

# Backgammon

Tobias Krönke

Technische Universität Darmstadt  
Fachbereich Informatik  
Fachgebiet Knowledge Engineering

Seminar zu Knowledge Engineering und Lernen in  
Spielen, 2010

# Gliederung

## Einführung

- Backgammon
- Lösungsansätze

## TD-Gammon

- Steine ziehen
- Den Einsatz verdoppeln

## Evaluation

- Spielstärke
- Warum es so gut funktioniert

## Zusammenfassung

### Einführung

- Backgammon
- Lösungsansätze

### TD-Gammon

- Steine ziehen
- Den Einsatz verdoppeln

### Evaluation

- Spielstärke
- Warum es so gut funktioniert

### Zusammenfassung

# Gliederung

## Einführung

Backgammon

Lösungsansätze

## TD-Gammon

Steine ziehen

Den Einsatz verdoppeln

## Evaluation

Spielstärke

Warum es so gut funktioniert

## Zusammenfassung

Backgammon

Tobias Krönke

### Einführung

Backgammon

Lösungsansätze

### TD-Gammon

Steine ziehen

Den Einsatz verdoppeln

### Evaluation

Spielstärke

Warum es so gut funktioniert

### Zusammenfassung

# Startaufstellung

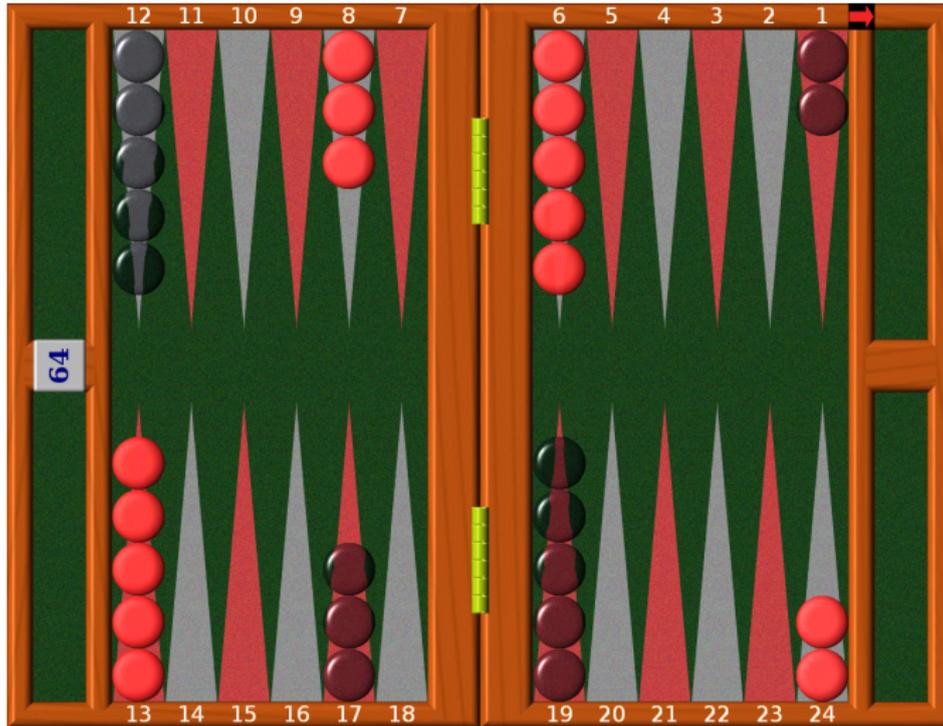


