

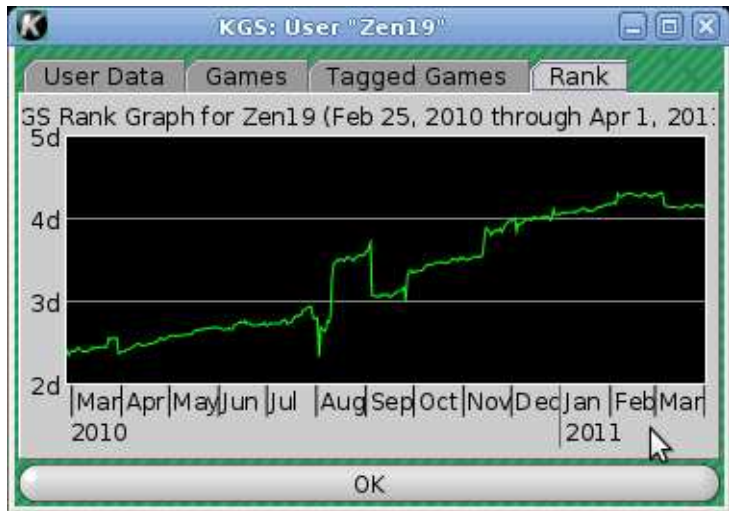
Durch Zufall zum Dan?

Die Monte-Carlo-Baumsuche

Johannes Waldmann
HTWK Leipzig

Potsdamer Panda, April 2011

Wer hätte das gedacht ...



(Zen, Programmierer: Ojima Yoji)

Spielbäume

zu jedem Spiel gehört ein Baum:

- ▶ Knoten = Spiel(-Position),
- ▶ Kante = Zug

Spielbäume

zu jedem Spiel gehört ein Baum:

- ▶ Knoten = Spiel(-Position),
- ▶ Kante = Zug

Abmessungen:

- ▶ Tiefe = maximale Länge einer Zugfolge
- ▶ Verzweigungsgrad = max. Anzahl verschiedener nächster Züge
- ▶ Anzahl Knoten = $\text{Grad}^{\text{Tiefe}}$

Spielwerte (Definition)

Spielwert (für jeweils anziehenden Spieler)
bestimmen von unten nach oben:

- ▶ Blatt: hat keine Nachfolger, kein weiterer Zug möglich,
Resultat auszählen, Gewinner bestimmen

Spielwerte (Definition)

Spielwert (für jeweils anziehenden Spieler)
bestimmen von unten nach oben:

- ▶ Blatt: hat keine Nachfolger, kein weiterer Zug möglich, Resultat auszählen, Gewinner bestimmen
- ▶ innerer Knoten:
 - ▶ ist gewonnen, falls ein Nachfolger verloren ist
 - ▶ ist verloren,

Spielwerte (Definition)

Spielwert (für jeweils anziehenden Spieler)
bestimmen von unten nach oben:

- ▶ Blatt: hat keine Nachfolger, kein weiterer Zug möglich,
Resultat auszählen, Gewinner bestimmen
- ▶ innerer Knoten:
 - ▶ ist gewonnen, falls ein Nachfolger verloren ist
 - ▶ ist verloren, falls jeder Nachfolger gewonnen ist

Spielwerte (Bestimmung)

Da das Spiel endlich ist, kann man

- ▶ für jeden Knoten (auch Startknoten) den Wert bestimmen
- ▶ und falls existiert, dann auch einen Gewinnzug

Spielwerte (Bestimmung)

Da das Spiel endlich ist, kann man

- ▶ für jeden Knoten (auch Startknoten) den Wert bestimmen
- ▶ und falls existiert, dann auch einen Gewinnzug

Theoretisch jedenfalls.

Praktisch ist der Baum beim Go aber viel zu groß.

Wie findet man dann den jeweils besten Zug?

Genäherte Bewertungen

den tatsächlichen Wert ersetzen durch etwas, das man (schnell) ausrechnen kann, und trotzdem nützliche Information liefert.

Genäherte Bewertungen

den tatsächlichen Wert ersetzen durch etwas, das man (schnell) ausrechnen kann, und trotzdem nützliche Information liefert.

