

Kognitive Robotik – Herausforderungen an unser Verständnis natürlicher Umgebungen

Cognitive Robotics – challenges for our understanding of natural environments

Marc Toussaint, Tobias Lang, Nikolay Jetchev, FU Berlin

Wir diskutieren einen Ansatz, der kognitive Robotik als die Erweiterung der Methoden der Robotik – insb. Lernen, Planen und Regelung – auf “äußere Freiheitsgrade” versteht. Damit verschiebt sich der Fokus: Weg von Gelenkwinkeln, Vektorräumen und Gauß’schen Verteilungen, hin zu den Objekten und der Struktur der Umwelt. Letztere können wir nur schwer formalisieren und in geeignete Repräsentationen und Priors übersetzen, mit denen effizientes Lernen und Planen möglich wird. Es wird deutlich, welche theoretischen Probleme sich hinter dem Ziel autonomer Systeme verbergen, die durch intelligente Exploration und Verallgemeinerung ihre Umwelt zu verstehen lernen und die gelernten Modelle zur Handlungsplanung nutzen. Die momentan diskutierte Integration von Logik, Geometrie und Wahrscheinlichkeiten – und damit die Überbrückung der klassischen Disziplinbarrieren zwischen Robotik, Künstlicher Intelligenz und statistischer Lerntheorie – ist eine der zwangsläufigen Herausforderungen der kognitiven Robotik. In diesem Kontext skizzieren wir eigene Beiträge zum relationalen Reinforcement-Lernen, zur Exploration und dem Symbol-Lernen.

Schlagwörter: Autonomes Lernen, relationales Reinforcement-Lernen, intelligente Exploration und Verallgemeinerung, Inferenz

Keywords: autonomous learning, relational reinforcement learning, intelligent exploration and generalization, inference

1 Einleitung

Anfang des 20. Jahrhunderts führte Wolfgang Köhler seine „Intelligenzprüfungen an Menschenaffen“ (Köhler, 1917) durch. Ein eindruckliches Experiment ist in Abb. 1 illustriert. In einem Käfig hängt eine Banane an der Decke, zu hoch für eine Schimpansin. Das Tier versucht mehrmals vergeblich an sie heranzukommen und springt immer wieder hoch. Schließlich gibt es auf und setzt sich in die Ecke. In einer anderen Ecke des Käfigs steht eine Holzbox. Köhler beschreibt, dass nach einiger Zeit der Blick der Schimpansin zwischen Banane und Box wechselt, sie dann zielgerichtet zur Holzbox läuft, diese unter die Banane zieht, darauf steigt und zugreift.

Was genau geht in dem Tier vor, wenn es in der Ecke sitzt und abwechselnd Banane und Holzbox anvisiert? Die Psychologie und die Künstliche Intelligenz formu-

lieren verschiedene Modelle zielgerichteten Verhaltens¹. Diese Modelle sind im Prinzip sehr allgemeingültig, in der Praxis hängen sie jedoch fundamental von der Wahl der Zustandsbeschreibung ab: Welche Repräsentationen (und *Priors*, siehe unten) nutzt das System, um intern zu planen oder Verhalten zu organisieren? Nimmt man im Falle des Stapelns von Holzboxen eine diskretisierte Repräsentation des Problems an, die sowohl Weltzustände als auch mögliche Aktionen deterministisch und symbolisch abstrahiert, so funktionieren KI Methoden effizient. Echte Welten sind jedoch nicht symbolisch, wir haben es mit kontinuierlichen Größen zu tun, mit Geometrie und Physik, mit Objekten und ihren Merk-

¹ D.h. planendes Verhalten, im Kontrast zu konditioniertem, durch wiederholte Belohnung habituiertem Verhalten. Siehe Niv et al. (2006) für eine Definition von zielgerichtetem vs. habituiertem Verhalten in der Psychologie. In der KI entspricht dies grob dem Unterschied zwischen modellbasiertem und modell-freiem Reinforcement-Lernen.

malen. Eine der großen und bisher ungelösten Herausforderungen ist es, solche fundamentalen Eigenschaften natürlicher Welten in geeignete Repräsentationen und Priors zu übersetzen.

Im Gegensatz zur KI ist für die Robotik der Umgang mit Geometrie und Physik alltäglich. Beispiele sind die klassischen Methoden der Pfadplanung unter Berücksichtigung der Geometrie, oder die Kraftregelung von Freiheitsgraden. Allerdings bringt die naive Integration solcher Methoden mit der Planung auf symbolischer Ebene unweigerlich Probleme: Sollte man zum Stapeln von Holzkisten geometrische Pfadplaner für jedes mögliche Objekt starten, ohne einen Zielort für das einzelne Objekt vorweg zu kennen? Wie kämen Pfadplaner auf "die Idee" eine Holzkiste zunächst heranzuziehen, um dann auf sie zu steigen? Während in symbolischen Repräsentationen die Kombinatorik voll exploriert werden kann, scheint das in geometrischen Repräsentationen ineffizient. Kaelbling and Lozano-Pérez (2011) diskutieren eindrücklich das Problem der Integration von geometrischer und symbolischer Planung. Die zentrale Herausforderung sind hier integrierte Methoden zum Umgang mit Geometrie, Logik (im Sinne der symbolischen Repräsentation) und Unsicherheit, die klassisch getrennt betrachtet werden.