Abbildung: GNU Backgammon aus schwarzer Sicht

Backgammon

Tobias Krönke

Einführung

Backgammon

Lösungsansätze

TD-Gammon

Steine ziehen

Den Einsatz verdoppeln

Evaluation

Spielstärke

Warum es so gut funktioniert

Zusammenfassung

# Spielziel

Backgammon

Tobias Krönke

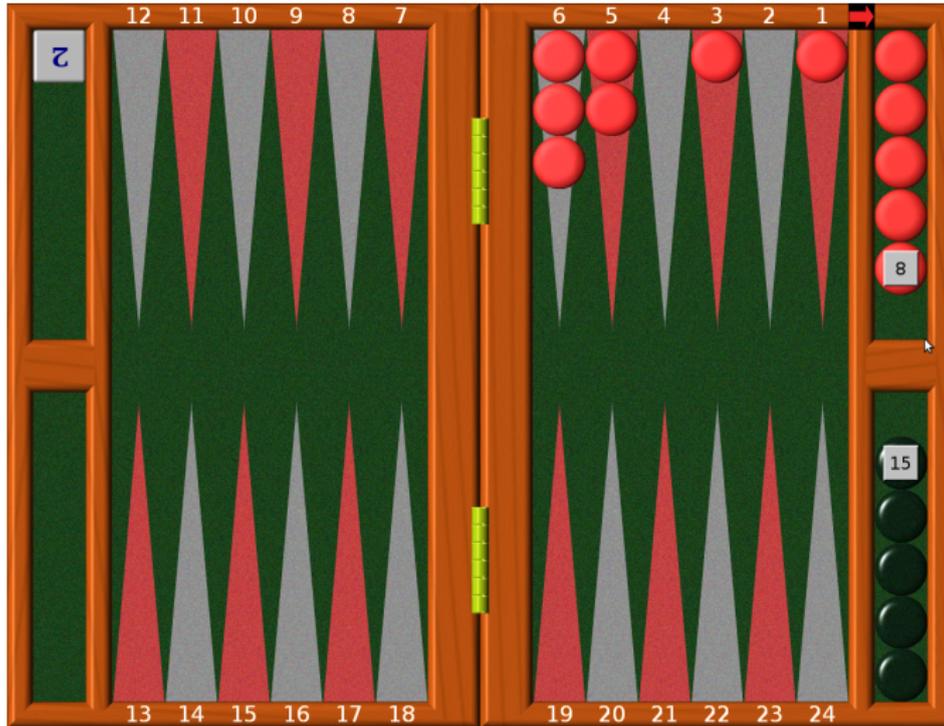


Abbildung: Schwarz gewinnt zwei Punkte

Einführung

Backgammon

Lösungsansätze

TD-Gammon

Steine ziehen

Den Einsatz verdoppeln

Evaluation

Spielstärke

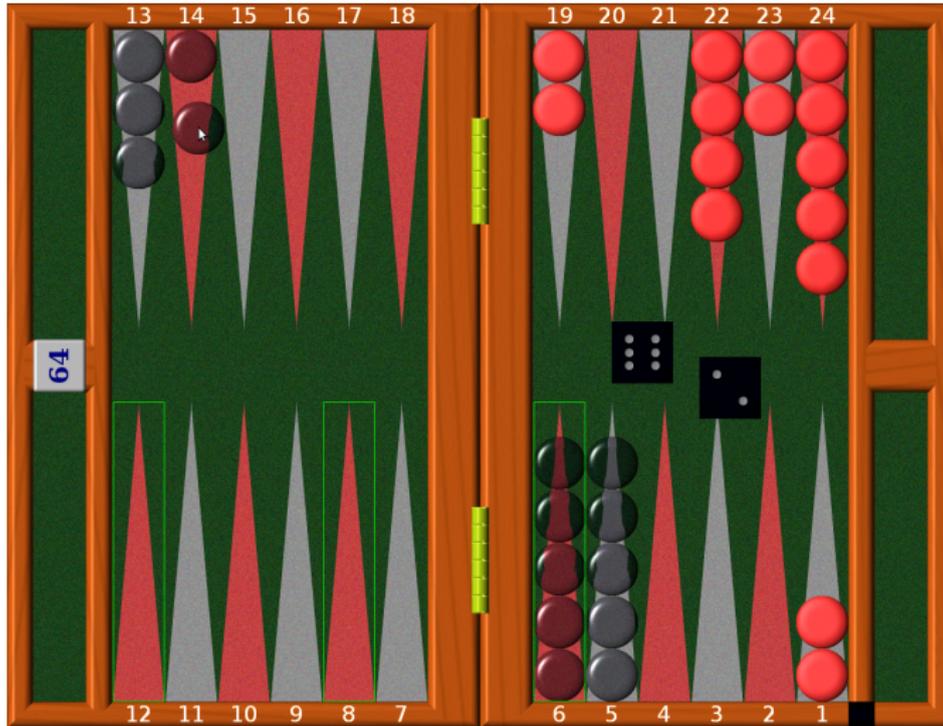
Warum es so gut funktioniert

Zusammenfassung

# Steine bewegen

Backgammon

Tobias Krönke



Einführung

Backgammon

Lösungsansätze

TD-Gammon

Steine ziehen

Den Einsatz verdoppeln

Evaluation

Spielstärke

Warum es so gut funktioniert

Zusammenfassung

Abbildung: Mögliche Züge für Schwarz von 14 aus

## Steine ziehen

- ▶ Zugzwang
- ▶ Einzelne Steine (= Blots) sind schlagbar
  - ▶ Sie müssen dann beim Start neu beginnen
  - ▶ Vorher darf kein anderer Stein bewegt werden

