Strategien in unendlichen Spielen mit Liveness-Gewinnbedingungen

Syntheseverfahren, Optimierung und Implementierung

Nico Wallmeier

Lehrstuhl Informatik 7
RWTH Aachen

06.09.2007

Überblick über die Arbeit

Die Arbeit umfasst die folgenden Themen:

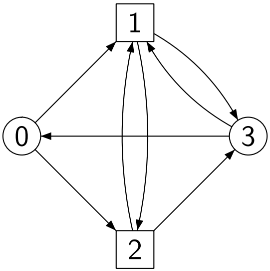
- Determinisierung von Büchi-Automaten
- Zwei-Personen-Spiele mit Streett-Gewinnbedingung und Anwendung auf Synthese von Live-Sequence-Charts
- Optimierung von Gewinnstrategien in Spielen mit Liveness-Bedingungen
- Implementierung einer Testplattform GASt: Determinisierung von Büchi-Automaten & unendliche Zwei-Personen-Spiele

Gliederung

1. Motivation und Hintergrund
   - Unendliche Zwei-Personen-Spiele
   - Request-Response-Spiele
   - Motivation
2. Definition und Eigenschaften von optimalen Strategien
   - Definition eines Partiewertes
   - Definition einer optimalen Strategie
   - Eigenschaften einer optimalen Strategie
3. Berechnung optimaler Strategien
   - Mean-Payoff-Spiele
   - Reduktion auf Mean-Payoff-Spiele

Unendliche Zwei-Personen-Spiele

- System-Modell: 2 Spieler
  - Kontroller (Spieler 0)
  - Umgebung (Spieler 1)
- **Spielgraph** (Arena)
  - Zustandsmenge $Q = Q_0 \dot\cup Q_1$
  - Transitionen $E \subset Q \times Q$ (jeder Zustand muss einen Nachfolger haben)



- **Partie** ρ ist eine unendliche Folge von Zuständen $\rho = \rho(0)\rho(1)\rho(2)\dots$ mit $(\rho(i), \rho(i+1)) \in E$
- **Gewinnbedingung** für Spieler 0

Request-Response-Gewinnbedingung

- **Request-Response-Gewinnbedingung**:
  - Mengen $P_i, R_i \subseteq Q$ for $1 \leq i \leq r$
  - (P_i, R_i) wird auch *RR-Paar* oder *RR-Bedingung* genannt
  - Die Gewinnbedingung ist gegeben durch
    - "Immer wenn ein Zustand von P_i besucht wird, wird jetzt oder später auch ein Zustand von R_i aufgesucht"
    - Oder in LTL: $\bigwedge_{i=1}^{r} \mathrm{G}(P_i \rightarrow \mathrm{F}\ R_i)$
- Verschärfung der Streett-Bedingung
- Spezialfall: Büchi Bedingung $(GF\ R_1)$

Reduktion auf Büchi-Spiele

Theorem

RR-Spiele sind auf Büchi-Spiele reduzierbar. Bei r RR-Paaren ergibt sich ein Blow-up von n auf $nr2^{r+1}$ Zustände.

Idee

Im erweiterten Spielgraphen werden die folgenden Informationen gespeichert:

- Welche Anforderungen sind noch nicht erfüllt worden?
- Welche Anforderung soll als nächstes erfüllt werden? (Zyklische Abarbeitung der Bedingungen)

Motivation

- RR-Gewinnbedingung stellt nur sicher, dass die Menge R_i irgendwann besucht wird (es wird nicht festgelegt wann).
- Im Kontext der Controllersynthese: Beschränkung/Minimierung der "Wartezeiten" erwünscht
- Beispiele: Aufzugsteuerung, Callcenter, ...

