Fiktives Spiel und Verlustminimierung zur Berechung optimaler Lösungen in der Pokervariante Texas Hold'em

Alexander Marinc alexander.marinc@gmx.de

Abstract

Diese Ausarbeitung gibt zentral zwei verschiedene Ausarbeitungen wieder, die versuchen möglichst genaue Nash Gleichgewichte in einem Poker Spiel (Texas Hold'em) zu berechnen. Der erste Ansatz trägt den Titel "Using Fictitious Play to Find Pseudo-Optimal Solutions for Full-Scale Poker" (Duziak, 2006). Die Autoren versuchen pseudooptimale Lösungen durch eine geschickte Art der Abstraktion des Spieles und anschließendem "Training" allgemeiner Konvertierungsmatrizen zu finden. Der zweite der beiden Ansätze mit dem Titel "Regret Minimization in Games with Incomplete Information" (M. Zinkevich & Piccione, 2007) beschreibt einen auf Verlustminimierung basierenden Gedanken, welcher von den Autoren als "Kontrafaktischer Verlust" bezeichnet wird. Es werden die Grundlagen beider Ausarbeitungen dargestellt und ihre Verwendung im Rahmen des Spiels Poker beschrieben.

1. Einleitung

Poker ist ein Spiel, welches in der letzten Zeit immer mehr Beachtung gefunden hat. Parallen zu "realen" Konkurenzsituationen, wie zum Beispiel in der Wirtschaft zwischen zwei Firmen, und eine dennoch annähernd berechenbare Komplexität von 10^{18} Spielzuständen, machen das Spiel auch für theoretische Betrachtungen im Bereich der künstlichen Intelligenz interessant. Zwei Ansätze spieltheoretische Optima für das Spiel Poker zu berechnen werden im folgenden dargestellt. Der Ansatz von (Duziak, 2006) reduziert die Komplexität des Spiels auf 10^7 und liefert laut den Autoren dennoch gute Ergebnisse auf dieser "pseudooptimalen Lösung". Der im Original als "Fictitious Play" bezeichnet englische Ausdruck wird im Rahmen dieser Arbeit als "Fiktives Spiel" übersetzt. Allgemein läßt sich feststellen, dass die Gewinnchancen eines Algorithmus verbessert werden, umso weniger er das originale Spiel abstrahiert und umso höher die berechenbare Komplexität wird. Der zweite in dieser Arbeit beschriebene Ansatz von (M. Zinkevich & Piccione, 2007), macht in seiner Gesamtheit einen wesentlich komplexeren Eindruck und verwendet eine viel formalere und mathematischere Ausdrucksweise. Entsprechend den Angaben im Dokument ist mit den hier verwendeten Techniken zur Verlustminimierung (im englischen als "Regret Minimization" bezeichnet) und Abstaktion lediglich eine Reduktion auf 10^{12} Zustände nötig. Zu beiden Ansätzen liegen Ergebnisse im Vergleich zu bekannten Pokeralgorithmen vor, jedoch leider kein Vergleich untereinander. Der Vergleich zwischen beiden Algorithmen wird sich demnach im Wesendlichen auf Inhaltliche Aspekte beziehen. Der Fokus liegt aber ohne hin auf der Beschreibung der beiden Algorithmen. Hierzu gehört die Erörterung der verwendeten Begriffe im Zusammenhang des jeweiligen Ansatzes, sowie die theoretischen Hintergründe (sowei angegeben) und zuletzt die konkrete Umsetzung für das Spiel Poker. Allgemeine Kenntnisse der Regeln und Abläufe im Texas Hold'em Poker werden

vorausgesetzt, sind aber zum Beispiel auch in der Einleitung des Textes von (Duziak, 2006) beschrieben.

2. Fiktives Spiel

2.1 Grundlagen

Um den Gedanken des Fiktiven Spiels erklären zu können, werden einige allgemeine Einführungen der Spieltheorie, welche von den Autoren aufgeführt werden, benötigt. Manche hier erörterten Begriffe werden in der Beschreibung des zweiten Algorithmus erneut dargestellt. Da sich jedoch die Art und die Komplexität der Darstellung in dem Paper über das Fiktive Spiel leichter verständlich und weniger präzise darstellt, ist der nachfolgende Teil als Einführung in die behandelten Problematiken besser geeignet und eine gewissen Redundanz durchaus hilfreich für das Verständnis.

