Zufall durch Spiel für mehr Vertrauen bei Entscheidungen

Seminar Künstliche Intelligenz
Marko Stankovic

Überblick

1. Vertrauen durch Zufall oder Spiel?
2. Spielbasierte Mechanismen
3. Bewertung und Ausblick

Überblick

1. Vertrauen durch Zufall oder Spiel?
    a. Motivation: Fairness
    b. Organspendeskandal
    c. Deterministische Methoden
    d. Klassischer Zufall - Vorteile
    e. Münzwurf
    f. Klassischer Zufall - Nachteile
    g. Mögliche Lösung: Spiel-basierter Ansatz von Toby Walsh
2. Spielbasierte Mechanismen
3. Bewertung und Ausblick

Motivation: Fairness

- Menschen treffen alltäglich Entscheidungen im Alltag
  - Oftmals Ziel: Möglichst Fair zu sein
- Entscheidungen mit hohen ethischen Anforderungen
- Bspw.
  - Schulplatzvergabe, Organspenden, generell Ressourcenzuteilung
- Fairness = zentrale gesellschaftliche Anforderung

  → Fairness resultiert in Vertrauen für das System

- Aber was ist Fair? Ist das immer möglich?

Organspendeskandal - Vertrauensverlust

- 2012: Manipulation von Wartelisten in Deutschland
- Folge: 30% Rückgang in Spendenbereitschaft
- → Enormer Vertrauensverlust der Bevölkerung

Anforderung:

- ➔ Vertrauen zurückgewinnen
- ➔ Mechanismus muss **vertrauenswürdig** , **transparent** und **überprüfbar** sein
- ➔ Wie können wir solche Entscheidungen treffen?

Deterministische Methoden

- Vollständigkeitshalber → Deterministische Methoden
- Regelbasierter Ansatz
- Ergebnis durch Regeln bestimmt
- Vorteile: Nachvollziehbar, Auditierbar

Nachteil: Keine Lösung für Gleichstände oder Unklarheiten

Wem allokiere ich nun die Ressource bspw. ?

Klassischer Zufall

- Zufall als Mittel zur fairen Auswahl
- Zwei oder mehrere gleichwertige Optionen
- Jeder hat in Theorie gleiche Wahrscheinlichkeit

➔ Im Gegensatz zu deterministischen Verfahren bietet Zufall weitere normative Eigenschaften

Klassischer Zufall - Vorteile

- **Pareto-effizient**
  - Niemand wird schlechter gestellt, ohne dass jemand besser gestellt wird
- **Strategiesicher**
  - Kein Anreiz zu lügen (Falsche Symptome)
- **Anonym**
  - Keine Bevorzugung durch Namen, Herkunft etc.

Münzwurf 🕹️

Münzwurf - Nachvollziehbarkeit

- Offensichtlich anfällig für **Manipulation**
- **Keine Nachvollziehbarkeit** von Außen → **BlackBox**

Klassischer Zufall - Nachteile

Weitere technischen Limitierungen:

- Echter Zufall **schwer zu erzeugen** (mit konventioneller Hardware)
  - oftmals nur **Pseudozufall**
- **Einmalige** Entscheidungen **kaum überprüfbar**
- Selbst bei Wiederholungen kein Beweis für Fairness
  - **Stichprobe** muss **groß** sein
- **Hohe Komplexität** und Hardwareauslastung

Mögliche Lösung: Spiel-basierter Ansatz von Toby Walsh

- Zufall stark in der Theorie
  - normative Eigenschaften, Fairness
- Problem: mangelnde Prüfbarkeit, Manipulierbarkeit

Toby Walsh bietet möglicherweise eine Lösung in seinem Paper “Mechanisms That Play a Game, Not Toss a Coin” an.