Plan

- Definition des Wertes einer Partie
- Definition optimaler Strategien in RR-Spielen
- Nachweis von optimalen Gewinnstrategien mit endlichem Speicher
- Schlüssel für die Berechnung optimaler Strategien besteht in der Nutzung von Ergebnissen über Mean-Payoff-Spiele

RR-Bedingungen

Definition

In einer Partie ρ eines RR-Spiels heißt die Bedingung (P_i, R_i) zur Zeit t

(a) *aktiv*, gdw. $\exists t' \leq t : \rho(t') \in P_i \wedge \forall s \in [t', t] : \rho(s) \notin R_i$

$$active(\rho, t, i) = \begin{cases} 1 & \text{zur Zeit } t \text{ ist Paar } i \text{ in Partie } \rho \text{ aktiv} \\ 0 & \text{sonst} \end{cases}$$

(b) gerade *aktiviert*, gdw. $\neg active(\rho, t-1, i) \wedge active(\rho, t, i)$

$$newActive(\rho, t, i) = \begin{cases} 1 & active(\rho, t, i) \wedge \neg active(\rho, t-1, i) \\ 0 & \text{sonst} \end{cases}$$

(c) gerade *erfüllt*, gdw. $active(\rho, t-1, i) \wedge \rho(t) \in R_i$

Wartezeit

Definition (Wartezeit)

Die Wartezeit einer Bedingung (P_i, R_i) in einer Partie ρ eines RR-Spiels zur Zeit t ist:

$$WZ(\rho, t, i) = \begin{cases} 0 & \neg active(\rho, t, i) \\ t - t' + 1 & \exists t' \leq t : newActive(\rho, t', i) \wedge \\ & \forall s \in [t', t] : active(\rho, s, i) \end{cases}$$

Definition (Warteschranke)

Die Warteschranke $wb(\rho, i)$ für die i-te RR-Bedingung in der Partie ρ ist das Maximum aller Wartezeiten dieser Bedingung, wenn dieses existiert, sonst ∞.

Wartezeit (2)

Bemerkung

Die Warteschranke $wb(\rho, i)$ ist unendlich, wenn in der Partie ρ

- die i-te Bedingung irgendwann nicht mehr erfüllt wird, oder
- die Folge der Wartezeiten unbeschränkt ist.

Konvention

In einer Partie ρ heißt Bedingung i unbeschränkt, wenn

- nach jeder Aktivierung irgendwann Erfüllung in ρ folgt, und
- die Wartezeit für diese Bedingung unbeschränkt ist.

Wartezeitvektor

Definition (Wartezeitvektor)

- Jeder Punkt einer Partie ρ eines RR-Spiels mit r Paaren kann um den aktuellen Wartezeitvektor (WZV) $\begin{pmatrix} WZ(\rho, t, 1) \\ \vdots \\ WZ(\rho, t, r) \end{pmatrix} \in \mathbb{N}^r$ zur Zeit t erweitert werden.
- Ein Spielgraph eines RR-Spiels wird entsprechend um die Wartezeitvektoren erweitert.
- Der resultierende Spielgraph G_r^+ ist i.a. unendlich.

Beobachtungen

Lemma (Bounded Waiting Time)

Sei σ eine Automatengewinnstrategie in einem RR-Spiel. Jede gemäß σ gespielte Partie ρ hat eine beschränkte Wartezeit bezüglich jeder RR-Bedingung.

Dieses Lemma gilt **nicht** für folgende Abschwächungen der Voraussetzungen:

- Gewinnstrategien von Spieler 0, die unendlichen Speicher haben;
- unendliche Spielgraphen.

Partiebewertungen: Erster Ansatz

Idee: durchschnittliche Anzahl aktiver Bedingungen

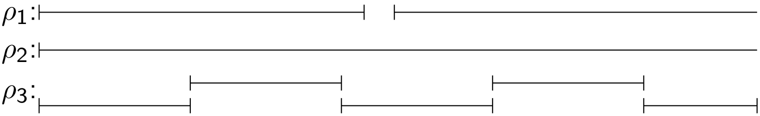
Definition (Partiewert)

Gegegeben sei ein RR-Spiel mit r Bedingungen. Dann ist der Partiewert $a(\rho)$ für eine Partie ρ definiert als

1. $a(\rho) = \limsup_{m \to \infty} \frac{1}{m} \sum_{t=1}^{m} \sum_{i=1}^{r} active(\rho, t, i)$

Diskussion von $a(\rho)$

- Keine Unterscheidung folgender Fälle:

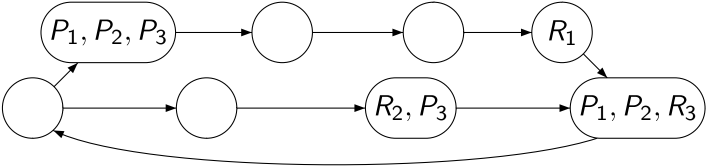


- Minimaler Partiewert $a(\rho)$ muss nicht gewinnbringend sein, obwohl Spieler 0 eine Gewinnstrategie hat.