2.1.1 Nash Gleichgewicht und dominierende/nicht dominierende Fehler

Eine optimale Lösung, oder auch Nash Gleichgewicht, eines gegebenen Spiels ist nach den Autoren des Textes ein aus intelligentem Verhalten abgeleitete Strategie, welche den Verlust eines Teilnehmers minimiert. Ein Spieler welcher strategisch dominante Fehler begeht, wird langfristig gegen eine optimale Strategie verlieren. Der einzige Nachteil hierbei ist die Voraussetzung, dass der Gegner Fehler machen muss, damit die optimale Strategie zum Erfolg führt. Anhand des Beispiels "Schere, Stein, Papier" werden die drei Begriffe Nash Gleichgewicht, dominierender und nicht dominierender Fehler verdeutlicht. Die optimale Strategie in diesem Spiel ist es (über die Summe der Spiele) jede der drei Möglichkeiten zu einem Drittel zu spielen. Immer nur zum Beispiel Stein zu spielen, wäre keine optimale Lösung, aber es besteht immer noch zu je einem Drittel die Möglichkeit zu verlieren, gewinnen und Gleichstand zu spielen. Der Fehler in der Strategie ist daher ein "nicht dominierender". Wenn man jetzt noch eine vierte Auswahlmöglichkeit hinzunehmen würde die nur einmal gewinnen und zweimal verlieren kann, wäre es ein dominierender Fehler diese zu wählen, da in der optimalen Strategie diese in null Prozent der Fälle zu wählen ist. In komplizierteren Spielen treten solche Fehler ausreichend oft auf, um zu deutlichen Schwächen in der Strategie führen zu können.

2.1.2 Definition eines Fiktives Spiels

Frei übersetzt aus Quelle (Duziak, 2006) ist ein "Fiktives Spiel" eine Menge von Lernregeln, entworfen damit Teilnehmer (eines Spiels) befähigt werden ihr Handeln einem Optimum anzunähern. In einem fiktiven Spiel gelten folgende vier Regeln:

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

Laut dem Autor des Artikels kann nicht immer eine stabile Lösung berechnet werden.

2.2 Abstraktionen für das Fiktive Spiel im Poker

Wie bereits einleitend erwähnt, wird die Komplexität von Texas Hold'em in diesem Ansatz von 10^{18} auf 10^7 reduziert. Jedoch werden die Schlüsseleigenschaften des Spiels durch eine angemessene Abstraktion erhalten. Lösungen auf dieser Abstraktion werden pseudooptimal genannt. Zwei grundlegenden Techniken sind hierbei das *position isomorph* und das *suit equivalenz isomorph*. Erstere besagt, dass die Reihenfolge der Karten in der Hand und im Flop keine Rolle spielt, und die Zweite ignoriert die Farben der Karten. Beide Techniken haben die Eigenschaft die Optimalität einer Lösung auf der Abstraktion nicht zu beeinflussen. Um jedoch das Problem für eine Berechnung genügend zu reduzieren, sind weitere Schritte notwendig. Eine Möglichkeit hierfür ist das so genannte *Bucketing*. Hierbei wird versucht Kartensätze mit der gleichen Gewinnwahrscheinlichkeit zu gruppieren. In dem vorliegenden Ansatz werden 169 solcher Buckets im Preflop und 256 in den folgenden Runden verwendet. Beispielsweise kommen die Hände "2h,4d,3c,5s,6s" und "2d,5c,4h,6d,3h" (d=diamonds,h=heart,s=spades,c=cross) in die gleichen Buckets, da sie sich nur in Farbe und Reihenfolge unterscheiden.
Weiterhin wird eine Technik namens *Chance Node Elimination* verwendet. Für FS benötigt man einen genau definierten und (nach guten Lösungen) durchsuchbaren Spielbaum. Allerdings hat jede der vier Runden im Poker (Preflop, Flop, Turn und River) zwischen zwei Spielern zehn Zustände, welche bedingt durch die Vielzahl der möglichen Karten der Spieler und im Flop, jeweils zu einer vollkommen anderen Nachfolgestruktur führen. Der zugehörige Baum zu jeder Runde wird als Domäne bezeichnet. Der Übergang zwischen diesen Domänen durch Check oder Call kann durch so genannte Zufallsknoten (engl.: Chance Node) beschrieben werden. Ein Zufallsknoten ist eine Struktur aus den Bayesschen Netzwerken (Stockhammer, 2006), welche Verwendung findet in Einflussdiagrammen. Bayesschen Netzwerke kombinieren Graphentheorie und Wahrscheinlichkeitsrechnung. Ein Zufallsknoten repräsentiert hierbei eine Zufallsvariable, welche eine bestimmte, sich gegenseitig ausschließende Anzahl von Zuständen annehmen kann, wobei jeder Zustand eine gewisse Eintrittswahrscheinlichkeit hat. Die hohe Menge der möglichen Zustände sorgt für ein schnelles, exponentielles Wachstum. Die Lösung in diesem Ansatz besteht daher darin diese Zustandsknoten zu eliminieren und durch Konvertierungsmatrizen zu ersetzen, welche die gleiche Funktion erfüllen, aber das exponentielle Wachstum einschränken. Jeder Knoten einer Domäne hat auf diese Weise nur noch einen Domänenunterbaum. Als Möglichkeit eine Konvertierungsmatrix darzustellen, wir im Text folgendes Beispiel angegeben:

