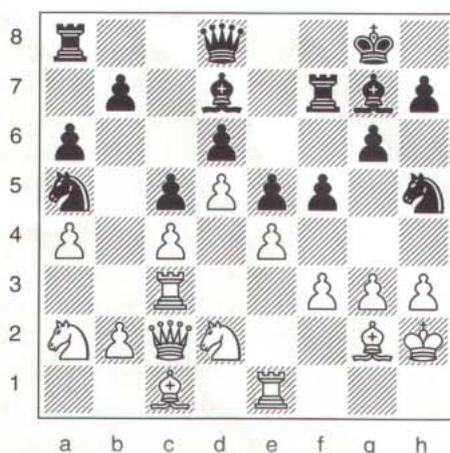
Der englische Computerwissenschaftler Don Beal widmete den Singular Extensions eine ausführliche Studie. Sein Resultat: Fügt man einem Programm, das ohne Sucherweiterungen rechnet, die Singular Extension-Technik hinzu, wird es tatsächlich um mindestens 100 Elo besser. Allerdings erweitern praktisch alle Programme bei Schachgeboten, die meisten auch bei Figurenabtäuschen und bei Drohungen. Durch diese tieferen Varianten, die frühzeitige Erkennung von Figurenverlust bzw. Matt, spielt das Programm gegenüber einer Version ohne Erweiterungen um 100 bis 200 Elo stärker. Die Kombination beider Techniken ist aber deutlich schlechter als die Singular-Extensions oder die Matt/Schlagzug-Erweiterungen alleine. Die Ergebnisse von Beal stimmen ziemlich genau mit meinen eigenen Experimenten zu diesem Thema überein. (Die vom Deep-Blue Team gemessene geringe Verbesserung hängt möglicherweise mit dem oben beschriebenen Geschwindigkeits-Effekt zusammen. Die Deep-Blue-Hardware rechnet ca. tausendmal schneller als ein Pentium-Pro.)

Zahlenmagie

Wahrscheinlich sind Schachprogrammierer nicht die einzigen, die sich in ihren Träumen den Schneewittchen-Spiegel wünschen. Die Verfasser von Testberichten sind in der Regel noch viel ärmer dran. Meistens bekommen sie kurz vor Redaktionsschluß vom verantwortlichen Redakteur ein Testexemplar des neuesten Programmes. Innerhalb von zwei Tagen sollen sie dann eine detaillierte, brilliant formulierte Beschreibung des Programmes nebst genauer Einschätzung der Spielstärke abliefern.

Rettung in dieser Not versprechen die Stellungstests. Nach 2 bis 3 Stunden Testarbeit erhält man einen genauen Testindex, der sich dann mittels Taschenrechner in die Elozahl umrechnen läßt. Zweifellos ergibt diese Methode einen besseren Eindruck, als in derselben Zeit zwei Testpartien zu spielen. Das Problem der meisten Tests besteht in ihrer einseitigen Ausrichtung auf bestimmte taktische Muster. Ein klassisches Testmuster ist z.B. eine durch ein Figurenopfer eingeleitete Kombination. Ein schönes Beispiel ist die Stellung Hsu-Nunn, Olympiade Manila, 1992.

Diagramm 1:
Hsu – Nunn, Olympiade Manila 1992



Schwarz am Zug

GM Nunn opferte hier zunächst mit **1...Sxg3! 2.Kxg4 Dh4+!** Springer und Dame. Nach **3.Kxh4 f4 4.Kg5 Tff8 5.b4 h6+ 6.Kxg6 Le8** ist Weiß matt.

