Fiktives Spiel und Verlustminimierung

Seminarvortrag von Alexander Marinc zur „TUD Computer Poker Challenge 2008“

Inhalt

- Einführung und allgemeine Übersicht
- Lösungsansatz des „Fiktiven Spiels“
- Lösungsansatz der „Verlustminimierung“
- Abschluss

Übersicht

Komplexität im Poker

- Poker (Texas Hold‘em) als Einstieg zur Betrachtung allgemeiner intelligenter Handlungsweisen in (komplexen) Spielsituationen
- Anzahl der Zustände im Spielbaum ist 10^{18}
- Reduktion durch Abstraktion auf 10^{7} bzw. 10^{12}
- Umso weniger Abstraktion umso besser die (pseudooptimale) Lösung

Zwei Ansätze

- Von mir behandelt wurden zwei Ansätze
  - Beide haben als Ziel die Berechnung eines spieltheoretischen Optimums (Nash Gleichgewicht)
- Der erste verwendet „Fiktives Spiel“
  - Durch Abstraktion (auf 10^7) und Vorberechnungen wird eine allgemeine Lösung bestimmt
- Die zweite verwendet „Verlustminimierung“
  - Engl.: Regret Minimization
  - Über Auflösung des Gesamtproblems in getrennt berechenbare Teile wird die Berechenbarkeit gewährleistet

Fiktives Spiel

Begriffsdefinitionen

- Nach dem Paper ist ein Nash Gleichgewicht eine Strategie die den Verlust eines Teilnehmers minimiert
- Es werden dominante (dF) und nicht dominante Fehler (ndF) in einer Strategie definiert
  - dF des Gegners führen langfristig zum Sieg
  - Bei ndF weicht zwar die Strategie von Optimum ab, aber führen langfristig nicht zwingend zu deren Schwächung
  - Beispiel: „Schere, Stein, Papier“
    - Beste Strategie: Alles zu 1/3 wählen.
    - Ein ndF wäre es immer Stein zu wählen.
    - Mit einer vierten Möglichkeit (z.B. Dynamit) die nur einmal gewinnt und zweimal verliert, wäre deren Wahl ein dF

Regeln des Fiktiven Spiels

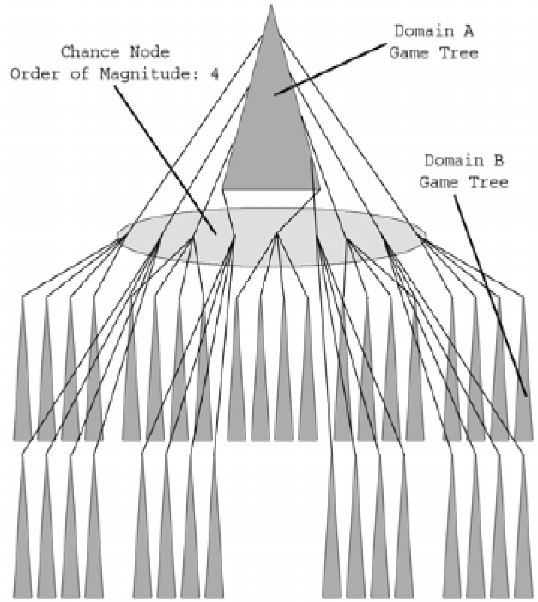
- Ein Fiktives Spiel ist im Grunde eine Menge von vier simplen Lernregeln:
  1. Jeder Spieler analysiert die Strategie des Gegners und erfindet eine beste Antwort
  2. Wurde eine beste Antwort berechnet, wird sie in die aktuelle Strategie eingesetzt (oder ersetzt diese)
  3. Jeder gegnerische Spieler führt ebenfalls Schritt 1 und 2 durch
  4. Die vorhergehenden Schritte werden wiederholt bis eine stabile Lösung erreicht wird

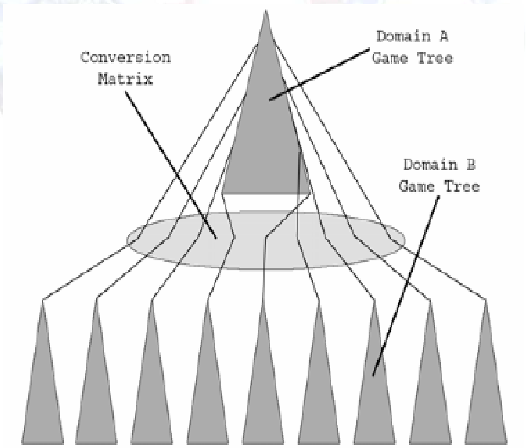
Abstraktion (1)