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

Ausgedrückt wird eine *Übergangswahrscheinlichkeit* (engl.: Transition Probabilitie) von einer Domäne 1 mit den möglichen Zuständen $\{A, B, C, D\}$ zu einer Domäne 2 mit den möglichen Zuständen $\{a, b, c, d\}$. Die Wahrscheinlichkeit, dass das Spiel im Zustand a in Domäne 2 angelangt, entspricht demnach der Summe der konjugierten Wahrscheinlichkeiten

zwischen a und allen Zuständen der Domäne 1 (da P(a|b)*P(b)=P(a,b)). Im Grunde wird der Übergang eines Bucket in Domäne 1 zu dem passenden in Domäne 2 beschrieben, oder anders gesagt von Kartenblättern gleicher Gewinnwahrscheinlichkeit einer Runde zu denen der nächsten. Für das (aufwendige) Berechnen dieser Matrizen werden zwei Beispiele angegeben. Beim *Abdeckenden Übergang* (engl.: Masking Transition) werden zunächst allgemeine Übergangswahrscheinlichkeiten von Domäne zu Domäne erstellt und dann mit speziellen Informationen über den Zustand der Zieldomäne überdeckt. Der nächste Ansatz des *Perfekten Übergangs* (engl.: Perfect Transition) hingegen verwendet Vorberechnungen zu jedem möglichen Spielzustand, welche während der Laufzeit verwendet werden können um perfekt angepasste Umwandlungsmatrizen, welche den Zustand von Domäne 1 und 2 mit einbeziehen, zu generieren. Auf Grund der vielen Spielzustände und der einfacheren Berechnung und Speicherung der Buckets, fiel die Auswahl auf die erste Möglichkeit.

2.3 Verwendung des "Fiktiven Spiel" Ansatzes

Die als *Adam* bezeichnete Implementation des Spiels, wurde nach den Prinzipien des FS trainiert (Anpassung einer allgemeinen Lösung). Dies erfolgte durch zwei Spieler, welche alles übereinander wissen. Für zufällige Spielsituationen wird mit dem Wissen beider Spieler die "korrekte" Handlungsweise bestimmt und eine allgemeine Lösung entsprechende angepasst. Für jeden Knoten des Entscheidungsbaumes wurde dieser Vorgang hunderttausende Mal angewendet, bis eine stabile (optimale) Lösung gefunden wurde in der dominante Fehler weitgehend Beseitigt sind und die ein annährendes Nashgleichgewicht darstellt. Im "realen" Spiel muss Adam auf verschiede Wege Ansätze zur Lösungsbestimmung finden. Im Preflop (welcher durch die Abstraktion unverändert bleibt) kann sich Adam noch vollkommen auf die vorberechnete Lösungen verlassen. Erst danach können diese mehr und mehr nur als Referenz verwendet werden um den Spielbaum effizienter nach der besten Lösung zu durchsuchen. Hierbei wird jeweils die Aktion gewählt, welche zu dem Unterbaum mit dem maximalen Gewinnwert führt.

3. Verlustminimierung

Der zweite hier behandelte Artikel (M. Zinkevich & Piccione, 2007) beschäftigt sich mit der *Verlustminimierung* (frei aus dem engl.: regret minimization) in umfangreichen Spielen (engl.: extensive games) mit unvollständigen Informationen. Um die Inhalte dieses Textes besser vermitteln zu können, folgen auch in diesem Abschnitt einige allgemeine Erörterungen. Interessant sind diese auch in Bezug auf das "Fiktive Spiel", da sich mit der im Anschluss verwendeten Terminologie auch diese Methode genauer spezifizieren lassen würde. Im Anschluss wird das Prinzip des "Kontrafaktischen Verlustes" und dessen Anwendung im Poker dargestellt.

3.1 Grundidee und Grundlagen

Wie der Name schon sagt, geht es darum den Verlust eines Spielers möglichst auf Null zu reduzieren, wobei es zuerst einmal nicht explizit um das Spiel Poker geht. Zusätzlich wird aber auch gezeigt, wie sich diese Minimierung nutzen läßt um ein Nashgleichgewicht für ein Spiel zu berechnen. Einige der im Folgenden genannten Beschreibungen zur allgemeinen

Spieltheorie wurden ergänzt mit der Hilfe eines weiteren Artikels (Osborne, 2006). Der Begriff des "Extensive Game" (EX) kommt aus der allgemeinen Spieltheorie. Der Kern hierbei ist zunächst einmal ein perfekter Spielbaum. Jeder Endzustand dieses Baumes (Blätter) ist mit einem bestimmten Gewinn/Verlust für alle Spieler assoziiert und jeder andere mit einem Spieler, welcher eine Aktion wählen muss. Der Zusatz "mit *unvollständigen Informationen*" gibt an, dass jedem Spieler nicht der gesamte (bisherige) Spielbaum während eines Spiels bekannt ist. Bestimmte Zustände kann er nicht voneinander unterscheiden und liegen daher in einer Informationsmenge (weil er z.B. im Poker die Karten der anderen Spieler nicht kennt). Genau definiert besteht ein "extensive game" aus mehreren Strukturen, welche in anschließender Liste beschrieben und in ihrer formalen Schreibweise wiedergegeben werden.