Diskussion von $a(\rho)$ - 2

- Um Optimalität zu erzielen, reichen Gewinnstrategien mit endlichem Speicher nicht aus.

  Ein-Personen-Spiel:



Untere Schleife: $\frac{1}{4} \sum_{t=1}^{4} \sum_{i=1}^{r} active(\rho, t, i) = \frac{1}{4} \cdot 8 = 2$

Obere Schleife: $\frac{1}{6} \sum_{t=1}^{6} \sum_{i=1}^{r} active(\rho, t, i) = \frac{1}{6} \cdot 15 = \frac{5}{2}$

Zweiter Ansatz

Idee: Aufteilung von Wartezeiten auf Aktivierungen ("durchschnittliche Wartezeit")

Definition

Gegegeben sei ein RR-Spiel mit r Bedingungen. Dann ist der Partiewert $t(\rho)$ für eine Partie ρ definiert als

2. $$t(\rho) = \limsup_{m \to \infty} \frac{\sum_{t=1}^{m} \sum_{i=1}^{r} active(\rho, t, i)}{\sum_{t=1}^{m} \sum_{i=1}^{r} newActive(\rho, t, i)}$$

Lemma (Existenz des Limes)

Falls Spieler 0 die Partie ρ gewinnt und die Partie gemäß einer Automatengewinnstrategie gespielt hat, existiert der Wert $t(\rho)$.

Diskussion von $t(\rho)$

- Strafe bei wachsender Wartezeit wächst linear
- Keine Unterscheidung folgender Fälle:

ρ_5:

vs.

ρ_6:

Dritter Ansatz

Idee: Quadratische Gewichtung der Wartezeiten

Definition

3. $w(\rho) = \limsup_{m\to\infty} \frac{1}{m} \sum_{t=1}^{m} \sum_{i=1}^{r} WZ(\rho, t, i)$

Lemma (Existenz des Limes)

Falls Spieler 0 die Partie ρ gewinnt und die Partie gemäß einer Automatengewinnstrategie gespielt hat, existiert der Wert $t(\rho)$.

Definition (Segmentwert)

$$w(\rho, t_1, t_2) = \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} \sum_{i=1}^{r} WZ(\rho, t, i)$$

Diskussion von $w(\rho)$

- Strafe bei wachsender Wartezeit wächst nicht-linear
- Die Zeit bis zur Erfüllung fließt quadratisch ein: $\left(\sum_{i=0}^{n} i = \frac{n \cdot (n+1)}{2}\right)$
- Somit haben längere Wartezeiten einen größeren Einfluß
- Unterscheidung der vorgegangenen Fälle
- "Optimal" bedeutet jetzt "optimal bezüglich $w(\rho)$"

Strategiewert

Zu Strategien σ, τ ist $\rho(\sigma, \tau, q_0)$ die induzierte Partie.

Definition (Strategiewert)

Der Wert einer Strategie σ von Spieler 0 für einen Zustand q_0 für ein gegebenes RR-Spiel ist definiert als

$$w(\sigma, q_0) = \sup_{\tau \in \text{Strategie von Spieler 1}} w(\rho(\sigma, \tau, q_0))$$

Korollar (Existenz des Supremums)

Für ein RR-Spiel, eine Automatengewinnstrategie σ von Spieler 0 und einen Anfangszustand $q_0 \in W_0$ existiert das Supremum $w(\sigma, q_o)$.