Änderung des Wertebereiches: von zweiwertig (gewonnen/verloren) zu mehrwertig

z. B. reelle Zahlen, die Tendenz ausdrücken sollen: je größer, desto besser für anziehenden Spieler.

Genäherte Bewertungen

den tatsächlichen Wert ersetzen durch etwas, das man (schnell) ausrechnen kann, und trotzdem nützliche Information liefert.

Änderung des Wertebereiches: von zweiwertig (gewonnen/verloren) zu mehrwertig

z. B. reelle Zahlen, die Tendenz ausdrücken sollen: je größer, desto besser für anziehenden Spieler.

aber was heißt "besser"?

wie koppelt man Baum und Zahlenbewertung?

Klassisches Bewertungsverfahren

für verkürzte Spielbäume

- ▶ im Blatt: statische Bewertung (nach Materialeigenschaften)
- ▶ in inneren Knoten:
$$\text{wert}(p) = - \min\{\text{wert}(q) \mid p \rightarrow q\}$$

Klassisches Bewertungsverfahren

für verkürzte Spielbäume

- ▶ im Blatt: statische Bewertung (nach Materialeigenschaften)
- ▶ in inneren Knoten:
$$\text{wert}(p) = - \min\{\text{wert}(q) \mid p \rightarrow q\}$$

Bestreben, die Kinder q wegzulassen die $\text{wert}(p)$ nicht beeinflussen.

Klassisches Bewertungsverfahren

für verkürzte Spielbäume

- ▶ im Blatt: statische Bewertung (nach Materialeigenschaften)
- ▶ in inneren Knoten:
$$\text{wert}(p) = - \min\{\text{wert}(q) \mid p \rightarrow q\}$$

Bestreben, die Kinder q wegzulassen die $\text{wert}(p)$ nicht beeinflussen.

im Schach schon sehr lange sehr erfolgreich, im Go überhaupt nicht.

keine vernünftige und einfache
Materialbewertung, Verzweigungsgrad zu hoch.

Bewertung durch Simulation

simple Idee: um den Wert einer Spielposition zu bestimmen, wird von dort sehr oft zufällig weitergespielt.

vgl. Simulationen zur Bewertung von komplexen dynamischen Systemen, Finanzprodukten

Bewertung durch Simulation

simple Idee: um den Wert einer Spielposition zu bestimmen, wird von dort sehr oft zufällig weitergespielt.

vgl. Simulationen zur Bewertung von komplexen dynamischen Systemen, Finanzprodukten

Wert = Gewinn-Wahrscheinlichkeit (für anziehenden Spieler)

Bewertung durch Simulation

simple Idee: um den Wert einer Spielposition zu bestimmen, wird von dort sehr oft zufällig weitergespielt.

vgl. Simulationen zur Bewertung von komplexen dynamischen Systemen, Finanzprodukten

Wert = Gewinn-Wahrscheinlichkeit (für anziehenden Spieler)

nicht: erwartete Punktdifferenz

Bewertung durch Simulation

simple Idee: um den Wert einer Spielposition zu bestimmen, wird von dort sehr oft zufällig weitergespielt.

vgl. Simulationen zur Bewertung von komplexen dynamischen Systemen, Finanzprodukten

Wert = Gewinn-Wahrscheinlichkeit (für anziehenden Spieler)

nicht: erwartete Punktdifferenz aber: welche Wahrscheinlichkeitsverteilung?

multi-armed bandit problem

man kann beliebig zwischen verschiedenen Glücksspielmöglichkeiten (Lotterien) wählen, kennt aber deren Gewinnwahrscheinlichkeiten nicht.

multi-armed bandit problem

man kann beliebig zwischen verschiedenen Glücksspielmöglichkeiten (Lotterien) wählen, kennt aber deren Gewinnwahrscheinlichkeiten nicht.

nach mehreren Versuchen kennt man aktuelle Werte für Erwartungswert m_i und Varianz (erwartete Abweichung vom Mittelwert) σ_i

multi-armed bandit problem

man kann beliebig zwischen verschiedenen Glücksspielmöglichkeiten (Lotterien) wählen, kennt aber deren Gewinnwahrscheinlichkeiten nicht.

