Prüfungsprotokoll

Diplomprüfung

Vertiefung: Künstliche Intelligenz

in den Fächern:

CSMR (Computer-Supported Modeling and Reasoning)
Advanced AI Techniques
Spieltheorie (WS 07/08)

26.April 2008
Prüfer: Prof. Dr. Nebel, Prof. Dr. Burgard
Note: 1,0
Kerstin Haring

Bemerkung:
Die Reihenfolge der Prüfungen war frei wählbar. Ich habe mit Spieltheorie begonnen. Die Fragen sind fett dargestellt, die Antworten normal und meine Kommentare im Nachhinein sind kursiv. Da ich dieses Protokoll im Nachhinein verfasst habe, kann ich mich nicht 100% an die Fragestellungen erinnern, bin mir aber sicher immer die relevanten Themengebiete abzudecken.

CSMR:
Schein eingereicht. Note: 1,3
Wer einen Schein einreicht sollte ihn auf jeden Fall dabei haben!

Spieltheorie:

Was ist denn Spieltheorie?
Spieltheorie ist die Analyse strategischer Entscheidungssituationen, in denen mehrere Spieler miteinander interagieren.
Dabei ist das Resultat eines Spiels von den Entscheidungen der Mitspieler abhängig. Damit stellt sich die Frage nach dem Ergebnis, das sich ergibt, falls alle Spieler "rational" handeln, d.h. ihren (erwarteten) Nutzen maximieren, wobei sie davon ausgehen, dass ihre Mitspieler ebenso rational handeln. Genau wegen dieser Rationalität ist die Spieltheorie für die Informatik interessant.

Kennen Sie denn ein Spiel, bei dem ein Spieler Interesse hat, nicht rational zu spielen?
Naja, wenn z.B. ein Spieler dem anderen das Schlechteste wünscht und dafür selber Einbußen in seinem Nutzen in Kauf nimmt.
Haben da noch ein bisschen diskutiert, ich bekam ein Hinweis auf ein weiteres Spiel, bei dem durch Unwissen irrational gespielt wurde, und mir fiel dann noch das 2/3 average Spiel ein, das wir sogar in der VL gespielt haben, das habe ich dann noch erklärt.

Was ist denn ein strategisches Spiel? Woraus besteht es?
- $G = \langle N, (A_i), (u_i) \rangle$

- N ist endliche Spielermenge
- nicht-leere Menge A_i von Aktionen/Strategien für jeden Spieler i
- Auszahlungsfunktion $u_i: A \to \mathbb{R}$ für jeden Spieler i
- $A = \prod_{i \in N} A_i$
- G heißt endlich, wenn A endlich ist.
- Beide Spieler legen vor Spielbeginn ihre Strategie fest, Ergebnis ist Strategiekombination.

Was sind dominate Strategien? Definition?
- Spiel $G = \langle N, (A_i), (u_i) \rangle$
- Eine Aktion $a_j^x \in A_j$ heißt strikt dominiert, falls es
- eine Aktion $a_j' \in A_j$ gibt, so dass
- $u_j(a_{-j}, a_j') > u_j(a_{-j}, a_j^x)$ für alle Profile $a \in A$ gilt
- Es ist nicht rational strikt dominierte Strategien zu spielen, weil man dann garantiert weniger bekommt als man bekommen könnte. Es ist auch nicht rational, strikt dominante Strategien nicht zu spielen.

Sind die Ergebnisse eindeutig?
Ja, das Ergebnis bei iterativer Elimination strikt dominierter Strategien ist eindeutig und damit unabhängig von der Reihenfolge der Elimination.

Was für eine Dominanz gibt es noch? Sind hier die Ergebnisse eindeutig?
Schwache Dominanz, ebenfalls mit vollständiger Definition erklärt und gesagt, dass die schwache Dominanz nicht eindeutig sein muss, weil das Ergebnis von der Eliminationsreihenfolge abhängt.

Was ist ein Nash-Gleichgewicht? Definition?
Ein NG ist ein Strategiekombination, in dem kein Spieler durch Abweichung einen Vorteil erlangen kann.

(1) Ein Profil $a^x \in A$ ist ein NG in einem strategischen Spiel $G = \langle N, (A_i), (u_i) \rangle$ gdw.
$u_i(a^x) = u_i(a_{-i}^x, a_i^x) \geqslant u_i(a_{-i}^x, a_i)$ *f.a* $a_i \in A_i$ *f.a.* $i \in N$

Welche Definition hatten wir noch? Wofür braucht man die denn?