Optimalität einer Gewinnstrategie

Definition

Eine Gewinnstrategie σ ist *optimal* für ein gegebenes RR-Spiel mit Anfangszustand q_0 gdw.

$$w(\sigma, q_0) \leq w(\tau, q_0) \text{ für alle Strategien } \tau$$

Bemerkung

Die Optimalitätsbedingung stellt nicht sicher, dass es eine eindeutige optimale Strategie für ein gegebenes RR-Spiel mit Anfangszustand q_0 gibt.

Obere Schranke des Strategiewerts

Korollar

In einem RR-Spiel mit n Zuständen und r Paaren gilt für alle Werte einer optimalen Strategie immer:

$$w(\sigma, q_0) \leq \frac{n \cdot r^2 + r}{2} \text{ für } q_0 \in W_0$$

Beweis.

- Büchi-Reduktion: Bedingung ist höchstens $n \cdot r$ Schritte aktiv
- Summe der Wartezeiten für eine Bedingung, die $n \cdot r$ Schritte aktiv ist: $\sum_{i=0}^{n \cdot r} i = \frac{n \cdot r \cdot (n \cdot r + 1)}{2}$
- Der Strategiewert ist kleiner oder gleich: $\frac{1}{n \cdot r} \cdot \frac{n \cdot r^2 \cdot (n \cdot r + 1)}{2} = \frac{n \cdot r^2 + r}{2}$

□

Hauptergebnis

Theorem

Für jedes RR-Spiel existiert eine optimale Gewinnstrategie, die eine beschränkte Wartezeit für alle RR-Paare garantiert.

<u>Annahme für den Spezialfall</u>: Strategie σ ist optimal und nur ein RR-Paar ist unbeschränkt.

<u>Motivation</u>: Verringere große Werte der ersten Komponente des WZV durch eine Strategieänderung.

Betrachte eine Gewinnstrategie σ, die

- optimal für Spieler 0 ist,
- bei der das erste RR-Paar (P_1, R_1) unbeschränkt auftritt und
- für $2 \leq i \leq r$ das i-te Paar durch s_i beschränkt ist.

Spielbaum für feste Strategie

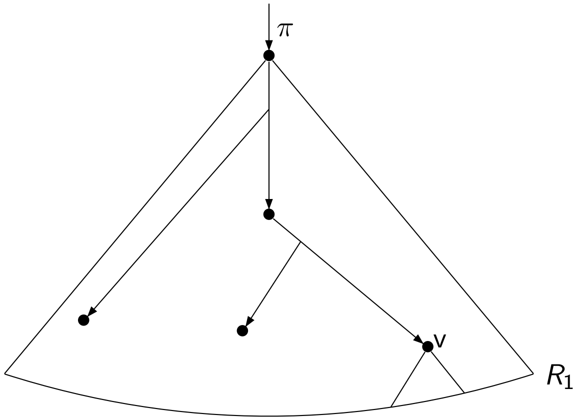
Definition

Für ein Partiepräfix $\pi = \pi(0) \dots \pi(n)$ und eine Strategie σ von Spieler 0 betrachte Spielbaum t_π^σ mit Knoten $\pi(n)$ als Wurzel. Alle gemäß σ möglichen Partien vom Präfix π bilden die Pfade in t_π^σ.

Um eine neue Gewinnstrategie σ' für ein Partiepräfix π zu definieren, betrachte den Spielbaum $\overline{t_\pi^\sigma}$, bei dem alle Pfade mit dem ersten Auftreten von R_1 enden.
Bemerkung: $\overline{t_\pi^\sigma}$ ist endlich, wenn $\pi(0) \in W_0$

Beispiel

Spielbaum $\overline{t_{\pi}^{\sigma}}$:



Strategiekonstruktion

- Betrachte alle Knoten im Baum, die mit Ausnahme der ersten Komponente des WZV identisch zur Wurzel sind.
- Wähle den Knoten v mit maximaler erster Komponente des WZV aus dieser Menge aus (ist dieser nicht eindeutig, wähle beliebigen davon)

Neue Gewinnstrategie σ' wird wie folgt definiert:

- Ist das erste Paar nicht aktiv zur Zeit t, verhält sich die Strategie σ' wie σ
- Ist $w(\tau, t, v) > w(\sigma, \tau(0))$: Strategie σ' verhält sich für das Partiepräfix wie die Strategie σ bei dem Knoten v und die erste Komponente des WZV wird auf den Wert von v gesetzt.
- Sonst verhält sich die Strategie σ' wie σ

Plan

Korollar

Endlicher Speicher reicht für eine optimale Strategie in einem RR-Spiel aus.