nach mehreren Versuchen kennt man aktuelle Werte für Erwartungswert m_i und Varianz (erwartete Abweichung vom Mittelwert) σ_i

Man spielt dann immer bei $\arg \max(m_i + \sigma_i)$

Selektive Baumsuche

Man betrachte jeden Knoten des Spielbaumes als multi-armed bandit problem,

d. h. man notiert für jedes Kind (Folgezug) Gewinnwahrscheinlichkeit und Varianz (Konfidenz-Intervall) $1/\sqrt{N}$ und spielt immer dort, wo Mittelwert – Konfidenz am kleinsten sind.

UCT = upper confidence bound for trees

Parameter für UCT

Mittelwert – C · Konfidenz

- ▶ C groß:

Parameter für UCT

Mittelwert – C · Konfidenz

- ▶ C groß: tendentiell Gleichverteilung auf Zügen (Exploration)
- ▶ C klein:

Parameter für UCT

Mittelwert – C · Konfidenz

- ▶ C groß: tendentiell Gleichverteilung auf Zügen (Exploration)
- ▶ C klein: Verteilung gehäuft auf wenige gute Züge (Exploitation)

Variante: C steigt (schwach) monoton mit Anzahl der Besuche im Vorgänger

Parameter für UCT

Mittelwert – C · Konfidenz

- ▶ C groß: tendentiell Gleichverteilung auf Zügen (Exploration)
- ▶ C klein: Verteilung gehäuft auf wenige gute Züge (Exploitation)

Variante: C steigt (schwach) monoton mit Anzahl der Besuche im Vorgänger (damit jeder Nachfolger nochmal drankommt).

Dann gilt: UCT konvergiert gegen exakten Spielwert

Selektive Expansion

Wenn man ab einem Knoten (Blatt) oft genug simuliert hat, wird dieses Blatt expandiert, d. h. alle Kinder-Blätter angehängt.

Selektive Expansion

Wenn man ab einem Knoten (Blatt) oft genug simuliert hat, wird dieses Blatt expandiert, d. h. alle Kinder-Blätter angehängt.

MC/UCT-methode: viele Playouts:

- ▶ in-tree: selektiere nach UCT
- ▶ below-tree: simuliere nach MC

Playouts

MC/UCT ist sehr allgemeines Verfahren. Um es für ein Spiel anzuwenden, braucht man nur:

Playouts

MC/UCT ist sehr allgemeines Verfahren. Um es für ein Spiel anzuwenden, braucht man nur:

- ▶ schnelle Simulationen (effiziente Zugausführung und Gewinnfeststellung)
Go: Freiheiten mitzählen, Steine schnell vom Brett nehmen, Gebiete schnell auszählen

Playouts

MC/UCT ist sehr allgemeines Verfahren. Um es für ein Spiel anzuwenden, braucht man nur:

- ▶ schnelle Simulationen (effiziente Zugausführung und Gewinnfeststellung)
Go: Freiheiten mitzählen, Steine schnell vom Brett nehmen, Gebiete schnell auszählen
- ▶ gute Simulationen (keine unsinnigen Züge in Playouts) — Go: nicht die eigene Augen zusetzen, Treppen auslesen

aber:

Playouts

MC/UCT ist sehr allgemeines Verfahren. Um es für ein Spiel anzuwenden, braucht man nur:

- ▶ schnelle Simulationen (effiziente Zugausführung und Gewinnfeststellung)
Go: Freiheiten mitzählen, Steine schnell vom Brett nehmen, Gebiete schnell auszählen
- ▶ gute Simulationen (keine unsinnigen Züge in Playouts) — Go: nicht die eigene Augen zusetzen, Treppen auslesen

aber: wenn man Simulationen zu gut (gründlich) macht, sind sie nicht mehr schnell genug

Typisches MC/UCT-Verhalten

- ▶ unsinnige Eröffnungszüge
(Gewinnwahrscheinlichkeiten aller
Anfangszüge nahe beieinander)