(2) Ein Profil $a^x \in A$ ist ein NG in einem strategischen Spiel $G = \langle N, (A_i), (u_i) \rangle$ gdw.
$B_i(a_i) = \{a_i \in A_i \mid u_i(a_{-i}, a_i) \geqslant u_i(a_{-i}, a_i')\}$ *f.a* $a_i \in A_i$ *f.a.* $i \in N$ und
$a_i^x \in B_i(a_{-i}^x)$ *f.a.* $i \in N$
hier hat Prof. Nebel gemeint, ich soll das mal wieder weg machen, das wäre keine Definition. Erinnerte mich aber an das Skript und habe dann natürlich erklärt, dass man das auf dem Weg zu der dritten Definition braucht. Das hat er dann auch eingesehen.

(3) Ein Profil $a^x \in A$ ist ein NG gdw.
$B(a^x) = \prod_{i \in N} B_i(a_{-i}^x)$ und $a^x \in B(a^x)$ (wird für Beweis des Satzes von Nash verwendet)

Wie lautet der Satz von Nash?
Jedes endliche strategische Spiel ein NG in seiner gemischten Erweiterung.

Was sind gemischte Erweiterungen?
Die gemischte Erweiterung eines Spiels $G = \langle N, (A_i), (u_i) \rangle$ ist $G = \langle N, (\Delta A_i), (U_i) \rangle$

- ΔA_i ist die Menge der Wahrscheinlichkeitsverteilungen über die Menge A
- $U_i: \prod_{j \in N} \Delta(A_j) \rightarrow \mathbb{R}$ jedem Profil α den erwarteten Nutzen von Spieler i unter der von α induzierten Wahrscheinlichkeitsverteilung zuordnet.

Hier irgendwo kamen wir auf das Elfmeterspiel, keine Ahnung wie genau, es kam auf jeden Fall zur Sprache.

Wie berechnet man da NGs?
Mit dem Support Lemma.

Support-Lemma Definition? Was folgt daraus?
Habe mich tatsächlich an einem lockeren Spruch versucht:
Also, es sagt, dass wir zuerst den support definieren müssen:
- Sei α_i eine gemischte Strategie. Der support von α_i ist die Menge
- $supp(\alpha_i) = \{a_i \in A_i \mid \alpha_i(a_i) > 0\}$

- Sei $G = \langle N, (A_i), (u_i) \rangle$ strategisches Spiel
- Dann ist $\alpha^x \in \prod_{i \in N} \Delta(A_i)$ ein NG in gemischten Strategien gdw. alle reinen $a_i \in supp(\alpha_i)$ für jeden Spieler *i* eine beste Antwort auf α^x_{-i} ist.

Wenn die anderen Spieler ihre gemischte Strategie beibehalten, ist es also für jeden einzelnen Spieler egal, ob er seine gemischte Strategie oder eine Einzelaktion daraus spielt.
Aktionen, die in dem support einer gemischten Strategie in einem NG sind, sind immer beste Antworten auf das NG Profil und haben damit auch den gleichen Nutzen!

Wie wird das dann berechnet?
Benutze stattdessen Instanzen des Linear Complementarity Problem.
Ich habe gehofft, dass er nicht weiter nachfragt

Und wie geht das dann konkret?
D`oh! Schwachstelle erwischt.
Habe was von linearen Ungleichungen erwähnt, bin kurz in die Lösung von NSS mit LP abgedriftet und es dann gemerkt und zugegeben, dass ich hier nicht weiterkomme.

Sollen wir lieber an einer anderen Stelle weitermachen?
Einfachste Frage der Prüfung.
Ja gerne

Ist ein NG bei allen Spielen aussagekräftig?
Nein, nicht bei extensiven Spielen:
(habe hier was von übersetzen in strategische Form und leeren Drohungen erwähnt)
$\Gamma = \langle N, H, P, (u_i) \rangle$
- endliche nicht-leere Menge N von Spielern

- Menge H von Sequenzen (Historien), Menge Z der terminalen Historien

- Spielerfunktion $P: H \setminus Z \rightarrow N$

- Auszahlungsfunktion $u_i: Z \rightarrow \mathbb{R}$ *f.a.* $i \in N$

Extensive Spiele können unendlich viele Verzweigungen haben und die Pfade können unendlich lang sein.

- Wir haben deshalb auch zwei Formen der Endlichkeiten, und zwar heißen Spiele endlich, wenn ihre Menge der Historien H endlich ist, und sie haben einen endlichen Horizont, wenn die Länge der Historien h nach oben beschränkt ist.

