

MODELLIERUNG VON BEWÄLTIGUNGSVERHALTEN MIT MERKMAL-WERTE-RELATIONEN

Matthias Wolff, Werner Meyer, Ronald Römer

*Brandenburgische Technische Universität
matthias.wolff@b-tu.de*

Kurzfassung: Kognitive Agenten sind nach dem aktuellen Stand der Technik in der Lage, in einer gegebenen Umwelt und mit einem gegebenen Repertoire von Aktionen eine Strategie zum Erreichen eines definierten Ziels zu erlernen und dieser Strategie entsprechend eine optimale Handlung – also eine zielgerichtete Folge von Aktionen – auszuführen. Es existiert jedoch bislang kein uns bekanntes Verfahren für Fälle, in denen das vordefinierte Aktionsrepertoire nicht zum Erreichen des Handlungsziels ausreicht. In einer solchen Konstellation wird ein Systemverhalten benötigt, das mit einer Bewältigungsstrategie (*coping*) vergleichbar ist. Wir definieren Bewältigungsverhalten technisch dadurch, dass sich der kognitive Agent bei Bedarf selbständig neue Aktionsmöglichkeiten erschließt und damit erfolgreich zu seinem (mit dem vordefinierten Aktionsrepertoire unerreichbaren) Ziel kommt. In unserem Beitrag stellen wir einen Algorithmus zur Realisierung einer solchen Art von autoplastischem Coping vor und diskutieren diesen an einem Minimalbeispiel.

1 Einführung

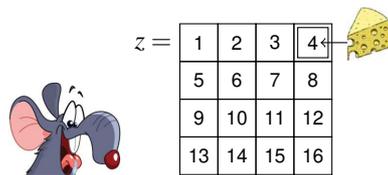
1.1 Bewältigungsverhalten

Den entwicklungsgeschichtlich ältesten Mechanismus der Verhaltenssteuerung nennt die Psychologie Instinkt. Dieser ist ein prärationaler Mechanismus, der ohne Einsicht und Erfahrung adaptiv auf einen Antrieb reagiert. Gelingt dem Organismus die Adaption, kann der Antrieb gelöscht und die Kontrolle an übergeordnete Instanzen abgegeben werden. Man unterscheidet zwei Klassen: Die erste Klasse zielt auf eine Antriebslöschung bei einer Endhandlung ab, hier fehlt jedoch die Rückmeldung, ob eine Endsituation eingetreten ist, in der das Bedürfnis auch tatsächlich befriedigt worden ist. Die zweite Klasse löscht den Antrieb erst durch das Erreichen einer Endsituation. Mit steigendem Evolutionsniveau verschiebt sich das Ziel der Instinkthandlung hin zur zweiten Klasse. Der nächste Entwicklungsschritt wurde in Situationen notwendig, in denen die Umgebung den Zugang zum Ziel verweigert und somit die Endhandlung blockiert. Die Fähigkeit, solche Zugangsbarrieren zu überwinden, wird Bewältigungsverhalten (*coping*) genannt. Dabei steht der Coping-Apparat vor dem Problem, dass sich die vom Bauplan vorgesehenen sensorischen Schemata und motorischen Radikale als untauglich erweisen [2].

1.2 Ein Minimalbeispiel

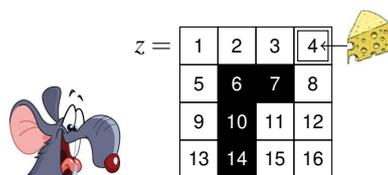
Eine Maus (kognitiver Agent) wird auf ein zufällig gewähltes Feld eines Labyrinths (Welt) gesetzt. Die Aufgabe der Maus (Handlungsziel, Endsituation) ist es, den in der rechten oberen

Ecke platzierten Käse zu finden. Dazu kann sie eine Folge von Aktionen (Handlung) ausführen. Die Beispielmaus soll ein Repertoire von nur zwei Aktionen haben: nach oben gehen (N) und nach rechts gehen (E). Die Maus „beherrscht“ das Labyrinth, wenn sie den Käse von *jedem* zufällig gewählten Anfangspunkt erreicht und zwar möglichst auf dem kürzesten Weg (optimale zielführende Handlung). Wir beginnen mit einer Welt ohne Hindernisse:

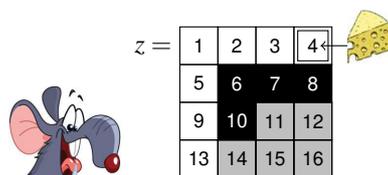


(1)

Die Felder sind mit z bezeichnet und durchnummeriert. Der Käse liegt im Feld $z = 4$. Die Maus kann den Käse auch mit ihrem beschränkten Aktionsrepertoire von jedem beliebigen Feld aus erreichen. Eine optimale Handlungsstrategie kann automatisch mit dem bekannten *Q-Learning*-Verfahren [8] erlernt werden. Nun fügen wir ein Hindernis ins Labyrinth ein und vereinbaren, dass die Maus nie auf einem Hindernis eingesetzt wird (in diesem Fall also nicht auf den Feldern $z \in \{6, 7, 10, 14\}$):



Die Maus muss jetzt zwar möglicherweise ihre erlernte Strategie – wiederum mit Hilfe des *Q-Learning*-Verfahrens – anpassen, kann das Labyrinth aber dann wieder beherrschen. Ein solches Systemverhalten kognitiver Agenten wird als Adaption bezeichnet. Zuletzt ändern wir das Hindernis wie folgt:



(2)

Da die Maus nur nach oben oder rechts gehen kann, reicht nun Adaption alleine nicht mehr aus, um das Labyrinth zu beherrschen. Es gibt Anfangspunkte (grau hinterlegt), von denen aus keine Lösung gefunden werden *kann*, weil dazu weitere Aktionen, nämlich nach unten (S) und nach links (W) gehen, nötig wären. Autoplastisches Bewältigungsverhalten erschließt dem Agenten in Fällen, in denen die Adaption versagt, *neue Aktionen* zur Beherrschung des Problems.

1.3 Formales Modell

Wir definieren einen *kognitiven Agenten* formal als sequenziellen Automaten, der in jedem Zeitschritt k in Abhängigkeit von einem inneren Zustand $c \in C$ und einer Beobachtung $o \in O$ (I) einen Zustandsübergang $c \rightarrow c'$ und (II) eine Aktion $a \in A$ ausführt. Dabei stehen $C = \{c_1, c_2, \dots\}$ für die Menge möglicher Agentenzustände, $O = \{o_1, o_2, \dots\}$ für die Menge möglicher Beobachtungen und $A = \{a_1, a_2, \dots\}$ für die Menge möglicher Aktionen. Wir bezeichnen eine Folge

$$\mathbf{a} = (a^1, \dots, a^K) \quad a^k \in A$$

als *endliche Handlung* aus K Aktionen. Das Handlungsziel ist bei Eintreten einer bestimmten Beobachtung $o_F \in O$ erreicht.

Die Aktionen des Agenten sind auf eine *Umwelt* bezogen, in welche die Aktionen eingegeben und aus welcher die Beobachtungen ausgelesen werden. Wir definieren die Umwelt formal als gewichteten endlichen Transduktor

$$\mathcal{W} = \{Z, I, \{z_F\}, A, O, S, w\} \quad (3)$$

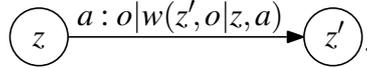
mit Gewichtshalbring $S = (\mathbb{K}, \oplus, \otimes, \bar{0}, \bar{1})$ und mit einer endlichen Menge Z von Zuständen, von denen genau ein Zustand $z_F \in Z$ das Handlungsziel des Agenten in der Welt und damit den einzigen Schlusszustand des Weltautomaten repräsentiert: $F = \{z_F\}$. Die Welt kann am Anfang in jedem Zustand $z \in I \subseteq Z$ sein, solange dieser kein „Hindernis“ ist.¹ Ihre Verhaltensfunktion