## Einsatz verdoppeln

- ▶ Höchstens einmal pro Zug vor dem Würfeln
- ▶ Besitzer des Verdopplungswürfels kann anbieten, den Einsatz zu verdoppeln
- ▶ Gegner kann ablehnen → verliert aktuellen Einsatz
- ▶ Gegner kann annehmen → wird Besitzer des Verdopplungswürfels

## Mathematisch

- ▶ Mehr als  $10^{20}$  Spielzustände
- ▶ Zwei 6er-Würfel  $\rightarrow$  21 Kombinationen
- ▶ Im Schnitt 20 Möglichkeiten pro Halbzug  $\rightarrow$  Hunderte Folgezustände

## Strategisch

- ▶ Blots schlagen
- ▶ Blockaden
- ▶ Endgame-Race
- ▶ ...

## Zustandsbewertungsfunktion

- ▶ Modell und Lernverfahren frei
- ▶ Ausgabe ist Wahrscheinlichkeit für „Schwarz / Rot gewinnt einfachen / doppelten Einsatz“
- ▶ Backgammons ignoriert, da zu selten
- ▶ Wähle Zug mit bestem Folgezustand

## Neurales Netz

- ▶ Rohe Boarddaten und strategische Konzepte als Inputs
- ▶ Backpropagation aufgezeichneter Expertenzüge → Supervised Learning
- ▶ Gewann die International Computer Olympiad 1989

## Wieder neurales Netz

- ▶ Inputs von Neurogammon
- ▶ Inferenz durch Spiele gegen sich selbst → Reinforcement Learning

## Trainingsdaten ohne Expertenwissen

- ▶ Sequenz von Spielzuständen  $x_1, x_2, \dots, x_f$
- ▶ Finale Punktevergabe  $Y_f$
- ▶ Spezielle Features als Ausnahme

**Problem:** Welche Halbzüge (= Zustandstransitionen) waren gut / schlecht? → Temporal Difference Learning

## Supervised Learning emulieren

- ▶ Verwende Voraussage für Folgezustand als Zielwert für aktuellen Zustand
- ▶ Propagiere den Fehler in der Zeit zurück

# Gliederung

## Einführung

Backgammon

Lösungsansätze

## TD-Gammon

Steine ziehen

Den Einsatz verdoppeln

## Evaluation

Spielstärke

Warum es so gut funktioniert

## Zusammenfassung

Backgammon

Tobias Krönke

### Einführung

Backgammon

Lösungsansätze

### TD-Gammon

Steine ziehen

Den Einsatz verdoppeln

### Evaluation

Spielstärke

Warum es so gut funktioniert

### Zusammenfassung

$$w_{t+1} - w_t = \alpha(Y_{t+1} - Y_t) \sum_{k=1}^t \lambda^{t-k} \nabla_w Y_k$$

Abbildung: TD-Gammon Updaterregel TD( $\lambda$ ) [Tesauro2002]

Symbol	Erklärung
$t$	Zeitpunkt in der Zustandssequenz
$w$	Gewichte des neuronalen Netzes
$Y_t$	Ausgabe bei Input $x_t$
$\alpha$	Lernrate
$0 \leq \lambda \leq 1$	Wie stark TD-Fehler der Vergangenheit korrigiert werden