- Grundlegende Abstraktionen durch ignorieren -
  - der Reihenfolge der Karten (position isomorphs)
  - der Farbe der Karten (suit equivalenz isomorph)
- Weiterhin durch „Bucketing“
  - Kartensätze mit gleicher Gewinnwahrscheinlichkeit werden (in „Buckets“) gruppiert
  - Hier werden 169 solcher Buckets im Preflop und 265 für die folgenden Runden verwendet
  - Beispiel: "2h,4d,3c,5s,6s" und "2d,5c,4h,6d,3h„ (d=diamonds,h=heart,s=spades,c=cross)

Abstraktion (2)

- „Chance Node Elimination“
  - „Zufallsknoten“: Struktur aus den Bayesschen Netzwerken (Kombination aus Graphentheorie und Wahrscheinlichkeitsrechnung)
  - Stehen für Zufallsvariable mit einer Menge von sich gegenseitig ausschließenden Zuständen
  - Jeder Zustand hat eine Eintrittswahrscheinlichkeit
  - Die Zufallsknoten beim Übergang zwischen Domänen (z.B.: Preflop zu Flop) hat sehr viel Zustände
  - Ersetzung durch Konvertierungsmatrix





Abstraktion (3)

- Übergangswahrscheinlichkeiten
  - Konvertierungsmatrix, welche den Übergang einer Domäne (1) mit z.B. den Zuständen (A,B,C,D) in eine Domäne (2) mit den Zuständen (a,b,c,d) beschreibt

$$\begin{bmatrix} P(a) \\ P(b) \\ P(c) \\ P(d) \end{bmatrix} := \begin{bmatrix} P(a|A) & P(a|B) & P(a|C) & P(a|D) \\ P(b|A) & P(b|B) & P(b|C) & P(b|D) \\ P(c|A) & P(c|B) & P(c|C) & P(c|D) \\ P(d|A) & P(d|B) & P(d|C) & P(a|D) \end{bmatrix} \begin{bmatrix} P(A) \\ P(B) \\ P(C) \\ P(D) \end{bmatrix}$$

  - Erstellt werden die Matrizen durch „Masking Transition“
    - Zuerst werden allgemeine Übergangswahrscheinlichkeiten von Buckets in (1) zu Buckets in (2) festgelegt
    - Dann werden diese „maskiert“ mit Informationen über den Zustand von (2)

Training und Spiel

- Das Training erfolgt nach dem Fiktiven Spiel
  - Zwei Spieler die alles übereinander wissen
  - Erzeugung zufälliger Situationen und jeder Spieler wählt eine Lösung auf Basis des bisherigen Wissens
  - Die Lösung des Gegners wird mit in die allgemeine Lösung übernommen
- Im Spiel wählt der Algorithmus (Adam) verschiedene Ansätze:
  - Im Preflop: Wählt nur nach vorberechneter Lösung
  - Sonst: Nutzt die Vorberechnungen um den Spielbaum effizienter durchsuchen zu können

Verlustminimierung

Extensive Game (1)

- Kern eines „Umfangreiches Spiel“ ist der Spielbaum
  - Blätter bedeuten einen Gewinn/Verlust für jeden Spieler
  - Knoten sind mit einem Spieler der am Zug ist verbunden
- Ist das Spiel mit „unvollständigen Informationen“, kann ein Spieler Knoten im Baum (und somit ganze Pfade in diesem) nicht unterscheiden
- Die Formale Definition ist in diesem Rahmen zu umfangreich → Hier so weit möglich eine allgemeine Beschreibung

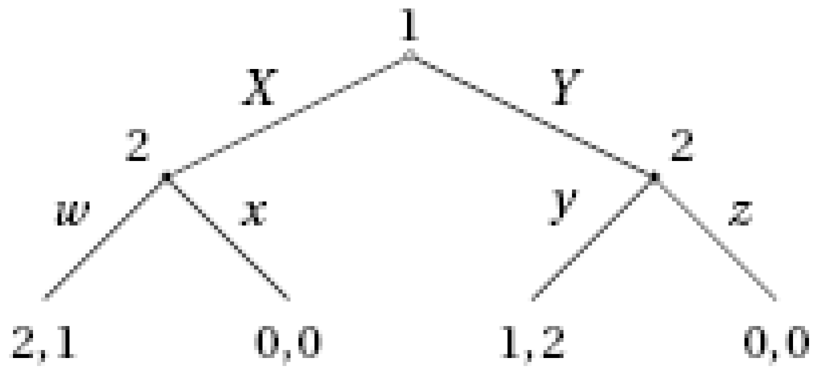
Extensive Game (2)

- Jedes Spiel hat eine Anzahl von N Spielern
- Betrachtet werden Sequenzen von Aktionen, *Historien* (*h*) genannt (Pfade durch den Spielbaum); *terminale Historien* enden in einem Blatt
- Funktionen geben die Menge der auswählbaren Aktionen (z.B. Check/Call), sowie den Spieler der nach einer *h* am Zug ist und den Gewinn an einem Blatt zurück
- Die *Informationspartition* (IP) eines Spielers besteht aus allen *h* nach den er am Zug ist, aufgeteilt in *Informationsmengen* (IM)
- Eine IM besteht aus allen h, welche der Spieler auf Grund seines unvollständigen Wissens nicht unterscheiden kann (die Menge auswählbarer Aktionen nach allen h einer IM ist gleich)