$$w : Z \times A \times O \times Z \rightarrow \mathbb{K}$$

ordnet jedem 4-Tupel (z, a, o, z') aus aktuellem Weltzustand $z \in Z$, vom Agenten eingegebener Aktion $a \in A$, vom Agenten beobachtbarer Reaktion $o \in O$ und neuem Weltzustand $z' \in Z$ ein Gewicht $w(z', o|z, a)$ zu. Eine Aktion a heißt wirkungslos, wenn sie *nie* zu einer Zustandsänderung in der Welt führt:

$$\forall z_i, z_j \in I, z_i \neq z_j : |\mathcal{P}(z_i, a, z_j)| = 0,$$

wobei $\mathcal{P}(z_i, a, z_j)$ die Menge aller Übergänge zwischen zwei Zuständen z_i und z_j darstellt, welche die Aktion a akzeptieren. Ein Zustandsübergang kann grafisch wie folgt dargestellt werden:

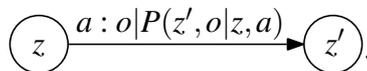


Wir bezeichnen das Beobachten einer Weltreaktion durch den Agenten als *Perzeption* und das Eingeben einer Aktion in die Welt als *Aktion*. Agent und Welt sind in Form eines *Perzeptions-Aktions-Zyklus* [4, S. 3] miteinander verbunden.

Wir unterscheiden zwischen *deterministischen Welten* mit $S = (\{0, 1\}, \vee, \wedge, 0, 1)$ (BOOLEscher Halbring), Übergängen der Form



und einem deterministischen Weltautomaten sowie *stochastischen Welten* mit $S = (\mathbb{R}^+, +, \cdot, 0, 1)$ (Wahrscheinlichkeitshalbring) und Übergängen der Form



In deterministischen Welten ist also der Zustandsübergang bei jeder eingegebenen Aktion eindeutig und der Agent kann den neuen Weltzustand direkt beobachten. Bild 1 zeigt zwei Beispiele für deterministische Weltautomaten. Zur besseren Lesbarkeit wurden die Ausgabezeichen der Übergänge weggelassen, da sie nach (4) immer gleich dem Folgezustand sind. Beide Automaten repräsentieren je ein Labyrinth. Die möglichen Handlungen $A = \{N, E, S, W\}$ sind Bewegungen. Jeder Zustand der Welt hat eine mit δ beschriftete Rückfallschleife, die jede Aktion akzeptiert, falls kein besser passender Übergang existiert. Dies modelliert den Umstand, dass ein kognitiver Agent in jeder vernünftigen Welt jederzeit alles tun kann, gegebenenfalls jedoch ohne irgendeine Wirkung zu erzielen. Rückfallübergänge sind Voraussetzung für den Coping-Algorithmus.

¹Die Definition von Hindernissen als Weltzustände ist formal nicht nötig. Wir bleiben hier jedoch dabei, um kompatibel mit dem einführenden Beispiel zu sein.