➔ **Normative Vorteile** behalten

➔ **Probleme** durch Spiel **lösen**

Überblick

1. Vertrauen durch Zufall oder Spiel?
2. Spielbasierte Mechanismen
   a. Grundkonzept von spiel-basierten Mechanismen
   b. Beispiel Parity Game
   c. Interaktives Spiel: Parity Game
   d. Die Drei Methoden
   e. Random Dictator - Beispiel für Game-First
   f. Random Dictator Interaktiv 🕹️
   g. Probabilistic Serial - Beispiel für Game-Last
   h. Probabilistic Serial Interaktiv 🕹️
   i. Resource Allocation - Beispiel für Game-Interleaved
3. Bewertung und Ausblick

Grundkonzept von spiel-basierten Mechanismen

- Idee: Zufall durch Spiel statt durch Hardware
- Wieso:
  - Normative Eigenschaften (Fairness, Strategiesicherheit etc.) beibehalten
  - Gleichzeitig Zufall nachvollziehbar machen
- Konzept: “Derandomisierung” in klassische Entscheidungssituationen einbinden

Methoden:

1. Game-First: Spiel → Entscheidung
2. Game-Last: Entscheidung → Spiel
3. Game-Interleaved: Abwechselnd Spiel & Entscheidung

Beispiel: Parity Game

Grundpfeiler für alle kommenden spielebasierten Ansätze

Regeln:

- Zwei Spieler A & B
- Jeder wählt 0 oder 1 → XOR
- Gleiche Zahlen: A gewinnt
- Ungleiche Zahlen: B gewinnt

Hoffnung:

➔ Spieler wählen zufällig
➔ Ergebnis resultiert in echtem Zufall

→ Was glaubt ihr?

Beispiel: Parity Game

- Spiel ist einfach, aber extrem mächtig
- Für beide Spieler ist die optimale Strategie zufällig zu spielen
  - kein systematisches bevorzugen einer Zahl
- Ein einziges Nash Gleichgewicht: Beide Spieler spielen zufällig

➔ Zufallsergebnis durch rationales Spiel

➔ Komplett nachvollziehbar
  - für alle Parteien

Spieler A	Spieler B	Ergebnis
0	0	**A** gewinnt
1	1	**A** gewinnt
0	**1**	**B** gewinnt
1	**0**	**B** gewinnt

Die drei spielbasierten Methoden

- Parity Game dient als Basis für andere Fälle
- Drei systematische Arten der Derandomisierung

Unterscheiden sich im Zeitpunkt:

➔ **Game-First:**

Spiel → Entscheidungsprozess → Ergebnis

➔ **Game-Last:**

Entscheidungsprozess → Spiel → Ergebnis

➔ **Game-Interleaved:**

Spiel → Entscheidungsprozess → Spiel ... → Ergebnis

Entscheidungsprozess → Spiel → Entscheidungsprozess → ... → Ergebnis

Die drei spielbasierten Methoden

- Jeweils ein Beispiel pro Methode
- Klassisches Entscheidungsproblem (ursprünglich kein Spiel)

→ Durch Derandomisierung versucht …

- transparenter
- nachvollziehbar
- normativ

… zu sein

Random Dictator - Game-First

Allgemein Random Dictator → Danach Derandomisierung mittels Spiel

- Jede Person hat **klare Präferenz** (Bspw. Restaurant)
- **Zufällig** wird **eine Person** als **Diktator** erkoren
  - (Würfelwurf ? Random-Zahlengenerator? ...)
- Diese **Person wählt dann für alle**
- Problem: Möglichkeit fürs mogeln (bspw. gezinkter Würfel)

➔ Lösungsansatz durch Game-First von Toby Walsh

Random Dictator Interaktiv 🕹️

- Jeder repräsentiert eine Zahl: 1 bis n
- Schreibt eine Zahl zwischen 0 und n

Die Summe dieser Zahl wird dann modulo n gerechnet und wählt den Diktator

Optionen:

A. Steakhouse **B. Burgerladen** **C. Salatbar**

Random Dictator – Einordnung

- Zufall durch Spielhandlung erzeugt
- Transparent und Nachvollziehbar
  - Nachverfolgbar wie Diktator gewählt wurde
- ! Bei wenigen Spielern nicht strategischer

→ Bei vielen Spielern: Optimales handeln durch zufällige Wahl

Probabilistic Serial - Game-Last

- Akteure ***n*** begehren ***m*** Objekte
- Jeder Akteur hat Präferenzliste als Rangfolge
    - Nicht teil des Spiels!
- Anhand Präferenzliste → Wahrscheinlichkeit
- Zuteilung mit gegebenen Wahrscheinlichkeiten

Akteur 1: 20% | Akteur 2: 80% für Objekt 1

→ In 4 von 5 Fällen kriegt Akteur 2 das Objekt

$$\frac{(n!)^m - 1}{(n!)^m}$$

$$\left\{0, \frac{1}{36}, \frac{2}{36}, \frac{3}{36}, \ldots, \frac{35}{36}\right\}$$

3 Akteure
2 Objekte

Probabilistic Serial Interaktiv 🕹️

- Probabilistic Serial → Jeder Akteur hat eine **Konsumrate**
- Desto länger ein Akteur ein Objekt “konsumiert”

→ Höhere Wahrscheinlichkeit es zugelost zu kriegen

Annahme: Konsumrate von 0.5 Einheiten die Sekunde

Probabilistic Serial Interaktiv 🕹️

- 3 Freiwillige erstellen ihre Präferenzliste:

Zur Auswahl:

Kuchen 🎂 & Eis 🍧

Mögliche Präferenzlisten:

➔ Kuchen, Eis
➔ Eis, Kuchen

Probabilistic Serial Interaktiv 🕹️

Konsumrate: 0.5 pro Sekunde
Akteure: 🔴, 🟡, 🔵

Option 1: Alle haben gleiche Präferenzliste

🔴, 🟡, 🔵 = Kuchen > Eis

Anfang:
🔴, 🟡,🔵 → 🎂

Nach 0,66 Sekunden:
🔵,🔴, 🟡 → 🍧

Wahrscheinlichkeit 🎂:

P🔴,🟡,🔵(🎂) = 1/3

Nach 1,33 Sekunden:

Wahrscheinlichkeit 🍧:

P🔴,🟡,🔵(🍧) = 1/3

Probabilistic Serial Interaktiv 🕹️

Konsumrate: 0.5 pro Sekunde
Akteure: 🔴, 🟡, 🔵

Option 2: Einer weicht ab

🔴, 🟡 = Kuchen > Eis

🔵 = Eis > Kuchen

Anfang:

🔴, 🟡 → 🎂

🔵 → 🍧

Nach 1 Sekunde:

🔵, 🔴, 🟡 → 🍧

Wahrscheinlichkeit 🎂:

P🔴,🟡(🎂) = 1/2

Nach 1,33 Sekunden:

Wahrscheinlichkeit 🍧:

P🔵(🍧) = 4/6

P🔴,🟡(🍧) = 1/6

Probabilistic Serial Interaktiv 🕹️ - Die Zuweisung

- 3 Akteure = n, 2 Objekte = m
- Jede Zahl in der Menge entspricht einem Index
  - 0 bis $(n!)^m - 1$
- Jeder Index entspricht einer Verteilung
  - Bspw. 0 = {🔴, 🎂, 🍧} und 1 = {🔴, 🎂} und {🔵, 🍧}
  - Wahrscheinlichkeiten entsprechend

$$\frac{(n!)^m - 1}{(n!)^m}$$

$$\left\{0, \frac{1}{36}, \frac{2}{36}, \frac{3}{36}, \ldots, \frac{35}{36}\right\}$$