- Es gibt eine endliche Anzahl von Spielern, welche durch N ausgedrückt wird.
- Eine endlichen Menge H von Sequenzen, welche die möglichen *Historien* von Aktionen darstellen. Historien stellen Pfade durch den Spielbaum dar, welche leer sein können (Pfad "zum Root"), zu einem beliebigen Knoten im Baum verlaufen können oder bis zu einem terminalen Knoten (Blatt) verlaufen. Letztere bezeichnet man als terminale Historien und werden durch die (echte) Untermenge Z von H angegeben.
- Die Funktion $A(h)$ gibt die Menge der auswählbaren Aktionen nach (z.B. Call und Check) einer nicht terminale Historie $h \in H$ $(H \backslash Z)$ zurück.
- Die Funktion $P(h)$ weißt jeder Historie $h \in H \backslash Z$ einen Spieler aus N zu, welcher eine mögliche Aktion (aus $A(h)$) auswählen muss. Die Menge der Spieler wird allerdings um einen Spieler c erweitert. Ist $P(h) = c$ wird eine zufällige Aktion ausgewählt (z.B. wenn eine neue zufällige Karte in den Flop kommt).
- Die Funktion f_c assoziiert mit jeder Historie h mit $P(h) = c$ (der Zufall bestimmt die nächste Aktion) eine Wahrscheinlichkeitsmessung $f_c(\bullet|h)$ auf $A(h)$. Die Funktion f gibt also für jede nach einer Historie auswählbare Aktion die Wahrscheinlichkeit an, dass sie auch tatsächlich gewählt wird (also z.B. wie hoch die Wahrscheinlichkeit ist, dass ein Bube gewählt wird).
- Die Untergliederung (Informationspartition) I_i eines Spielers i beinhaltet alle Historien h in welchen $P(h) = i$ gilt. Alle Historien h' mit den gleichen auswählbaren Aktionen $(A(h) = A(h'))$ müssen hierbei in der gleichen Teilmenge der Informationspartition stehen. Diese Teilmengen werden mit $I_i \in \mathrm{I}_i$ dargestellt und als eine Informationsmenge von Spieler i bezeichnet. $A(I_i)$ gibt die Menge der auswählbaren Aktionen nach den Historien in I_i an und $P(I_i)$ den Spieler. Verdeutlichend beschrieben umfasst eine Informationsmenge die Historien, welche für einen Spieler auf Grund seines unvollständigen Wissens im Spielbaum nicht zu unterscheiden sind.
- Zuletzt gibt es für jeden Spieler einer Funktion u_i welche die terminalen Zuständen in Z auf relle Werte abbildet. Gleichen sich die Funktionswerte aller Spieler an jeweils allen terminalen Knoten auf Null aus, wird das Spiel als ein "zero-sum extensive Game" bezeichnet. Die maximale Spanne zwischen kleinstem und größten Wert der terminalen Knoten für Spieler i wird mit $\Delta_{u,i}$ dargestellt.

Bezeichnung	Beschreibung
σ	Beschreibt ein vollständiges Strategieprofil, also eine feste Strategie für jeden Spieler
σ_{-i}	Alle Strategien in σ außer σ_i
$\pi^{\sigma}(h)$	Wahrscheinlichkeit, dass die Historie h erscheint, wenn die Spieler entsprechend dem Profil σ ihre Aktionen wählen
$\pi_i^{\sigma}(h)$	Wahrscheinlichkeit, dass wenn Spieler i nach σ spielt, er in Historie h die entsprechende Aktion wählt wie in den Präfixen h' von h an denen er auch am Zug war ($P(h') = i$), wobei $\pi^{\sigma}(h) = \prod_{i \in N \cup \{c\}} \pi_i^{\sigma}(h)$ gilt (Produkt aller Einzelbeiträge der Spieler)
$\pi_{-i}^{\sigma}(h)$	Produkt der Beiträge aller Spieler, außer dem von i
$\pi^{\sigma}(I)$	Ist die Wahrscheinlichkeit, dass wenn alle Spieler nach dem Profil σ spielen, Informationsmenge I erreicht wird. Formal ausgedrückt gilt $\pi^{\sigma}(I) = \sum_{h \in I} \pi^{\sigma}(h)$ (Summe der Wahrscheinlichkeiten aller Historien in I)
$u_i(\sigma)$	Erwartete Auszahlung für Spieler i bei Profil σ in dem daraus resultierenden Endknoten, formal dargestellt durch $u_i(\sigma) = \sum_{h \in Z} u_i(h)\pi^{\sigma}(h)$ (Summe aller Gewinnwerte für Spieler i im Spielbaum, gewichtet mit der Wahrscheinlichkeit, dass die jeweilige terminale Historie in σ auftritt)

Table 1: Begriffsdefinitionen für Strategien in "Extensive Games"

3.1.1 Strategien und Nash Gleichgewicht

Eine Strategie σ_i ist eine Funktion, welche Aktionen aus $A(I_i)$ auswählt für alle Informationsmengen des Spielers i. Zur besseren Übersicht erfolgt die Definition über die weiteren verwendeten Zeichen im Zusammenhang mit Strategien in Tabelle 1.