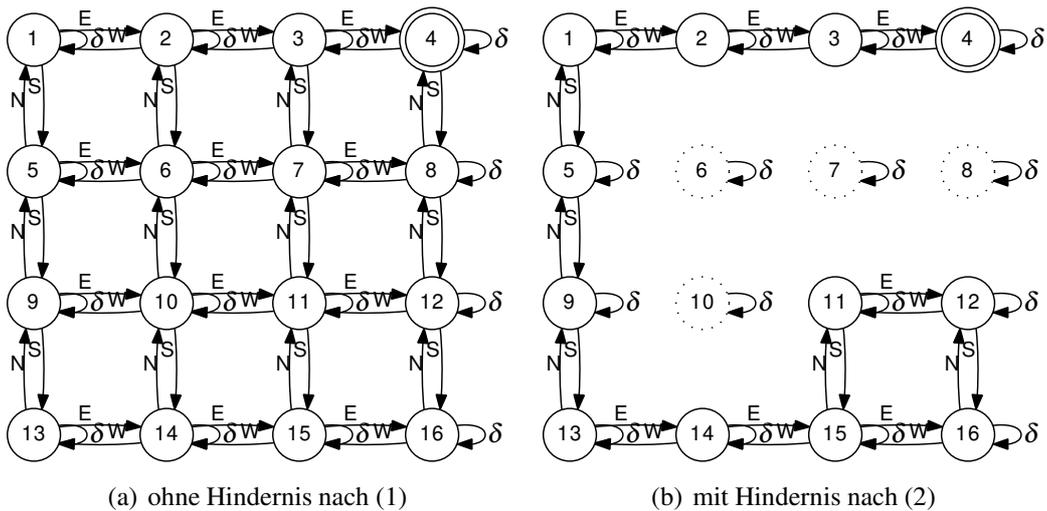


Bild 1 - Labyrinth-Welten (endliche, deterministische Aktionsakzeptoren).

Algorithmus 1 - Naives Coping

- | | |
|--|--|
| 1: $\tilde{A} \leftarrow \emptyset, \mathcal{I} \leftarrow \emptyset$
2: repeat
3: $\tilde{A} \leftarrow \tilde{A} \cup \{a_{ \tilde{A} +1}\}$
4: Q-Learning
5: for all $e = (z, a, z') \in Q$ do
6: $\mathcal{I} \leftarrow \mathcal{I} \cup e$
7: end for
8: until $\forall z \in \mathcal{I} : \mathcal{P}(z, z_F) \geq 1$ | ▷ Aktionsrepertoire, inneres Weltmodell

▷ füge neue Aktion hinzu
▷ führe <i>Q-Learning</i> aus
▷ Zustandsübergänge beim <i>Q-Learning</i>
▷ erweitere inneres Weltmodell

▷ Lösung von allen Anfangszuständen gefunden |
|--|--|
-

2 Naives Coping

Wir betrachten im Folgenden nur deterministische Welten. Der Agent ermittelt, wie oben beschrieben, mit Hilfe des *Q-Learning*-Verfahrens eine optimale Strategie zum Erreichen des Zielzustands $z_F \in Z$ in der Welt von jedem anderen Anfangszustand $z \in I \setminus \{z_F\} \subseteq Z$ aus. Dabei zeichnet er alle beobachteten Weltzustände und alle durchgeführten Zustandsübergänge $e = (z, a, z') \in Z \times A \times Z$ in einem *inneren Weltmodell* auf (siehe Bild 2). Für deterministische Welten kann dieses durch einen ungewichteten endlichen Akzeptor

$$\mathcal{I} = (Z, I, \{z_F\}, A)$$

ausgedrückt werden (vgl. Gl. 3).² Der Agent beherrscht die Welt, wenn er das Ziel von jedem Anfangszustand aus erreichen kann. Im inneren Weltmodell muss dann also mindestens ein durchgehender Weg von jedem Zustand $z_i \in I \setminus \{z_F\}$ zum Schlusszustand z_F existieren:

$$\forall z_i \in I \setminus \{z_F\} : |\mathcal{P}(z_i, z_F)| \geq 1. \quad (5)$$

Wenn diese Bedingung für mindestens einen Anfangszustand nicht erfüllt ist, versucht der Agent, das Problem zu bewältigen. Da sein Aktionsrepertoire offensichtlich nicht ausreichend ist, müssen neue Aktionen „erfunden“ werden. Dazu fügt der Agent einfach neue Aktionen

²Für stochastische Welten wird ein stochastischer Automat verwendet, dessen Zustände Annahmen (*beliefs*) über Weltzustände sind.