Akteure nennen eine Zahl zwischen 0 und $(n!)^m - 1$

→ Summe mod $(n!)^m$ = Index

→ Index → **Endgültige Verteilung**

Probabilistic Serial - Einordnung

- Zufall erfolgreich mit “Game-Last” ersetzt
- Deterministisch und nachvollziehbar

- Allg. Probabilistic Serial → **Nicht Strategiesicher**
    - Anreize eigentliche Präferenz zu unterschlagen
    - Weniger populäres Objekt als erstes sinnvoller

Resource Allocation - Game-Interleaved

- Entscheidungsmechanismus und Spiel wechseln sich ab

- Maximal Zwei Akteure
- Angabe der Wichtigkeit für Ressource mit Limit
- Bei gleichstand Zufall üblicherweise

→ Hier Game-Interleaved Derandomisierung

Resource Allocation - Game-Interleaved

Annahme: Zwei Studenten **A** und **B** bieten für 4 Gegenstände mit 20€

- ➔ vertrauliche Angabe von Geboten für Objekte
- ➔ Immer nur für ein Objekt bieten
- ➔ Höheres Gebot bekommt Objekt
- ➔ Bei Gleichstand Spiel → Parity Game
    - ◆ Zufällige Zahl 1 oder 0
    - ◆ XOR → 0 = A, XOR → 1 = B

Resource Allocation - Game-Interleaved

	1. Stift	2. Taschenrechner	3. Altklausuren	4. Deutschlandticket
A:	1€	5€	9€	5€ ✅
B:	5€ ✅	5€ ✅	9€ ✅	1€

	→ Spiel 0 XOR 1 → B	→ Spiel 1 XOR 0 → B	

Resource Allocation - Einordnung

- Spiel erzeugt fairen Zufall
- Deterministisch und nachvollziehbar
- Robust gegen Manipulation
  - keine einzelne Partei hat alleinige Kontrolle

- Resource Allocation in der Form

➔ → Nicht Strategiesicher!

➔ → Spieler können wenig bieten falls → anderer Spieler kein Interesse hat
  - Nur Vermutung normalerweise
  - Budget limitiert Strategie

Überblick

1. Vertrauen durch Zufall oder Spiel?
2. Spielbasierte Mechanismen
3. Bewertung und Ausblick
    a. Zusammenfassung
    b. Evaluation + Eigene Meinung

Zusammenfassung

- Ursprungsziel: Mehr Vertrauen in Entscheidungen festigen!
- Derandomisierung durch Spiele
  - Statt klassischem Zufall → Zufall durch Spiel

➔ Beibehaltung normativer Eigenschaften
  - **Pareto-Effizienz, Strategiesicherheit? und Anonymität?**

➔ Nachvollziehbarkeit und Transparenz

➔ Drei Methoden für klassische Probleme
  - Game-First, Game-Last und Game-Interleaved

Evaluation

Pareto-Effizienz:

→ Grundlegend in den Beispielen ja

Strategiesicher:

→ Spielmechanischer Anteil ja

→ Ursprungsprobleme aber nicht

Anonymität:

→ Im klassischen Sinne nicht, aber → Keine Diskriminierung möglich

Evaluation + Eigene Meinung

Ziel des Papers: Erfüllt

- Fairness
- Robustheit gegen Manipulation
- Transparenz
- +/- Normative Eigenschaften

Kritikpunkte:

- Praktische Umsetzbarkeit schwierig
- Ethische Bedenken → "Spiel für Ressourcen wie Organspenden"
- Können Menschen "echten" Zufall

Vielen Dank für eure Aufmerksamkeit!

Gibt es noch Fragen?

Quellen

Inhalt:

Walsh, Toby : **Mechanisms That Play a Game, Not Toss a Coin** . In: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24), 2024, Seiten 3005–3013. https://arxiv.org/abs/2308.10413

Bilder:

https://www.pngwing.com/en/free-png-nokzt

https://www.pngwing.com/en/free-png-mebkn/download