Ein Nash Gleichgewicht wird hier entsprechend der in Tabelle 1 beschriebenen formalen Ausdrücke dargestellt. Bei zwei Spielern bedeutet dies:

$$u_1(\sigma) = max_{\sigma_1' \in \sum_1} u_1(\sigma_1', \sigma_2) \qquad \text{und} \qquad u_2(\sigma) = max_{\sigma_2' \in \sum_2} u_1(\sigma_1, \sigma_2')$$

Ein Nash Gleichgewicht ist demnach ein Profil, in dem jeder Spieler bei gegebener Strategie der Gegner (hier nur einer) von seinen möglichen Strategien ($\sum_i$) die Auswählt, welche seinen Gewinn maximiert. Keiner kann demnach mehr eine andere Strategie wählen, ohne seinen Gewinn zu schmälern. Interessant in diesem Zusammenhang ist die Eigenschaft von "Extensive Games", dass diese nicht zwingend ein Nash Gleichgewicht besitzen müssen oder sogar viele Existieren (für Beispiele siehe auch (Osborne, 2006)). Ein Nash Gleichgewicht bei dem das Maximum nicht vollständig erreicht wird, nennt man auch ϵ-Nash Gleichgewicht.

Zum besseren Verständniss werden die definierten Begriffe zunächst einmal durch das in Abbildung 1 auf der linken Seite dargestellte Beispiel aus (Osborne, 2006) eines "Extensive Games" mit vollständigen Informationen verdeutlicht. Das Spiel hat die terminalen Historien $Z = \{(X, w,), (X, x), (Y, y), (Y, z)\}$. Mit h als leere Historie ist $P(h) = 1$ (also Spieler 1) am Zug) und $A(h) = \{X, Y\}$. Wenn $h = \{X\}$ oder $h = \{Y\}$ ist $P(h) = 2$. Spieler 1 hat in diesem Beispiel nur zwei Strategien, und zwar X und Y. Spieler 2 hat vier

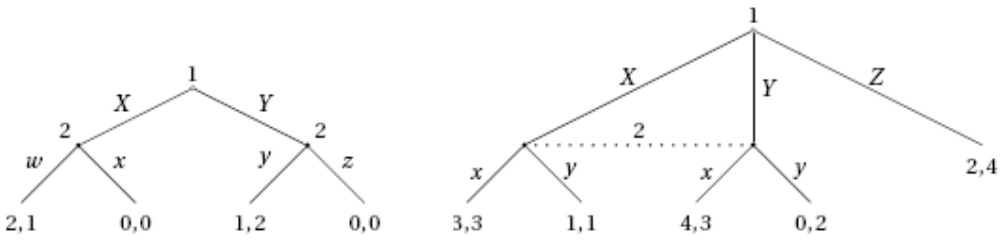

Figure 1: Beispiele für "Extensive Games" aus (Osborne, 2006)

Strategien und zwar wy, wz, xy, und xz, wobei zuerst immer die Wahl von Spieler 2 steht, wenn Spieler 1 X gewählt hat und dann die wenn Spieler 1 Y gewählt hat. Eine Strategie ist demnach immer eine vollständige Beschreibung der Handlungsweise in jedem Spielzustand. Nashgleichgewichte in diesem Spiel sind (X, wy), (X, wz), und (Y, xy), da in jedem Fall der gewählten Strategien kein Spieler eine andere Strategie wählen kann um seinen Gewinn zu vergrößern (z.B. im letzten Fall, wenn Spieler zwei sich schon für xy entschieden hat, würde Spieler 1 durch Wahl von X nicht erhalten). Es ist festzustellen, dass unter Berücksichtigung der Abhängigkeit der tatsächlichen Entscheidung von Spieler 1, nur das erste Nashgleichgewicht optimal ist. In der weiteren Betrachtung spielt dies jedoch keine Rolle. Auf der rechten Seite von Abbildung 1 sehen wir ein Beispiel für ein partitioniertes "Extensive Game" mit unvollständigen Informationen. Die Historien zu den durch eine Linie verbundenen Zuständen sind hierbei nicht unterscheidbar für Spieler 2. Sein Informationspartition ist demnach $\{\{X, Y\}, \{Z\}\}$, mit den zwei Informationsmengen $\{X, Y\}$ und $\{Z\}$ und X,Y,Z als Historien. Besonders zu beachten ist das $A(X) = A(Y) = \{x, y\}$ gilt wie vorgeschrieben. Strategien müssen jetzt auf Basis der Informationsmengen definiert werden, da die Wahl einer Aktion nach allen zugehörigen Historien gleich ist (aus Sicht des Spielers).