Strategien und Nash Gleichgewicht

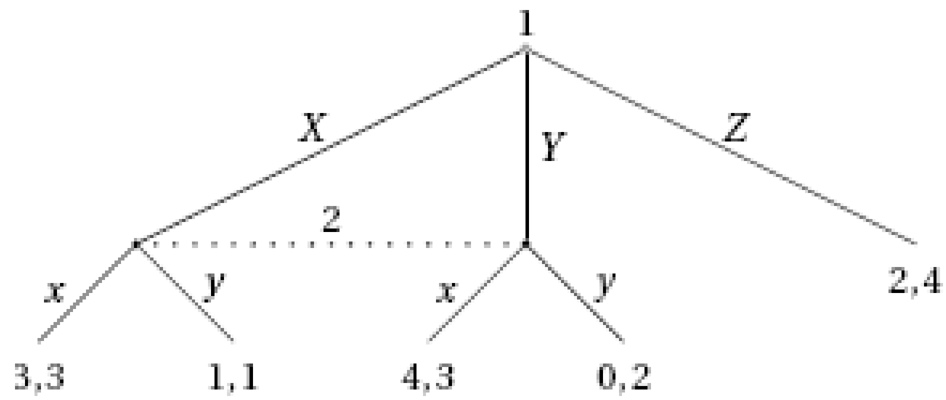
- Eine Strategie ist eine Funktion, welche für jede IM eines Spielers eine Aktion auswählt
- Ein Nash Gleichgewicht ist eine Menge von Strategien, bei der keiner der Spieler nur durch Wechsel seiner Strategie einen höheren Gewinn erzielt
- Ein Spiel kann keins, eins oder auch viele Nash Gleichgewichte beinhalten

„Extensive Game“: Beispiel 1



- Terminale h: {(X,w), (X, x), (Y, y), (Y, z)}
- Für die leere h gilt: Spieler 1 am Zug und die Menge der auswählbaren Aktionen = {X,Y}
- Strategien
  - Spieler 1: X oder Y
  - Spieler 2: wy, wz, xy, und xz
- Nash Gleichgewichte: {(X,wy), (X,wz), (Y, xy)}

„Extensive Game“: Beispiel 2



- Informationspartition für Spieler zwei ist {{X,Y},{Z}} mit {X,Y} und {Z} als IM und X,Y,Z als Historien
- Die auswählbaren Aktionen nach den Historien X und Y sind gleich ({x,y}).
- Für X und Y gelten demnach für Spieler 2 die gleichen Bedingungen → werden zusammen behandelt

Verlustminimierung Allgemein

$$R_i^T = \frac{1}{T} \boxed{max_{\sigma_i^* \in \sum_i}} \sum_{t=1}^{T} \left(\boxed{u_i(\sigma_i^*, \sigma_{-i}^t)} - \boxed{u_i(\sigma^t)} \right)$$

- Das „Gesamtbedauern" $R^T{}_i$ einer Spielers *i* wird mit obiger Formel berechnet (Runde 1…T)
- Rot = Gewinn von Spieler *i* im Nashgleichgewicht in Runde *t*
- Blau = Gewinn von *i*, wenn alle nach den Strategien des Nashgleichgewicht spielen und nur *i* eine andere Strategie wählt
- Grün = Diejenige der möglichen Strategien wird gewählt, welche den Term maximiert
- Für jede IM und jede Aktion von dieser kann eine „durchschnittliche Strategie" der Runden 1…T berechnet werden

→ **Satz:** Geht „das Bedauern" beider Spieler in einem Nullsummenspiel gegen Null wenn t gegen Unendlich geht, sind die gewählten Strategien ein Nashgleichgewicht

Kontrafaktischer Verlust (CV) (1)

- Dieser neue Ansatz berechnet das Bedauern über die einzelnen IM
- „Kontrafaktisch“ beschreibt Betrachtungen „wie etwas geworden wäre, wenn man etwas anders gemacht hätte“
- Der CV beschreibt das Bedauern eines Spielers *i* nach dem erreichen einer IM, wenn er von Anfang an gespielt hätte um diese zu erreichen
- Der als „immediate counterfactual regret“ (ICV) bezeichnete Verlust pro IM definiert sich folgender Maßen:

$$R^T_{i,imm}(I) = \frac{1}{T} \boxed{max_{a \in A(I)}} \sum_{t=1}^{T} \boxed{\pi^{\sigma^t}_{-i}(I)} \left(\boxed{u_i(\sigma^t|_{I \to a}, I)} - \boxed{u_i(\sigma^t, I)} \right)$$