**Tabelle:** Erläuterung zu TD( $\lambda$ )

Einführung

Backgammon

Lösungsansätze

TD-Gammon

Steine ziehen

Den Einsatz verdoppeln

Evaluation

Spielstärke

Warum es so gut funktioniert

Zusammenfassung

## Einführung

Backgammon

Lösungsansätze

## TD-Gammon

**Steine ziehen**

Den Einsatz verdoppeln

## Evaluation

Spielstärke

Warum es so gut funktioniert

## Zusammenfassung

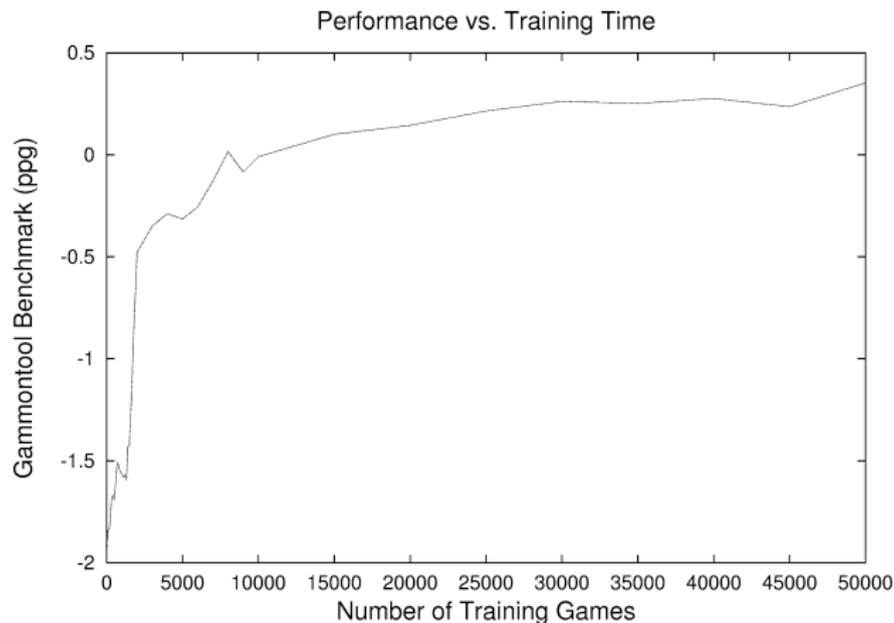


Abbildung: Lernkurve gegen Gammontool [Tesauro2002]

## Den Spielbaum durchsuchen

- ▶ Bis zu 3 Halbzüge Suchtiefe in Echtzeit („3-ply“)
  - ▶ Nicht in voller Breite
  - ▶ Forward Pruning des ersten Halbzugs gemäß 1-ply Vorhersage
  - ▶ Danach greedy Suche mit durch 1-ply Suche bestimmten Entscheidungen in allen  $21^2$  Würfelsequenzen
- ▶ Ergebnis ist der gewichtete Durchschnitt aller Blätter

## Annehmen

- ▶ Bei  $E[\text{Annahme}] \geq E[\text{Ablehnen}] = -\text{Einsatz}$
- ▶ In der Praxis weniger, da Besitz des Würfels die Equity erhöht

## Anbieten

- ▶ Gegner ist mathematisch gezwungen, anzunehmen (oder kurz davor)
- ▶ Einsatzserhöhung kompensiert kurzfristige Varianz

## Vorgehen

1. Varianz  $v$  („Volatilität“) und Bewertung  $\vec{x}$  des aktuellen Zustands mit  $n$ -ply Suche
2. Angepasste Heuristik von Zadeh–Kobliska als Funktion von  $v \rightarrow 3$ -D Entscheidungsebene  $E$
3. Auf welcher Seite von  $E$  liegt  $\vec{x}$ ?

## Verbesserungen

- ▶ Empirische Anpassungen, da zu aggressiv
- ▶ „Veto“-Entscheidungsebene

Einführung

Backgammon

Lösungsansätze

TD-Gammon

Steine ziehen

Den Einsatz verdoppeln

Evaluation

Spielstärke

Warum es so gut funktioniert

Zusammenfassung

# Gliederung

## Einführung

Backgammon

Lösungsansätze

## TD-Gammon

Steine ziehen

Den Einsatz verdoppeln

## Evaluation

Spielstärke

Warum es so gut funktioniert

## Zusammenfassung

Backgammon

Tobias Krönke

## Einführung

Backgammon

Lösungsansätze

## TD-Gammon

Steine ziehen

Den Einsatz verdoppeln

## Evaluation

Spielstärke

Warum es so gut funktioniert

## Zusammenfassung

## Computer-Wettkämpfe

- ▶ Große Samplesize möglich
- ▶ Keine Herausforderung für TD-Gammon
- ▶ Absolute Spielstärke?