Typisches MC/UCT-Verhalten

- ▶ unsinnige Eröffnungszüge
(Gewinnwahrscheinlichkeiten aller Anfangszüge nahe beieinander)
- ▶ spektakuläre Fehler im Mittelspiel
(weil Gruppenstatus falsch eingeschätzt wird)

Typisches MC/UCT-Verhalten

- ▶ unsinnige Eröffnungszüge
(Gewinnwahrscheinlichkeiten aller Anfangszüge nahe beieinander)
- ▶ spektakuläre Fehler im Mittelspiel
(weil Gruppenstatus falsch eingeschätzt wird)
- ▶ Verzweiflungszüge (wenn Wahrscheinlichkeit deutlich unter $1/2$)

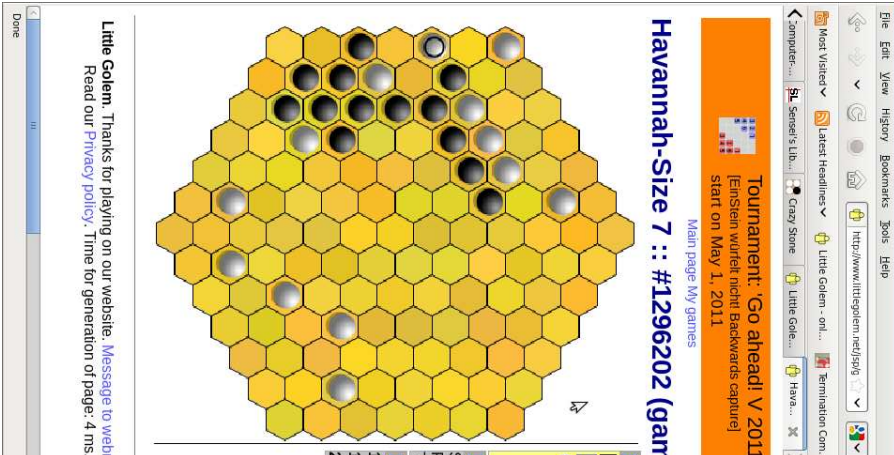
Typisches MC/UCT-Verhalten

- ▶ unsinnige Eröffnungszüge
(Gewinnwahrscheinlichkeiten aller Anfangszüge nahe beieinander)
- ▶ spektakuläre Fehler im Mittelspiel
(weil Gruppenstatus falsch eingeschätzt wird)
- ▶ Verzweiflungszüge (wenn Wahrscheinlichkeit deutlich unter $1/2$)
- ▶ ultra-sichere Züge (wenn Wahrscheinlichkeit deutlich über $1/2$)
(Punkte verschenken und mit 0.5 gewinnen)

Das Spiel Havannah

(Erfinder: Christian Freeling, 1979)

- ▶ Sechseck-Brett, Sechser-Nachbarschaft
- ▶ Spielziel: 1 Kreis, 2 Ecken, 3 Seiten



The screenshot shows a web browser window with a Havannah game board and tournament information. The game board is a large hexagonal grid of yellow and orange hexagons, with several black and white stones placed on it. The tournament information is displayed in an orange banner at the top of the page, reading: "Tournament: 'Go ahead! V 2011 [EinStein würfelt nicht! Backwards capture] start on May 1, 2011". Below the banner, the text "Main page: My games" is visible. The browser's address bar shows the URL "http://www.littlegolem.net/spg/". The browser's toolbar includes buttons for "Most Visited", "Latest Headlines", "SL Sensei's Lib...", "Crazy Stone", "Little Golem...", "Little Golem...", "Hava...", and "Termination Com...". The browser's status bar at the bottom shows "Done".