Was sind denn leere Drohungen?
Wenn eine Strategie ein Verhalten festlegt, dass im Falle des Erreichens eines bestimmten Knotens nicht rational ware.

Beispiel 56 Skript:
Was damit gemeint ist, erkennt man besonders gut am Nash Gleichgewicht (B,L) .

- 1 würde natürlich am liebsten A wählen, in der Hoffnung, 2 werde dann R spielen und 1 würde eine Auszahlung von 2 erhalten.
- 2 droht aber damit, auf A mit L zu antworten, so dass es für 1 günstiger ist, B zu wählen und wenigstens eine Auszahlung von 1 zu erhalten.
- Würde 1 aber ungeachtet der Drohung doch A wählen, wäre es für 2 besser statt die Drohung auszuführen und L zu spielen, was ihm eine Auszahlung von 0 einbrächte, lieber entgegen seiner Ankündigung R zu wählen und eine Auszahlung von 1 zu erhalten.
- Ist 1 von der Rationalität von 2 überzeugt, wird er der Drohung also nicht glauben.

Wie sind TPG definiert?
Es stellt eine Verfeinerung des NGs dar, d.h.: jedes teilspielperfekte Gleichgewicht ist auch ein NG. Ein NG ist teilspielperfekt, wenn es ein NG in jedem Teilspiel von G induziert.

Wie heißt der Satz?
Kuhn: Jedes endliche extensive Spiel mit perfekter Information hat ein TPG

Wieso nur bei endlichen Spielen?
Habe die 2 Gegenbeispiele aus dem Skript aufgemalt und erklärt

Was hat das ganze mit Minimax und NSS zu tun?

- Das ist die Erweiterung von NSS und dem Minimax-Algorithmus aus er KI -VL
- Rückwärts Induktion ist Verallgemeinerung des Minimax-Algorithmus

Ähnlichkeiten Kuhn zu Minimax:
Lösung wird gefunden, indem der Baum von unten nach oben abgesucht wird und in jedem Knoten der optimale Spielzug gewählt wird und die Werte nach oben weiter-propagiert werden.

Unterschiede: Im Fall der Rückwärts Induktion sind mehr als 2 Spieler möglich und der Nutzen (payoff) ist nicht nur eine einzelne Zahl sondern ein ganzes Nutzenprofil.

Welche Erweiterungen hatten wir da? Was passiert mit dem Satz von Kuhn?
Zufallszüge: Kuhn gilt, man muss mit Erwartungswert rechnen
Prof. Nebel wollte wissen, woher der Zufall kommt, nach Klarstellen der Frage habe ich was von exogen gesagt, also von außen und er war zufrieden
Simultane Züge: Kuhn gilt nicht mehr, Gegenbeispiel: Matching Pennies

Prof. Nebel sieht Prof. Burgard an und meint: Ach ja, da hatten wir was neues in der VL, das muss sie dir nun erklären:

Was ist denn Mechanismus Design?

- Teilbereich der Spieltheorie, bei dem es um die Synthese von Spielen geht.
- Ziel ist es, eine soziale Entscheidung als Lösung eines Spiels zu implementieren, also Spiele so zu definieren, dass deren Lösungen (Gleichgewichte) gerade die gewünschten Ausgänge sind.

Beispiele für soziale Entscheidungen sind Wahlen, Auktionen, Festlegungen von Policies.

- Nicht immer werden die Präferenzen der beteiligten Personen ehrlich angegeben
- Mechanismusdesign implementiert die Bestimmung der sozialen Entscheidungen in einer strategischen Umgebung, in der die Präferenzen der Teilnehmer nicht öffentlich sind.

Was ist eine soziale Entscheidung?
Aggregiert alle Präferenzen der Wähler in einzelne soziale Entscheidung:

- Eine soziale Wohlfahrtsfunktion aggregiert alle Präferenzen von allen Wählen in eine totale soziale Ordnung der Kandidaten: $F : L^n \to L$
- Eine soziale Entscheidungsfunktion aggregiert alle Präferenzen von allen Wählen in eine einzelne soziale Entscheidung für einen Kandidaten: $F : L^n \to A$
- L ist die Menge der linearen Ordnungen auf A, sprich alle Elemente können paarweise verglichen werden

Wie lautet der Satz von Arrow?
Jede soziale Wohlfahrtsfunktion über einer Menge von mehr als zwei Alternativen, die totale Einstimmigkeit und UIA erfüllt, ist diktatorisch.