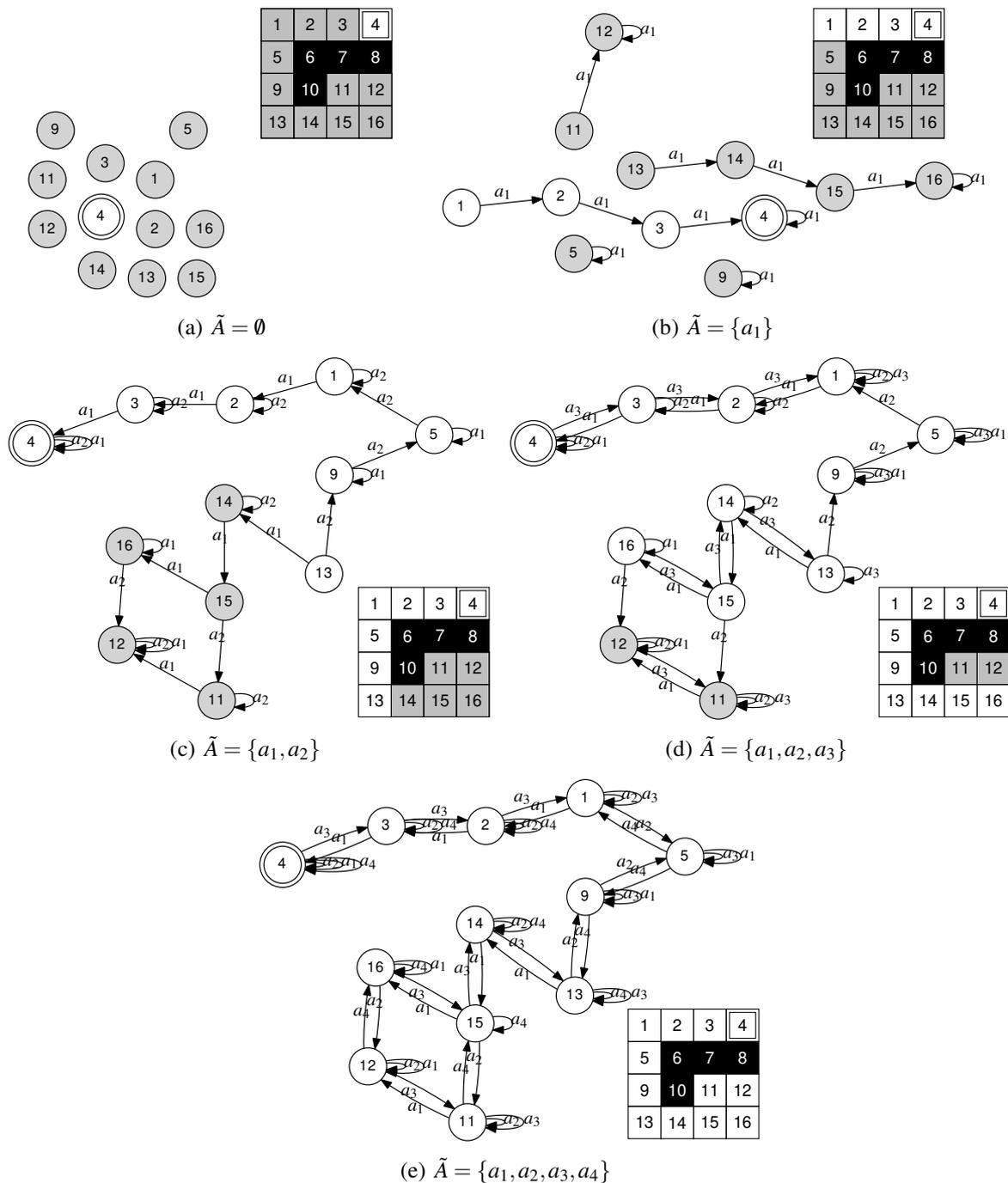


Bild 2 - Innere Weltmodelle \mathcal{I} des kognitiven Agenten nach *Q-Learning* in einer deterministischen Welt nach Bild 1(b) mit zunehmender Anzahl durch Coping erschlossener Aktionen. Von den grau hinterlegten Zuständen aus führt keine mögliche Handlung zum Ziel z_4 . Die inneren Weltmodelle wurden automatisch ohne Verwendung topologischer Informationen aus Bild 1(b) mit dem Werkzeug *fdp* aus dem GraphViz-Paket [1] gezeichnet. Einzige Randbedingungen waren möglichst gleiche Kantenlängen und eine möglichst kompakte Anordnung der Zustände.