Havannah-Size 7 :: #1296202 (game)

Main page: My games

Tournament: 'Go ahead! V 2011
[EinStein würfelt nicht! Backwards capture]
start on May 1, 2011

Little Golem. Thanks for playing on our website. Message to webmaster.
Read our Privacy policy. Time for generation of page: 4 ms

Computer-Havannah, Computer-Go

Havannah hat viele Aspekte von Go, evtl. einfacher:

- ▶ viele Zugmöglichkeiten
- ▶ keine Schlag-Züge
- ▶ keine offensichtliche Bewertungsfunktion

Computer-Havannah, Computer-Go

Havannah hat viele Aspekte von Go, evtl. einfacher:

- ▶ viele Zugmöglichkeiten
- ▶ keine Schlag-Züge
- ▶ keine offensichtliche Bewertungsfunktion

seit 2009 Havannah bei
ICGA-Computer-Olympiade,

einige Programme spielen auf
<http://www.littlegolem.net>

Computer-Havannah, Computer-Go

Havannah hat viele Aspekte von Go, evtl. einfacher:

- ▶ viele Zugmöglichkeiten
- ▶ keine Schlag-Züge
- ▶ keine offensichtliche Bewertungsfunktion

seit 2009 Havannah bei
ICGA-Computer-Olympiade,

einige Programme spielen auf
<http://www.littlegolem.net>

konkreter Anreiz: eine Wette über 1000 EUR
(Christian : Computer 10 : 0) Angebot bis 2012.

Havannah-Playouts

- ▶ Muster (lokal gute Form)
- ▶ Treppen (direkte und Springer-Treppen)

All Moves As First

Ein Playout liefert Information über den ersten Zug. — Schade, viel Wissen verschenkt.

Deswegen: Zugumstellungen annehmen, jeder schwarze Zug könnte der erste von Schwarz gewesen sein

All Moves As First

Ein Playout liefert Information über den ersten Zug. — Schade, viel Wissen verschenkt.

Deswegen: Zugumstellungen annehmen, jeder schwarze Zug könnte der erste von Schwarz gewesen sein

Informationen über alle Züge im Playout im Spielbaum propagieren

All Moves As First

Ein Playout liefert Information über den ersten Zug. — Schade, viel Wissen verschenkt.

Deswegen: Zugumstellungen annehmen, jeder schwarze Zug könnte der erste von Schwarz gewesen sein

Informationen über alle Züge im Playout im Spielbaum propagieren

AMAF-Werte zur Initialisierung der Werte neuer Blatt-Knoten

Nützliche Eigenschaften UCT

robust, flexibel:

- ▶ aus der Wurzel kann man immer eine guten Zug ablesen (in min/max oder alpha/beta: erst am Ende der Suche)
- ▶ wenn man mehr Zeit hat, kann man einfach weitersuchen (mehr Playouts)

Nützliche Eigenschaften UCT

robust, flexibel:

- ▶ aus der Wurzel kann man immer eine guten Zug ablesen (in min/max oder alpha/beta: erst am Ende der Suche)
- ▶ wenn man mehr Zeit hat, kann man einfach weitersuchen (mehr Playouts)

skalierbar, parallelisierbar: wenn man mehr Rechenkerne hat, dann simultane Playouts

- ▶ in Blättern (eines Baumes)
- ▶ in der Wurzel (getrennte Bäume)

Dynamisches Komi

In Spielen mit Vorgabe-Steinen denkt das Programm (anfangs):

- ▶ jeder schwarze Zug gewinnt, jeder weiße Zug verliert,

und spielt Unsinn.

Dynamisches Komi

In Spielen mit Vorgabe-Steinen denkt das Programm (anfangs):

- ▶ jeder schwarze Zug gewinnt, jeder weiße Zug verliert,

und spielt Unsinn.

Ausweg: Nullpunktverschiebung durch Komi, bis Gewinnwahrscheinlichkeit etwa bei $1/2$ liegt.

Komi bei jedem Zug neu bestimmen

Gewinnabstände bei Havannah

- ▶ wieviel Steine muß P_1 setzen, um zu gewinnen, wenn P_2 immer aussetzt?

Gewinnabstände bei Havannah

- ▶ wieviel Steine muß P_1 setzen, um zu gewinnen, wenn P_2 immer aussetzt?
- ▶ ... um einen **Rahmen** zu erhalten

Rahmen: untrennbare Konfiguration (Rösselsprünge), die zum Gewinn führen wird.