Ein Programm schneidet bei diesem Testmuster gut ab, wenn es die anfänglichen Opfer stark erweitert und daher bereits auf niedriger Suchstufe in der Lage ist, solche Kombinationen durchzurechnen. Da sie das Salz in der Schachsuppe sind, gehen die Stellungstests davon aus, daß sich die tiefere Berechnung von Opfern auf die Spielstärke positiv auswirkt. Der wesentlichste Unterschied zwischen der Nimzo-Weihnachtsversion 95 und der im Handel erhältlichen Version 3.0 besteht in der Behandlung dieser Opfer. Die Weihnachtsversion erweitert in

der Stellung von Diagramm 1 bereits den „Einschlag“ Sxg3. Die Version 3.0 ignoriert dagegen etwaige Erweiterungen in den beiden ersten Halbzügen. D.h., die im Handel erhältliche Version ist in derartigen Opferkombinationen in der Regel schwächer als die ominöse Entwicklungsversion. Was hat mich zu diesem auf den ersten Blick kurios erscheinenden Schritt veranlaßt?

Erstens fehlt einem Schachprogramm das Stellungsgefühl eines John Nunn. Es erweitert auch vollkommen aussichtslose Opfer. Die tiefere Berechnung dieser sinnlosen Varianten kostet natürlich Zeit, das Programm erreicht eine geringere Suchtiefe. Unter Umständen werden daher andere Kombinationen oder Spielzüge nicht gefunden.

Der zweite, entscheidende Grund ist wesentlich subtiler. Nach meinen Beobachtungen hängt der Spielstil eines Schachprogramms nicht nur von der Bewertungsfunktion, sondern im gleichen Ausmaß von der Art der Erweiterungen bzw. der Selektivität im Suchbaum ab. Erweitert man Schlagzüge sehr stark, mutiert das vorher friedliche Programm plötzlich zum Holzfäller, der keinen Schlagabtausch an sich vorbeigehen läßt. Erweitert man Schachgebote stärker, beginnt es nach dem Motto „gib immer Schach, es könnte Matt sein“ zu spielen.

Eine Frage des Stils

Varianten mit vielen Erweiterungen bieten dem Programm einfach mehr Möglichkeiten, sich die Welt nach seinen Vorstellungen herzurichten. So ist z.B. eine ungerade Anzahl von Halbzügen in der Hauptvariante für das Programm besser als eine gerade Anzahl, da es dabei einen Halbzug mehr als der Gegner ausführt. Kommt das nicht hin, wird es versuchen, durch ein (unnötiges) Zwischenschach oder einen in Wirklichkeit gar nicht so günstigen Figurenabtausch diesen Zustand herbeizuführen.

Erweitert man daher Abtausche zu Beginn der Variante, so verflacht das Spiel meist sehr schnell. Ohne diese Erweiterungen hält das Programm die Spannung besser aufrecht. Es legt lieber Zwischenzüge ein, bevor es ans Abholzen geht. Diese Zwischenzüge können zwar manchmal auch ein Zuviel des Guten sein, aber im Großen und Ganzen glaube ich doch, daß die Verbesserung des Spielstils den Nachteil, ab und zu ein Opfer später (zu spät) zu sehen, aufwiegt. Die Technik hat auch keine negative Auswirkungen auf das Erkennen von gegnerischen Opfern, da diese bereits genügend weit von der Ausgangsstellung entfernt sind.

Diese Umstellung war in der Nimzo-Crew nicht unumstritten. Ein Teil meinte „du schaufelst dir dein eigenes Grab, die Tester werden dich zerreißen“,

während die andere Fraktion mit dem Argument „es spielt jetzt aber wirklich besser“ dafür war.

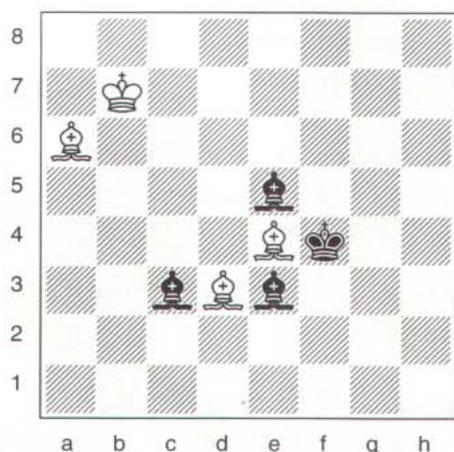
Meiner Meinung nach ist Nimzo keineswegs ein Teststellungs-Versager. Zum Beispiel liegt es im großen Endspieltest von F. Maresch mit 36 aus 40 an erster Stelle. Der Wiener Problemkomponist und ursprüngliche Nimzo-Skeptiker G. Semrad hat bereits eine eigene Mappe mit dem Titel NT (wie „Nimzo Triumph“) angelegt. Die Mappe enthält alle Teststellungen und Kompositionen, in denen Nimzo brilliert. Die NT-Mappe ist wesentlich dicker als die entsprechende NS- („Nimzo-Schwach“) Mappe.