Was ist ein Diktator?
Die soziale Präferenz ist immer die des Diktators, ungeachtet der Präferenzen aller anderen Wähler

Was heißt strategisch manipulierbar?
Eine soziale Entscheidungsfunktion ist manipulierbar, wenn ein Wähler i, der b vor a präferiert, b erzwingen kann, wenn er statt seiner wahren Präferenz $<_i$ eine davon verschiedene Präferenz $<_i'$ angibt.
f heißt anreizkompatibel, wenn f nicht manipulierbar ist.

Was besagt das Erweiterungslemma?
Falls f eine anreizkompatible, surjektive und nicht-diktatorische soziale Entscheidungsfunktion ist, so ist ihre Erweiterung F eine soziale Wohlfahrtsfunktion, die Einstimmigkeit, Unabhängigkeit von irrelevanten Alternativen und nicht-diktatorische Entscheidung erfüllt.
Satz von Gibbard-Satterthwaite ist das Analogon zum Satz von Arrow für soziale Entscheidungsfunktionen

Was folgt aus dem Satz von Gibbart & Satterthwaite?
Er scheint alle Hoffnungen zu zerstören, man könnte anreizkompatible soziale Entscheidungsfunktionen entwerfen.

Wie konnten wir das umgehen?
Man kann das Modell ändern. Es gibt zwei gängige Möglichkeiten, die Hinzunahme von Geld und die Einschränkung der zulässigen Präferenzrelation
Bei Prof. Nebel sollte man auf jeden Fall Ahnung haben wovon man spricht und/oder die Definitionen sauber lernen, er fragt da gerne mal nach (nicht immer, aber wenn es ihm gerade in

den Sinn kommt, z.B. ob die Definition bei strikter Dominanz für alle Strategien gilt). Es ist in jeder Hinsicht sehr freundlich und geduldig, stellt seine Fragen ruhig und präzise und erklärt sie bei Unverständnis gerne noch einmal.

Advanced AI Techniques:

Was ist denn der Unterschied von Reinforcement Learning zu Spieltheorie?
Wir haben nur einen Agenten

Was ist denn RL?
Man beobachtet, wie der Agent durch Erfolg und Misserfolg mit Belohnung durch die Interaktion mit der Umgebung lernen kann.

Wieso ist es so wichtig? Was hat man denn hier für Annahmen, die man in der Spieltheorie nicht hat? Wieso ist RL besser als Minimax?
In der Spieltheorie nimmt man an, dass der Gegner optimal spielt. Im RL kann man lernen die Schwächen eines suboptimalen Gegners auch auszunutzen.

Was suchen wir im RL? Was ist gegeben?
Man versucht anhand der beobachteten Belohnungen die optimale Strategie zu finden. Die optimale Strategie maximiert den erwarteten Gesamtgewinn.

Was ist Bellmann? *(also so hat er nicht gefragt, aber an irgendeiner Stelle habe ich Bellman erklärt)*
Wertefunktionen im RL erfüllen bestimmte rekursive Bedingungen. Die Gleichung beschreibt eine Konsistenz zwischen dem Zustand und allen möglichen Nachfolgezuständen und einer Strategie.

Was ist DP?
Man benutzt die Wertefunktion um nach guten Strategien zu suchen.
Dann habe ich lang und breit Policy evaluation und Policy Improvement erklärt, wie beides funktioniert und daran dann die Policy iteration mit diesen 2 Schritten indem man Bellman als Zuweisung benutzt also einen vollständigen policy evaluation Schritt hingeschrieben, dann denn policy improvement Schritt und erklärt, dass das endet wenn die Policy stabil ist. Auch dass der Policy evaluation -Schritt mehrere sweeps durch den Zustandsraum braucht und das dann bei der value iteration abgeschnitten wird, weil die optimale Policy oft vor der Konvergenz der Wertefunktion erreicht ist.

Was ist der Unterschied zu MC?
Bei MC: Kein bootstrapping, also Verletzung der Markov-Eigenschaft nicht so schlimm, aktualisiert Werte erst am Ende der Episode, also nur für episodische tasks geeignet, braucht kein vollständiges Umgebungsmodell, lernt also direkt aus der Erfahrung, kann man einfach und effizient auf Untermengen anwenden

Was ist GPI?
Generelle Idee zweier interagierender Prozesse, die sich der Policy und der Wertefunktion annähern. Der eine Prozess nimmt die Strategie als gegeben und ändert die Wertefunktion, dass sie zur policy passt, der andere Prozess nimmt die Wertefunktion als gegeben und verbessert die Strategie damit. Die Prozesse konkurrieren in dem Sinne, dass sie immer die Basis des anderen ändern, aber insgesamt arbeiten sie zusammen an einer gemeinsamen Lösung. Wenn sie sich gegenseitig nicht mehr ändern, sind sie optimal.