## Gegen Menschen

- ▶ Kaum mehr als 100 Spiele
- ▶ Viel zu hohe Varianz bei ähnlicher Spielstärke

**Lösung:** Rollout-Analyse aller Halbzüge

**Ziel:** Equity-Verlust  $E[\text{gemachter Zug}] - E[\text{bester Zug}]$  in Points per Game (ppg)

## Besten Zug bestimmen

- ▶ Monte Carlo Simulation aller möglichen Halbzüge
- ▶ Mehrere 1000 Spielverläufe simulieren
  - ▶ z. B. bis Tiefe 11
  - ▶ Mit oder ohne Verdopplungswürfel
- ▶ Equity ist Mittelwert der Ergebnisse aller Simulationen

# Rollout-Ergebnisse

## Steine ziehen

Snowie Rollouts	$\emptyset$ Equity	$\emptyset$ F	$\emptyset$ GF
Bill Robertie	-0,188 ppg	2,21	0,47
TD-Gammon 2.1	-0,163 ppg	1,67	0,20

**Tabelle:** Spielserie 1993 [Tesauro2002]

Snowie Rollouts	$\emptyset$ Equity	$\emptyset$ F	$\emptyset$ GF
Malcolm Davis	-0,183 ppg	1,85	0,48
TD-Gammon 3.1	-0,050 ppg	0,59	0,04

**Tabelle:** AAI Hall of Champions 1998 [Tesauro2002]

## Einführung

Backgammon

Lösungsansätze

## TD-Gammon

Steine ziehen

Den Einsatz verdoppeln

## Evaluation

**Spielstärke**

Warum es so gut funktioniert

## Zusammenfassung

# Rollout-Ergebnisse

Verdoppeln anbieten, ablehnen / annehmen

## Malcolm Davis

- ▶  $-0.022$  ppg bis  $-0.031$  ppg

## TD-Gammon 3.1

- ▶  $-0.002$  ppg bis  $-0.020$  ppg
- ▶ Leichter Vorteil

# Was haben wir erreicht?

## Einführung

Backgammon  
Lösungsansätze

## TD-Gammon

Steine ziehen  
Den Einsatz verdoppeln

## Evaluation

Spielstärke  
Warum es so gut  
funktioniert

## Zusammenfassung

## Wissen und Nicht-Wissen

- ▶ Neurale Netze offensichtlich gut für Backgammon geeignet
- ▶  $w$  allein gibt aber keine neue Spielerkenntnis
  - ▶ Immerhin lassen sich Spielzüge analysieren
- ▶ Training: Masse statt Klasse?
- ▶ Reinforcement Learning mit TD( $\lambda$ ) die neue Wunderwaffe?

## Hill-climbing Ansatz aus [PollackBlair1997]

1. Initialisiere  $w = \vec{0}$
2. Herausforderer  $m = w +$  normalverteiltes Rauschen
3. Testspiele ( $< 10$ ) zwischen  $w$  und  $m$
4.  $w = 0.95 * \text{Sieger} + 0.05 * \text{Verlierer}$
5. Go to 2.

## Ergebnis

- ▶ Erstaunlich gute Spielstärke
- ▶ Die Wahrscheinlichkeit, einen besseren Herausforderer  $m$  zu sampeln, **steigt** mit der Stärke von  $w$

# Gliederung

## Einführung

Backgammon

Lösungsansätze

## TD-Gammon

Steine ziehen

Den Einsatz verdoppeln

## Evaluation

Spielstärke

Warum es so gut funktioniert

## Zusammenfassung

Backgammon

Tobias Krönke

### Einführung

Backgammon

Lösungsansätze

### TD-Gammon

Steine ziehen

Den Einsatz verdoppeln

### Evaluation

Spielstärke

Warum es so gut funktioniert

### Zusammenfassung

## Fazit

- ▶ TD-Gammon auf übermenschlichem Niveau
- ▶ Einige Spielansichten umgekrempelt
- ▶ Nahezu einzigartig erfolgreiche Anwendung von Reinforcement Learning

## Ausblick und Ziele

- ▶ Verdoppeln direkt lernen
- ▶ Den Erfolg bei anderen Problemen wiederholen

 Gerald Tesauro.  
Programming backgammon using self-teaching  
neural nets.

*Artificial Intelligence*, 134(1-2):181–199, 2002.

 Gerald Tesauro.  
Practical Issues in Temporal Difference Learning.  
*Machine Learning*, 1(1):257–277, 1992.

 Jordan B. Pollack und Alan D. Blair  
Why did TD-Gammon Work?.  
*Advances in Neural Information Processing Systems*,  
9(1):10–16, 1997.

 Richard S. Sutton und Andrew G. Barto.  
Reinforcement Learning: An Introduction (Adaptive  
Computation and Machine Learning).  
*The MIT Press*, 1998.