Kontrafaktischer Verlust (CV) (2)

- Für die Menge aller ICV gilt laut Paper:

$$R_i^T \leq \sum_{I \in \mathrm{I}_i} R^T_{t,imm}(I)$$

- Die Summe aller ICV stellt also eine obere Schranke für das Gesamtbedauerns dar
- Für alle IM und alle nach diesen auswählbaren Aktionen wird der CV berechnet mit:

$$R_i^T(I,a) = \frac{1}{T} \sum_{t=1}^{T} \pi^{\sigma^t}_{-i}(I)\left(u_i(\sigma^t|_{I \to a}, I) - u_i(\sigma^t, I)\right)$$

Kontrafaktischer Verlust (CV) (3)

- Mit diesen Informationen lässt sich folgende Strategie ableiten:

$$\sigma_i^{T+1}(I)(a) = \begin{cases} \frac{R_i^{T,+}(I,a)}{\sum_{a \in A(I)} R_i^{T,+}(I,a)}, \text{ wenn } \sum_{a \in A(I)} R_i^{T,+}(I,a) > 0 \\ \frac{1}{|A(I)|} \text{ sonst} \end{cases}$$

- In Worten: Die Wahl einer Aktion ist proportional zum positiven CV den man hinnimmt, wenn man diese Aktion nicht wählt
- Gib es nur negative CV bestimmt der Zufall die nächste Aktion

Kontrafaktischer Verlust (CV) (4)

- Mit dieser Strategie sind alle ICV wie folgt nach oben beschränkt: $R_{i,imm}^{T}(I) \leq \Delta_{u,i}\sqrt{|A_i|}/\sqrt{T}$
- Mit $|A_i|$ = max. Anzahl auswählbarer Aktionen nach allen IM, und $\Delta_{u,i}$ = max. Differenz zwischen möglichen Gewinnen und Verlusten von Spieler i
- Mit $R_i^T \leq \sum_{I \in \mathrm{I}_i} R_{i,imm}^{T}(I)$ wird deutlich, dass so auch das Gesamtbedauern beschränkt wird

➔ Durch Minimierung der ICV kann auch das Gesamtbedauern minimiert werden

Anwendung: Abstraktion

- Karten werden nach ihrer „Quadrierten Handstärke“ (QHS) in Buckets einsortiert
- Handstärke = Gewinnwahrscheinlichkeit nur nach den Karten welche der Spieler gesehen hat
- Sequenzen (Historien) der 1. Runde werden nach ihrer QHS auf zehn Buckets verteilt
- In Runde 2 werden alle Sequenzen die einen Bucket der 1. teilen nach der QHS in einen von 10 neuen Buckets einsortiert
- Dies wird für jede neue Runde wiederholt, woraus sich *Bucketsequenzen ergeben (also im Grunde eine IM nicht mehr unterscheidbarer Historien)*

Anwendung: Strategieentwicklung

- Zwei Spieler spielen immer wieder gegeneinander mit der entwickelten Strategie
- Alle $R^t_i(I,a)$ müssen in jeder Runde gespeichert und danach aktualisiert werden
- Verbesserung durch Ignorieren des „Zufallspielers“ möglich
- Nach einer bestimmten Anzahl Iterationen werden die durchschnittlichen Strategien beider Spieler als Lösung akzeptiert

Abschluss

- Wir haben zwei Methoden kennen gelernt
  - Eine welche durch „Fiktives Spiel“ Übergangsmatrizen zwischen verschiedenen Domänen adaptiert
  - Eine andere die durch Aufteilung von „Bedauern“ auf Informationsmengen effektiv ein Nashgleichgewicht berechnen kann
- Es gibt Gemeinsamkeiten zwischen den Ansätzen
  - Beide müssen das Originalspiel abstrahieren
  - Beide verwenden eine Art von Bucketing
  - Auch wenn dies im „Fiktiven Spiel“ nicht explizit angegeben wird, arbeiten bei auf der Theorie des „Extensive Games“

Quellen

- Duziak, W. (2004). Using ctitious play to nd pseudo-optimal solutions for full-scale poker. In Proceedings of the 2006 International Conference on Articial Intelligence (ICAI-2006)
- M. Zinkevich, M. Bowling, M. J., & Piccione, C. (2006). Regret minimization in games with incomplete information. Advances in Neural Information Processing Systems (NIPS 2007), Seiten 374-380
- Osborne, M. J. (2006). Strategic and extensive games. Department of Economics, University of Toronto
- Stockhammer, P. (2006). Einführung in Bayessche Netzwerke. Hauptseminar Wirtschaftsinformatik- TU Claustal

Ende

Besten Dank für ihre Aufmerksamkeit!