Die Lieblingsstellung

Im Mai 1996 fand in Wien eine große Ausstellung zur Schachgeschichte statt. Die Exponate aus den letzten 1000 Jahren wurden an Hand der Partie Rubinstein-Grünfeld, Karlsbad 1928, präsentiert. In einem Radiointerview meinte der Ausstellungsgestalter, E. Strouhal, daß der letzte und entscheidende Zug von Rubinstein für Schachprogramme viel zu schwierig sei.

Bei dieser Ausstellung wurde Nimzo-3 erstmalig der Öffentlichkeit vorgestellt. Zum Gaudium der Anwesenden setzte ich ihm die Schlußstellung dieser berühmten Partie vor (siehe Diagramm 2). Nach ca. fünf Sekunden erschien in der Anzeige das für einen Computer „unmögliche“ Lb1!! mit einer Bewertung von fünf Bauern Vorteil für Weiß. Rubinstein wird es auch nicht wesentlich schneller gefunden haben. Seit dieser Episode ist die Position meine Lieblingsteststellung, die ich aus der Tasche ziehe, falls jemand es wagen sollte, die Qualitäten meines Geisteskindes in Frage zu stellen.

Diagramm 2:
Rubinstein – Grünfeld, Karlsbad 1928



Weiß am Zug gewinnt mit Lb1

Nimzo-3 liegt in der schwedischen Liste ein Stückchen vor Fritz. In den drei gemeinsam bestrittenen Turnieren waren sie einmal gemeinsam an der Com-



Unser Bild zeigt Chrillys Mitbewohnerin „Kamikaze“, die im bekannten BT2630-Test (vgl. CSS 2/94) auf 1730 Elo kommt. Glauben Sie nicht? Dann rechnen Sie doch einmal nach!

puterspitze (Aegon 96), zweimal hatte Nimzo die Nase vorn (Mikro-Weltmeisterschaft 95 und 96). Wenn daher z.B. jemand behauptet, Nimzo liege in seinem Test hinter Fritz, dann beweist das eigentlich nur, daß er einen schlechten Test verwendet. Der Test ist offensichtlich nicht in der Lage, die Spielstärke eines etwas anders gebauten, nicht auf Teststellungen hingetrimmten Programmes vorauszusagen.

Teststellungen sind meiner Meinung nach dazu da, Probleme in Schachprogrammen aufzeigen. Sie regen das Denken des Programmierers und des Testers an. So enthält die von Jürgen Faas ausgewählte Stellung in Diagramm 1 eine sehr harte und interessante Nuß, den „stillen“ Zwischenzug 4...Tff8. Vielleicht habe ich in den nächsten Jahren einmal eine gute Idee, wie man diese Nuß knackt.

Man sollte sich aber von der reinen Testzahl nicht übermäßig beeindruckt lassen. Nach einem anstrengenden Programmiertag spielen meine psychologische Betreuerin Kamikaze und ich noch gerne eine Partie Schach. Kamikaze hat 1730 BT-Elo, ich bin mit nur 1700 nationalen Elo geringfügig schwächer. Trotzdem gewinne ich fast jede Partie (Kamikaze wirft oft die Figuren um, wenn ihr die Stellung nicht mehr gefällt). Meine Frau meint, man könne doch von einer einfachen Hauskatze nicht mehr verlangen. Meinen Einwand, daß Kamikaze schließlich mehr Elopunkte hat als ich und nach allen Regeln der Statistik das Spiel mindestens ausgewogen gestalten müßte, versteht sie in typisch weiblicher Ignoranz nicht so recht.