Gewinnabstände bei Havannah

- ▶ wieviel Steine muß P_1 setzen, um zu gewinnen, wenn P_2 immer aussetzt?
- ▶ ... um einen **Rahmen** zu erhalten

Rahmen: untrennbare Konfiguration (Rösselsprünge), die zum Gewinn führen wird.

statische Positionsbewertung durch Vergleich dieser Zahlen möglich, aber riskant:

Gewinnabstände bei Havannah

- ▶ wieviel Steine muß P_1 setzen, um zu gewinnen, wenn P_2 immer aussetzt?
- ▶ ... um einen **Rahmen** zu erhalten

Rahmen: untrennbare Konfiguration (Rösselsprünge), die zum Gewinn führen wird.

statische Positionsbewertung durch Vergleich dieser Zahlen möglich, aber riskant:

- ▶ langer Rahmen verliert gegen kurze Kette

Gewinnabstände bei Havannah

- ▶ wieviel Steine muß P_1 setzen, um zu gewinnen, wenn P_2 immer aussetzt?
- ▶ ... um einen **Rahmen** zu erhalten

Rahmen: untrennbare Konfiguration (Rösselsprünge), die zum Gewinn führen wird.

statische Positionsbewertung durch Vergleich dieser Zahlen möglich, aber riskant:

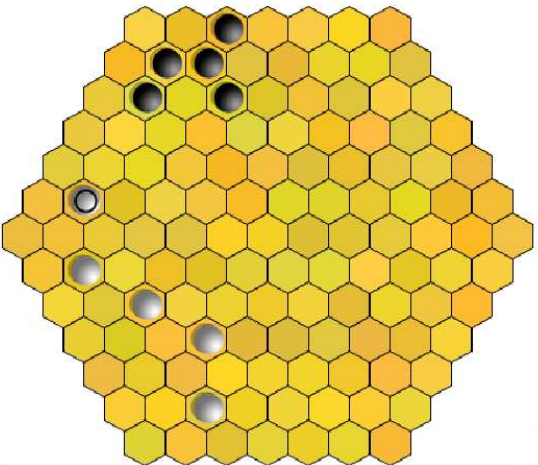
- ▶ langer Rahmen verliert gegen kurze Kette
- ▶ lokale (Kreis-)Drohungen können Abstände verändern

Lokal gegen Global (Aufgabe)

2010 September Reversi
monthly cup is finished
Winner: HVAGO
[Main page](#) [My games](#)

Summary
Player: 6/6
Your games: 6/6
Players/Games: 22691/131

Ring von
22691/131



Ring von
22691/131
Start
Finis
Tour
1.A3
15.C3
27.B7

Little Golem. Thanks for playing on our website. [Message to webmaster](#)
[Privacy policy](#). Time for generation of page: 2 ms.

Done

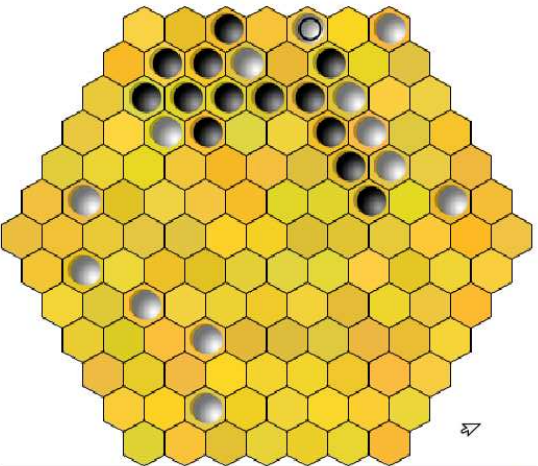
Lokal gegen Global (Lösung)



Tournament: 'Go ahead! V 2011'
[Einstein würfelt nicht Backwards capture]
start on May 1, 2011

Main page My games

Havannah-Size 7 :: #1296202 (game t



Pl
? Ring
G
Start
Finist
Tour
LI
LA3
15.C3
27.B7

Little Golem. Thanks for playing on our website. Message to webmaster
Read our [Privacy policy](#). Time for generation of page: 4 ms.

Done