zu seinem Repertoire hinzu:

$$A \leftarrow A \cup \{a_{|A|+1}\}.$$

Das ist formal möglich, da alle möglichen Aktionen elementar sind und die Welt in *jedem* Zustand *jede* Aktion akzeptiert (siehe Erläuterung der Rückfallübergänge oben). Nach Erweiterung des Aktionsrepertoires und erneutem *Q-Learning* wird wieder geprüft, ob Bedingung (5) erfüllt ist. Wenn nicht, werden solange neue Aktionen „erfunden“, bis dies der Fall ist. Algorithmus 1 stellt dieses Vorgehen zusammenhängend dar. Bild 2 zeigt das wachsende Aktionsrepertoire und die jeweiligen inneren Weltmodelle für Iterationen des Coping-Algorithmus in der Welt nach Bild 1(b). Dieses naive Coping-Modell hat zwei Nachteile:

- Falls überhaupt keine Lösung existiert, terminiert der Algorithmus nicht.
- Die Modellierung von Aktionen als Mengenelemente scheint zu trivial für reale Welten.

Immerhin wird aber das Grundproblem, also bei Versagen der Adaption neue Aktionen zu erschließen, gelöst. Es ist sogar möglich, jeder gefundenen Aktion eine Semantik zuzuordnen:

1. Betrachtet man den Weltzustand z vor einer Aktion als Vorbedingung (*antecedent*), den Zustand z' nach der Aktion als Konsequenz (*consequence*) und die Aktion a selbst als Verhalten (*behavior*), ist die Menge $\{(z, a_i, z')\}$ die *extensionale* Semantik einer bestimmten Aktion a_i im Sinne des SKINNERSchen ABC-Schemas [7].
2. Im Beispiel sind die Aktionen a_1, a_3 sowie a_2, a_4 jeweils antagonistisch. Wenn ein Übergang $z \xrightarrow{a_1} z'$ bzw. $z \xrightarrow{a_2} z'$ existiert, so existiert immer auch ein Übergang $z' \xrightarrow{a_3} z$ bzw. $z' \xrightarrow{a_4} z$.
3. Außerdem sind im Beispiel die Aktionspaare $\{a_1, a_3\}$ und $\{a_2, a_4\}$ orthogonal in dem Sinne, dass zwei Zuständen entweder durch das antagonistische Paar $\{a_1, a_3\}$ oder das antagonistische Paar $\{a_2, a_4\}$ (oder keines von beiden) verbunden sind.

Die letzten beiden Punkte beschreiben Aspekte der *intensionalen* Semantik der Aktionen, also eine innere Struktur der Aktionsmenge. Sie gelten zwar so nur für das Beispiel, man kann aber erwarten, dass sich ähnliche Befunde auch in inneren Modellen über andere Welten ergeben.

3 Coping mit intensionaler Aktionssemantik

In realistischen Agent-Welt-Szenarien sind Aktionen weder im Allgemeinen elementar, noch können sie in abstrakter, symbolischer Form in die Welt „eingegeben“ werden. Genauso wenig kann der Weltzustand symbolisch beobachtet werden. Die Kommunikation zwischen Agent und Welt erfolgt *immer* über Signale. Der Agent hat also zusätzlich das Problem, adäquate Aktuatorsignale zu synthetisieren und Sensorsignale, die er aus der Welt aufnimmt und die den Zustand der Welt ausdrücken, zu analysieren. Hinsichtlich des Bewältigungsverhaltens bedeutet dies, dass der Agent nur dann neue Aktionen „erfinden“ und erproben kann, wenn er auch in der Lage ist, adäquate Aktuatorsignale zu synthetisieren. Da nicht triviale Signale in realen Welten innere Strukturen besitzen [9], benötigt der Agent ein Modell, das zwischen Aktionssemantik und Aktuatorstruktur übersetzt.³ Wir unterstellen, dass Teile des Aktuatorsignals Teilbedeutungen tragen und die Gesamtbedeutung von den Teilbedeutungen abhängt (FREGE-Prinzip

³Das gleiche gilt natürlich auch für die Übersetzung zwischen Sensorsignalstruktur und Perzeptionssemantik, welche hier aber nicht betrachtet wird.