3.1.2 Verlustminimierung

Das Prinzip der Verlustminimierung ist ein allgemein bekanntes Spielkonzept, welches auf Lernen beruht. Betrachtet wird das wiederholte Spielen eines "Extensive Games", wobei σ_i^t die von Spieler i in Runde t verwendete Strategie darstellt. Der Gesamtverlust eines Spielers über alle Runden ergibt sich aus der Strategie, die im Schnitt über alle Runden die Differenz zum Nachgleichgewicht des Strategieprofils der Runden möglichst positiv werden läßt. Formal nach dem Paper bedeutet dies:

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

Mit $\bar{\sigma}_i^T$ als durchschnittliche Strategie bis zum Zeitpunkt T für Spieler i wird für jede Informationsmenge des Spielers und jede auswählbare Aktion von dieser folgendes berechnet:

$$\bar{\sigma}_i^T(I)(a) = \frac{\sum_{t=1}^{T} \pi_i^{\sigma^t}(I)\sigma^t(I)(a)}{\sum_{t=1}^{T} \pi_i^{\sigma^t}(I)}$$

Bezeichnung	Beschreibung
$u_i(\sigma, h)$	Erwarteter "Nutzen" für Spieler i nachdem Historie h erreicht wurde und alle Spieler die Strategie entsprechend von σ verwendet haben
$u(\sigma, I)$	Wird von den Autoren als "Kontrafaktischer Nutzen" (engl.: Counterfactual utility) bezeichnet. Es steht für den Nutzen welcher zu erwarten ist nachdem I erreicht wurde, wenn alle Spieler Strategie σ verwenden und Spieler i nicht gespielt hat um I zu erreichen.
$\sigma\|_{I \to a}$	Definiert für alle $a \in A(I)$ ein zu σ identisches Strategieprofil, außer dass Spieler i immer Aktion a wählt, wenn er in der Informationsmenge I ist.

Table 2: Begriffsdefinitionen für "Counterfactual Regret"

Also die durchschnittliche Wahrscheinlichkeit, dass a in I gewählt wird (wenn I überhaupt erreicht wird). Ein Algorithmus, welcher den Verlust eines Spielers möglichst minimiert, muss in jeder Runde t auf die Art eine Strategie σ_i^t für Spieler i wählen (unabhängig von der Strategien der anderen Spieler), dass R_i^T gegen Null geht, wenn t gegen Unendlich geht. Da der Verlust über die Differenz des Gewinns einer Runde zum Gewinn im Nash Gleichgewicht der Runde definiert ist und in einem Null-Summen Spiel sich Gewinn und Verluste auf Null ausgleichen, kann ein solcher Algorithmus auch direkt verwendet werden um ein gesamtes Nashgleichgewicht anzunähern (also eine "durchschnittliche Strategie" auf Basis des Wissens eines Spielers, welche ein Nashgleichgewicht annähert). In einem Nullsummenspiel ist $\bar{\sigma}^T$ (also das Strategieprofil zum Zeitpunkt T) demnach ein Nashgleichgewicht, wenn der Gesamtverlust beider Spieler kleiner ϵ ist (die Strategie ist dann ein 2ϵ-*Nash Gleichgewicht*)

3.2 Kontrafaktischer Verlust

Der von den Autoren das Papers (M. Zinkevich & Piccione, 2007) definierte Begriff des *Kontrafaktisches Verlustes* (engl.: Counterfactual Regret), beschreibt die Aufteilung des Gesamtverlustes auf einzelne Informationsmengen, welche unabhängig von einander minimiert werden können. Die Summe der Verluste der einzelnen Informationsmengen stellte eine Schranke für den Gesamtverlust dar. Im Folgenden werden die in Tabelle 2 definierten Begriffe verwendet.

Der Begriff "Kontrafaktisch" beschreibt den Umstand, dass man etwas betrachtet wie es hätte sein können. Kontrafaktisches Denken beschreibt somit zum Beispiel einen Gedankengang wie in etwa "Wenn ich damals Lotto gespielt hätte, wäre ich heute Millionär". Tatsächlich beschreibt der von den Autoren beschriebene *immediate counterfactual regret* (Abk.: ICR) den Verlust welcher ein Spieler in einer bestimmten Informationsmenge gehabt hätte, wenn er von Anfang an versucht hätte diesen zu erreichen. Formal dargestellt mit den in Tabelle 2 beschriebenen Mitteln sieht dies dann folgender Maßen aus:

$$R_{i,imm}^T(I) = \frac{1}{T} max_{a \in A(I)} \sum_{t=1}^{T} \pi_{-i}^{\sigma^t}(I)(u_i(\sigma^t|_{I \to a}, I) - u_i(\sigma^t, I))$$

Der Verlust nach einer Informationsmenge I definiert sich also über die Aktion in $A(I)$, welche optimal zu wählen gewesen wäre in den vorherigen Runden, gewichtet mit der Wahrscheinlichkeit, dass I auch ohne das Wirken von Spieler i anhand der Strategien

der anderen Spieler, erreicht worden wäre. Laut den Autoren ist der tatsächliche Verlust eines Spielers R_i^T immer kleiner gleich $\sum_{I \in \mathrm{I}_i} R_{t,imm}^T(I)$, also kleiner als die Summe aller seiner ICR (der Entsprechende Beweis ist im Paper vorhanden).
Jetzt kann ein Nash Gleichgewicht gefunden werden, nur durch Minimierung der einzelnen ICR, was den Vorteil hat, dass sich diese minimieren lassen nur durch Kontrolle von $\sigma_i(I)$ (der Strategie eines Spielers). Für alle Informationsmengen I der Informationspartition eines Spielers i, wird für alle $a \in A(I)$ entsprechend der Formel für ICR folgendes berechnet:

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

Dies entspricht dem Kontrafaktischen Verlust bei der Informationsmenge I und Aktion a, wie schon zuvor beschrieben. Die Autoren definieren jetzt $R_i^{T,+}(I,a) = max(R_i^T(I,a), 0)$, also auf den maximalen Kontrafaktischen Verlust und bilden daraus eine Strategie für Spieler i in der nächste Runde $T+1$. Auch dies erfolgt erst einmal in der formalen Definition des Papers:

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

Wie auch im Paper treffend ausgedrückt bedeutet dies, dass "Aktionen proportional zu der Menge an positivem Kontrafaktischen Verlust ausgewählt werden, der Auftritt wenn man nicht diese Aktion wählt. Hat keine Aktion einen positiven Wert, wird zufällig eine ausgewählt". Da wenn ein Spieler seine Aktionen entsprechend der letzten Definition auswählt $R_{i,imm}^T(I) \leq \Delta_{u,i}\sqrt{|A_i|}/\sqrt{T}$ gilt und entsprechend auch $R_i^T(I) \leq \Delta_{u,i}|I_i|\sqrt{|A_i|}/\sqrt{T}$ (mit $|A_i| = max_{h:P(h)=i}|A(h)|$), kann diese Wahl der Strategie laut den Autoren verwendet werden ein Nash Gleichgewicht zu berechnen (die Beweise sind auch hier im Paper vorhanden). Mit eigenen Worten bedeutet dies, weil die ICR durch $\Delta_{u,i}\sqrt{|A_i|}/\sqrt{T}$ (Max. Spanne im Verlust des Spielers, mal der maximalen Anzahl aller auswählbaren Zustände in denen Spieler i nach einer Historie h am Zug ist, mal der Wurzel des Nummer der aktuellen Runde) beschränkt sind, gilt zusammen mit $R_i^T \leq \sum_{I \in \mathrm{I}_i} R_{t,imm}^T(I)$ (wie zuvor auch schon definiert), dass auch der tatsächliche Gesamtverlust beschränkt wird. Die Herleitung eines Nash Gleichgewichtes ergibt sich aus dem schon aufgeführten Grund, dass sich in einem Nullsummenspiel Gewinn und Verlust ausgleichen.

3.3 Kontrafaktischer Verlust im Poker

Nachdem nun im Allgemeinen beschrieben wurde wie man ein Nash Gleichgewicht berechnen kann, bezieht sich das Paper direkt auf das Spiel Poker zwischen zwei Spielern in der Variante Texas Hold'em.

3.3.1 Abstraktion

Zunächst einmal wird die Art der Abstraktion des Spiels beschrieben, deren Ziel es ist die Anzahl der Informationsmengen auf eine berechenbare Größe zu reduzieren. Die Wettregelen werden hierbei voll übernommen und die Abstraktion bezieht sich nur auf die gespielten Karten. Die Karten werden entsprechend der *quadrierten Handstärke* (engl.: hand strenght squared) gruppiert. Die Handstärke ist hierbei die Gewinnwahrscheinlichkeit nur nach den

Karten, die ein Spieler aktuell sehen kann. Mit der quadrierten Handstärke ist das Quadrat der Handstärke nach der zuletzt aufgedeckten Karte gemeint. Durch das Quadrieren erhalten starke Blätter einen zusätzlichen Vorteil. Für die Abstraktion werden zu Beginn die aus den Startkarten resultierenden Sequenzen (Historien, z.B. Bube und Dame auf der Hand durch den Zufallsspieler) entsprechend ihrer quadrierten Handstärke in einen von 10 gleich großen Buckets einsortiert. Dann werden alle Sequenzen der Runde zwei, welche Präfixe aus den gleichen Buckets der Runde eins haben, in einen von 10 neuen Buckets einsortiert (wieder nach quadrierter Handstärke). In Runde zwei sind demnach alle Sequenzen einer Partition als ein paar von Zahlen darstellbar: Der Nummer des Buckets der vorherigen Runde und dem in der aktuellen (welcher von dem der vorherigen abhängt). Die wird nun in jeder Runde wiederholt, wobei die Einordnung immer wieder nach der Übereinstimmung in den vorherigen Buckets geschieht, wodurch Kartensequenzen in *Bucketsequenzen* (im Grunde Informationsmengen, also nicht mehr unterscheidbaren Historien) partitioniert werden. Das hieraus resultierende, abstrakte Spiel hat nach den Autoren eine Anzahl von 10^{12} Zuständen und 10^7 Informationsmengen.