(Anm. Die korrekte Übersetzung von Policy lautet Taktik, ich habe aber oft Strategie oder Policy gesagt, also ein gewisses "switchen" zwischen den Sprachen stellt kein Problem dar)

Eine letzte Frage: Was ist das exploration-exploitation Dilemma?
Man kann nicht immer ausbeuten, man kann nicht immer explorieren, aber beides muss getan werden um ein optimales Ergebnis zu erzielen. Das Dilemma ist nun, dass man beides nicht mit einem Schritt durchführen kann.

(Glaube diese beiden "wirklich" letzten Fragen waren die finale Entscheidung für die bessere Note)
In der realen Welt haben wir ja nicht immer solche Eigenschaften wie im RL gefordert ist?
Nein, da kennt der Agent den Zustand in dem er sich befindet nicht. Das nennt sich dann POMDP.

Und löst man das nun anders als vorher, oder wie macht man das?
Nein, nicht arg viel anders, denn man kann POMDP für einen Zustandsraum auf die Lösung eines MDPs reduzieren, man verwendet dann den zugehörigen Glaubenszustand. Allerdings wurden POMDPs bisher nur auf kleinen Zustandsräumen angewendet, denn die Anzahl der linearen Bedingungen wächst mit jeder Iteration exponentiell.
(Das war Gott sei Dank alles, was er wissen wollte, den mehr hätte ich nicht sagen können)

Und wie heißt dann die Gleichung?
(geraten, und zwar gut ;-))
Das ist die Recursive Bayes Filtering Gleichung.
Gleichung aufgeschrieben und kurz erklärt (Sensor-Modell, Aktionsmodel, dass man Zustand x in dynamischen System schätzen möchte)

Bei Prof. Burgard sollte man sich auch auf jeden Fall ein bisschen über den Bezug von RL und Spieltheorie kümmern (also zumindest bei dieser Kombination oder bei der Prüfung in AAIT) und die ganzen Prinzipien ebenfalls gut verstanden haben sowie die Definitionen und Formeln sicher können. Auch er stellt seine Fragen ruhig, aber manchmal war mir nicht ganz klar worauf er hinaus wollte, aber er gibt gerne eine kleine Hilfestellung wenn man nicht direkt antworten kann. Für die 1,0 reicht es nicht, nur den RL-Teil zu lernen, POMDP und Recursive Bayes Filtering sollte man sich dazu zumindest mal kurz angesehen haben. Was auf jeden Fall sitzen muss ist Bellmann, alle Backup-Schritte von den einzelnen Verfahren und alle anderen Grundbegriffe wie GPI oder Markov-Eigenschaft.

Zur Vorbereitung empfehle ich den regelmäßigen Besuch der Vorlesung und das gründliche Durcharbeiten der Vorlesungs-Aufzeichnungen und des Skripts in Spieltheorie. Wichtige Definitionen und Formeln muss man auf jeden Fall sicher auswendig können. Die Übungsblätter habe ich mir zur Vorbereitung auf diese Prüfung nicht ein einziges Mal angesehen, dafür aber andere Prüfungsprotokolle und mir auch selber alle möglichen Fragen ausgedacht, die drankommen könnten und sie vor allem auch beantwortet. Ohne einen gewissen Lernaufwand ist bei dieser Kombination eine gute Note sicher nicht drin. Außerdem legt schon die Länge des Protokolls dar, dass ich doch ziemlich viel zu sagen hatte, was den Vorteil bietet, dass dann weniger Fragen kommen. Mit genauerem Nachfragen zu dem Erzählten muss man auf jeden Fall immer rechnen.
Ach ja, es wurde mir gesagt, dass ich Mut zur Lücke bewiesen habe, weil ich zu den LCP-Ungleichungen überhaupt nichts vernünftiges zustande gebracht habe, aber habe dennoch eine 1,0 bekommen, also während der Prüfung von solchen Kleinigkeiten nicht aus der Ruhe bringen lassen.

Ich wünsche allen viel Glück bei Ihrer Prüfung!