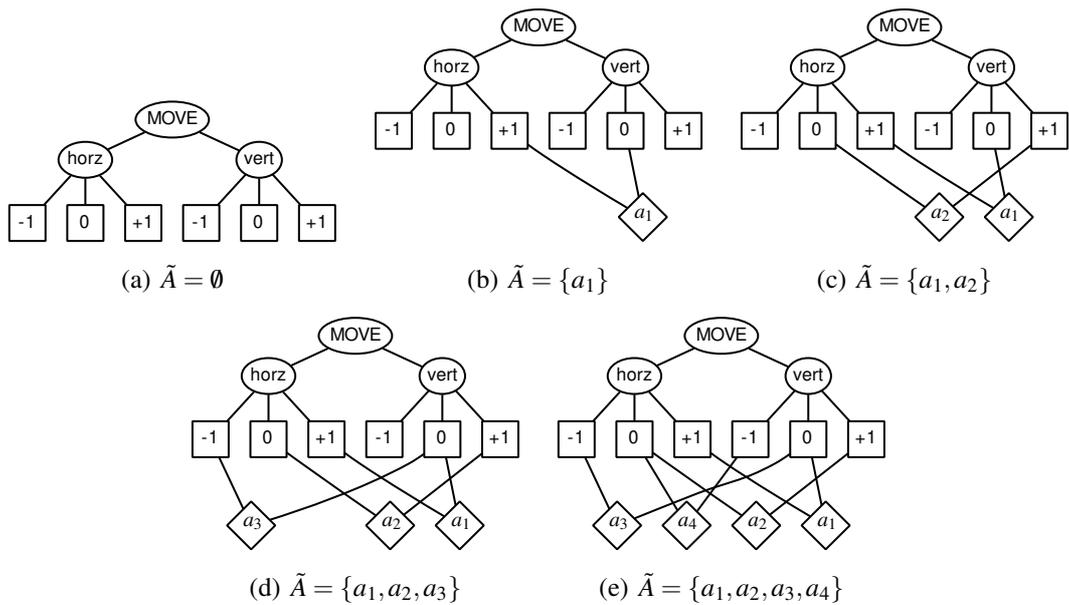
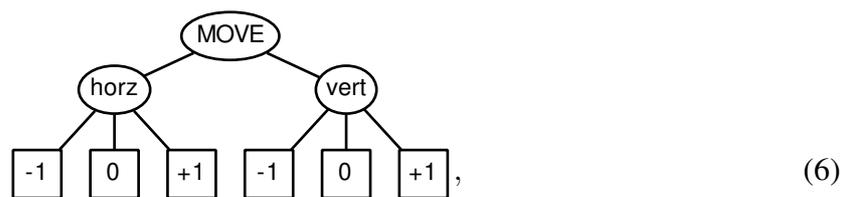


Bild 3 - Intensionale Aktionssemantiken mit zunehmender Anzahl durch Coping erschlossener Aktionen in der Welt nach Bild 1(b). Die inneren Weltmodelle entsprechen Bild 2. Die einzelnen Aktionssemantiken sind: $a_1 = \{h=+1, v=0\}$, $a_2 = \{h=0, v=+1\}$, $a_3 = \{h=-1, v=0\}$ und $a_4 = \{h=0, v=-1\}$. Alle weiteren erschließbaren Aktionen (diagonale und keine Bewegung) sind in der Welt wirkungslos.

[3]). In unserem Beispiel können die Teilbedeutungen eine „horizontale“ und eine „vertikale“ Bewegungskomponente sein. Abgesehen davon, dass diese Wahl in unseren Labyrinth-Welten offensichtlich plausibel ist, hat die Analyse der inneren Weltmodelle nach naivem Coping ja auch zwei orthogonale Paare jeweils antagonistischer Aktionen gefunden. Die Zuordnung der Paare zur „horizontalen“ oder „vertikalen“ Richtung ist aus Sicht des Agenten willkürlich. Eine Bewegungsaktion kann nun durch folgende Mermal-Werte-Relation dargestellt werden:



wobei Ellipsen Merkmale und Quadrate Werte kennzeichnen. Die horizontale und vertikale Bewegungskomponente stehen im Semantikmodell ausdrücklich *nicht* in einer bestimmten syntagmatischen Beziehung zueinander. Strukturen nach (6) werden als beschriftete Halbordnungen (*labeled partial orders*, LPO) bezeichnet. In [5, 6] zeigen wir, dass diese in Zeichenfolgen und damit in syntagmatische Signalstrukturen übersetzt werden können.