3.3.2 Minimierung des Kontrafaktischen Verlustes im Poker

Die Minimierung wird nun auf dem abstrakten Spiel durchgeführt. Im Prinzip Spielen zwei Gegner ständig gegeneinander und benutzen die abgeleitete Strategie jeweils für die Runde $T + 1$. Nach einer bestimmten Anzahl Iterationen wird die auf diese Weise berechneten Strategien $(\bar{\sigma}_1^T, \bar{\sigma}_2^T)$ als das angenährte Ergebnisgleichgewicht verwendet. Im Grunde müssten alle $R_i^t(I, a)$ bei jeder Iteration gespeichert und danach aktualisiert werden. Hier verwenden die Autoren jedoch einen zusätzlichen Trick, welcher die Anzahl an zu aktualisierenden Informationsmengen deutlich reduziert. Kern ist hierbei der Zufallsspieler c, dessen Handlungen (da sie zufällig sind) keinen Einfluss auf die Strategien der Spieler haben (Zumindest in Bezug auf die Abstraktion durch die Buckets). Die Handlungen der echten Spieler werden daher in einer als "Joint Bucket Sequenz" bezeichneten Struktur zusammengefasst, für welche anschließend nur noch die Updates gemacht werden müssen. Genauere Angeben werden hierzu leider nicht gemacht.

4. Vergleich und Nutzen

Die beiden vorgestellten Algorithmen, "Fiktives Spiel" und "Verlustminimierung", schließen beide mit einem Abschnitt ab, indem sie ihren jeweiligen Algorithmus gegen jeweils andere antreten lassen. Leider gibt es in beiden keinen direkten Vergleich untereinander. Beide messen sich allerdings mit einem gemeinsamen Gegner mit dem Namen PsOpti. Ebenfalls gewinnen beide gegen diesen Gegner. Da die vorgestellten Ergebnisse allerdings sehr verschieden sind, lässt sich nur auf Grund der Paper keine konkrete Aussage darüber machen welcher der bessere ist. Einzig die höhere Komplexität der Abstraktion bei der Verlustminimierung spricht dafür, dass dieser Algorithmus im direkten Vergleich besser sein könnte.
Was wir von beiden Algorithmen lernen können ist, dass die als "Bucketing" (oder erstellen von Informationsmengen) bezeichnete Methode zur Abstraktion des Spielbaumes, allgemein sinnvoll zu sein scheint. Während der Text zum "Fiktiven Spiel" hier mehr allgemeine Erklärungen liefert, erhalten wird bei dem Text zur Verlustminimierung sehr exakte

Definitionen und für den der den Algorithmus genau überprüfen will auch die nötigen Beweise. So erfahren wir, dass Buckets nicht anders als Informationsmengen sind (nach dem Wortlaut von (M. Zinkevich & Piccione, 2007)), welche verschiedene Pfade durch den Spielbaum als einen erscheinen lassen. Letzten Endes verwenden beide Ansätze ein ähnliches Gedankengut, nur dass es im Fall der Verlustminimierung genauer spezifiziert wird. Was im Fiktiven Spiel als "Chance Node Elimination" bezeichnet wird und durch Konvertierungsmatrizen aufgelöst wird, stellt auch nur eine Möglichkeit dar Informationsmengen zu erstellen und so Historien zusammenzulegen und den Zustandraum zu reduzieren. Allgemein ist leicht zu sehen, dass irgendeine Art der Abstraktion notwendig ist, um die hohe Komplexität des Spiels in den Griff zu bekommen. Auch wenn der Autor von "Fiktives Spiel" keine Angaben zur Laufzeit seines Algorithmus macht, dürfte auch hier sehr viel Zeit benötigt werden. Für die Technik der Verlustminimierung wurde eine Anzahl von 100 Millionen Iterationen in 33 Stunden durchgeführt um auf eine stabile Lösung zu kommen. Ohne eine angemessene Abstraktion wird also (noch) kein Algorithmus erfolgreich eine Lösung in absehbarer Zeit berechnen. Für den Ansatz der Verlustminimierung spricht an dieser Stelle allerdings auch, dass keine aufwendige Berechnung von Konvertierungsmatrizen notwenig ist um den Übergang zwischen verschiedenen Buckets zu bestimmen.
Als abschließendes Wort sei bemerkt, dass der Ansatz der Verlustminimierung in Hinsicht auf eine mögliche Nachimplementation die bessere Wahl zu sein scheint. Zwar wirkt die Theorie komplexer, aber bedingt durch die genauen Spezifikationen, fällt eine Umsetzung, wenn der Stoff erst einmal verstanden ist, entsprechend leichter und im Anhang ist sogar fertiger Programmkode vorhanden.

References

Duziak, W. (2006). Using fictitious play to find pseudo-optimal solutions for full-scale poker. *In Proceedings of the 2006 International Conference on Artificial Intelligence (ICAI-2006)*.

M. Zinkevich, M. Bowling, M. J., & Piccione, C. (2007). 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. *University of Toronto, Department of Economics in its series Working Papers with number tecipa-231*, Seiten 1–7 und 15–23.

Stockhammer, P. (2006). Einführung in bayessche netzwerke. *Hauptseminar Wrtschaftsinformatik - TU Claustal*.