- Optimale Strategien hängen nur von aktuellen Wartezeiten ab.
- Schranke: $w_{max}(n, r) = \mathcal{O}\left(n^{\mathcal{O}((r-1)!)}\right)$

Theorem

Für ein RR-Spiel ist eine optimale Strategie für Spieler 0 berechenbar.

Idee:

- Erweiterung des Spielgraphen um die Wartezeitvektoren, die durch $w_{max} := w_{max}(n, r)$ für alle Paare beschränkt sind.
- Reduktion des RR-Spiels auf ein Mean-Payoff-Spiel.

Mean-Payoff-Spiele

Definition (Mean-Payoff Game, Ehrenfeucht, Mycielski '79)

Sei $G = (Q, E)$ ein Spielgraph mit Kantenbewertung $w : E \rightarrow \{-W, \ldots, 0, \ldots, W\}$. Spieler 0 versucht, $\liminf\limits_{n\rightarrow\infty} \frac{1}{n} \sum\limits_{i=1}^{n} w(\rho_i)$ zu maximieren, während Spieler 1 versucht, $\limsup\limits_{n\rightarrow\infty} \frac{1}{n} \sum\limits_{i=1}^{n} w(\rho_i)$ zu minimieren.

Theorem (Zwick, Paterson '96)

In einem Mean-Payoff-Spiel $G = (Q, E)$ mit einer Gewichtsfunktion $w : E \rightarrow \{-W, \ldots, W\}$ sind positionale optimale Strategien für beide Spieler in Zeit $\mathcal{O}(|Q|^4 \cdot |E| \cdot \log(|E|/|Q|) \cdot W)$ berechenbar.

Erweiterter Spielgraph

- $Q'_i := Q_{1-i} \times \{0, \ldots, w_{max} + 1\}^r$
- $((q, i_1, \ldots, i_r), (q', i'_1, \ldots, i'_r)) \in E' :\Leftrightarrow$
  - $(q, q') \in E$
  - $i'_k = \begin{cases} i_k + 1, & \text{falls } 1 \leq i_k \leq w_{max} \wedge q' \notin R_k \\ 0, & \text{falls } q' \in R_k \wedge i_k \leq w_{max} \\ 1, & \text{falls } i_k = 0 \wedge q' \in P_k \\ i_k, & \text{sonst} \end{cases}$
- $f : Q \to Q'$ wie folgt:

  $f(q) = (q, i_1, \ldots, i_r)$ mit $i_k = \begin{cases} 1 & \text{falls } q \in P_k \\ 0 & \text{sonst} \end{cases}$

Reduktion auf Mean-Payoff-Spiele

- Die Gewichtsfunktion w summiert für alle aktiven RR-Paare die Wartezeit seit ihrer Aktivierung auf, solange diese kleiner w_{max} ist.
- Für RR-Paare, bei denen die Wartezeit den Wert w_{max} überschritten hat, wird der Wert $(r \cdot w_{max} + 1)$ hinzugefügt.
- Die positionalen optimalen Gewinnstrategien im Mean-Payoff-Spiel gemäß Zwick/Paterson liefern optimale Automatengewinnstrategien im gegebenen RR-Spiel.
- Komplexitätsproblem: Hyperexponentielle Expansion bei der Reduktion

Zusammenfassung und Ausblick

Spiellösung nicht als Entscheidungsproblem (wer gewinnt?) sondern als Optimierungsproblem (wie am besten?).

- Verschiedene Ansätze für Partiewerte eines RR-Spiels wurden diskutiert
- Optimalität von Strategien in RR-Spielen wurde definiert
- Berechnung optimaler Strategien wurde vorgestellt

Ausblick:

- Heuristiken für praktische Lösung
- Betrachtung komplexerer Gewinnbedingungen
- Betrachtung von Zeitautomaten