Intensionale Semantiken der Bewegungsaktionen können mit (6) durch Kombinationen je eines Werts für die horizontale und vertikale Komponente definiert werden. Der Coping-Algorithmus (Alg. 1) erzeugt im dritten Schritt also nun eine neue Aktion in Form einer neuen Wert-Wert-Relation und nicht mehr als nicht näher definiertes Element einer Aktionsmenge. Bild 3 zeigt den Fortschritt des Copings basierend auf dem intensionalen Semantikmodell (6). Die Reihenfolge der neu aufgestellten Wert-Wert-Relationen ist durch den Algorithmus nicht definiert. Die angegebene Reihenfolge entspricht der des Beispiels für das naive Coping. Der modifizierte Coping-Algorithmus endet mit Sicherheit, wenn das Semantikmodell endlich viele Merkmale

und Werte umfasst (im Beispiel spätestens nach Generierung aller neun möglichen Wert-Wert-Relationen).

4 Zusammenfassung und Ausblick

Wir haben Bewältigungsverhalten für kognitive Agenten mathematisch modelliert und einen entsprechenden Coping-Algorithmus angegeben. In seiner naiven Form löst dieser zwar das Grundproblem, geht aber vom unrealistischen Konzept ausschließlich elementarer Aktionen aus. Eine erweiterte Version ist auch bei einer realistischen Agent-Welt-Kommunikation über Signale anwendbar. In diesem Fall wird ein intensionales Semantikmodell (Merkmal-Werte-Relation) für Aktionen benötigt. Ist dieses nicht im Vorhinein bekannt, scheint es prinzipiell auch aus den inneren Weltmodellen des Agenten beim naiven Coping ableitbar zu sein. Wir werden noch untersuchen, ob und wie intensionale Semantiken mittels naiven Copings „erlernt“ werden können. Hauptprobleme dabei sind, automatisch Strukturen in der Aktionsmenge zu finden und geeignete Aktionen, die ja in realistischen Welten nicht elementar sein können, für das naive Coping zu erzeugen. Das letzte Problem könnte vielleicht über Imitation des Verhaltens anderer Agenten gelöst werden.

Literatur

- [1] *GraphViz - Graph Visualization Software*. Online: graphviz.org. Abgerufen: 2.2.2015.
- [2] BISCHOF, N.: *Psychologie, ein Grundkurs für Anspruchsvolle*. Verlag Kohlhammer, 2009. 2. Auflage.
- [3] FREGE, G.: *Über Sinn und Bedeutung*. Zeitschrift für Philosophie und philosophische Kritik, (NF 100):25–50, 1892.
- [4] HAYKIN, S.: *Cognitive Dynamic Systems: Perception-action Cycle, Radar and Radio*. Cambridge University Press, 2012.
- [5] HUBER, M. und R. RÖMER: *Modellierung des Semantik-Syntax-Grenzübergangs kognitiver Systeme am Beispiel des „Mouse-Maze“-Problems*. In diesem Tagungsband, 2015.
- [6] RÖMER, R. und M. WOLFF: *Konzeption eines kognitiven Systems für den experimentellen Einsatz in Forschung und Lehre*. In diesem Tagungsband, 2015.
- [7] SKINNER, B. F.: *Verbal Behaviour*. Copley Publishing Group, Acton, 1957.
- [8] WATKINS, C.: *Learning from Delayed Rewards*. Doktorarbeit, Cambridge University, Cambridge, England, 1989.
- [9] WOLFF, M. und R. HOFFMANN: *An Approach to Intelligent Signal Processing..* In: A. ESPOSITO ET AL. (Hrsg.): *Behavioral Cognitive Systems*, Bd. 7403 d. Reihe LNCS, S. 1–18. Springer-Verlag, Berlin Heidelberg, 2012.