Neben der Planung sind das Lernen und die Generalisierung weitere zentrale Punkte, die entscheidend von der Wahl der Repräsentation abhängen. Es ist nicht plausibel, dass der Affe bereits ein Modell besitzt, welches dezidiert die Nutzung einer Holzkiste zum Erreichen einer Banane enthält. Stattdessen muss er seine bisher gemachten Erfahrungen verallgemeinern. Das wirft das Problem der Generalisierung auf, beispielsweise die Nutzung von Objekten zu neuen Zwecken in vollkommen neuen Situationen. Aus der Lerntheorie wissen wir, dass Generalisierung Unsicherheit bedingt und durch den Prior bestimmt wird, der seinerseits durch die Wahl der Repräsentation induziert wird². Die Wahl der Repräsentation und der Umgang mit Unsicherheit sind also auch beim Lernen von zentraler Bedeutung.

Dies zeigt die starke Verzahnung der Forschungsfragen, die üblicherweise in der Robotik, der KI und beim maschinellen Lernen getrennt behandelt werden. In diesem Artikel wollen wir an einigen Beispielen aufzei-

² In der Bayesianischen Sicht auf ein lernendes System wird „Wissen“ über X durch die Bayes'sche Regel $P(X|D) \propto P(D|X)P(X)$ aktualisiert, wobei D die gesehenen Daten sind. Generalisierung bedeutet hier grob, dass selbst wenn die Daten nur wenige Komponenten von X direkt betreffen, auch das „Wissen“ bzgl. anderer Komponenten aktualisiert wird. Die Struktur des Priors bedingt dabei maßgeblich diese Struktur des Posteriors. Bei all dem ist die entscheidende Frage, wie das „Wissen“ $P(X)$ bzw. $P(X|D)$ repräsentiert wird – ob explizit oder implizit, als Verhaltensregeln oder logische Ausdrücke, als neuronales Netzwerk oder logistische Regression mit Kern-Merkmalen. Wenn wir im Folgenden also den Begriff der Repräsentation nutzen, so verweisen wir gleichzeitig auf die Struktur des Priors und der entsprechenden Generalisierung erlernten Wissens. Siehe auch Pearl (1988).



Bild 1: Eine „Intelligenzprüfung am Menschenaffen“ (Köhler, 1917)

gen, welche prinzipiellen Herausforderungen sich ergeben, will man die Methoden dieser drei Wissenschaften vereinen. Das Ziel sind künstliche Systeme, die Probleme autonom und auf Basis gelerntem Wissen lösen können, etwa wie es der Schimpansin mit der Holzkiste gelang. Gelegentlich werden solche Fähigkeiten auch als „kognitiv“ bezeichnet. In Abschnitt 2 werden wir eine Begriffsdefinition von „kognitiv“ vorschlagen, die den Fokus auf den qualitativen Unterschied zwischen der Handhabung innerer und äußerer Freiheitsgrade legt. Diese Sichtweise stellt den Bezug zwischen kognitiven Fähigkeiten und den oben diskutierten Wechselwirkungen her, die zwischen den geometrischen und physikalischen Eigenschaften natürlicher Welten, zwischen symbolischer und geometrischer Planung sowie beim Lernen von Modellen der Umwelt existieren. In Abschnitt 3 geht es um Forschungsprojekte zur Objektmanipulation in natürlichen Umwelten. Während hier viele beeindruckende Leistungen demonstriert wurden, bleiben prinzipielle Forschungsfragen weiterhin offen. In den Abschnitten 4 bis 6 gehen wir auf drei solcher Fragen ein: des Lernen & Planen in relationalen Welten, das Symbol-Lernen und die autonome Exploration. Wir schließen mit einer Diskussion weiterer offener Fragestellungen und einem Ausblick.

2 Innere vs. äußere Freiheitsgrade

Der Begriff „Kognition“ wird oft benutzt, um zwischen der motorischen (niederen) Ebene und der Ebene abstrakter Aktionen zu unterscheiden. Wir wollen stattdessen zunächst den Schwerpunkt auf die qualitativen Unterschiede zwischen inneren und äußeren Freiheitsgraden legen. Typische Beispiele für innere Freiheitsgrade in der Robotik sind der Vektor $q \in \mathbb{R}^n$ der Gelenkwinkel (oder (q, \dot{q})), oder die Position-Orientierung-

Geschwindigkeit (x, θ, v) eines Fahrzeugs. Die traditionell in der Robotik entwickelten Methoden beziehen sich im Wesentlichen auf Modelle zur Prädiktion, Zustandsschätzung und Regelung dieser inneren Freiheitsgrade. Auch die Anwendung von Lernmethoden in der Robotik sind bzgl. dieser Freiheitsgrade besonders erfolgreich (Vijayakumar et al., 2005). Können solche Methoden innerer Freiheitsgrade direkt auf Freiheitsgrade der Umwelt verallgemeinert werden? Was als Objektmanipulation bezeichnet wird, kann man als eine Form der „Regelung“ äusserer Freiheitsgrade – der Positur von Objekten – ansehen. Einer der vielen qualitativen Unterschiede liegt jedoch darin, dass dieser externe Freiheitsgrad zunächst nicht regelbar ist, wenn das Objekt nicht gegriffen ist.³ Das Greifen wird damit zum zentralen Gegenstand der Forschung in der Objektmanipulation.

Das Greifen ist nur ein Beispiel dafür, dass die Handhabung externer Freiheitsgrade qualitativ neue Herausforderungen zu denen innerer Freiheitsgraden stellt. Innere Freiheitsgrade lassen sich oft in einen endlichdimensionalen Vektorraum einbetten. Dagegen ist der Zustandsraum äußerer Freiheitsgrade komplex und seine Struktur schwer formal abzubilden. Auf inneren Freiheitsgraden haben wir Methoden, um effiziente Wahrscheinlichkeitsmodelle zu formulieren – etwa Gaussverteilungen über dem Vektorraum. Dies liefert letztlich die Grundlage für effiziente Inferenz (Zustandsschätzung und -prädiktion) und Lernmethoden. Das Formulieren von Wahrscheinlichkeitsmodellen über die äußere Umwelt ist schwer. Beschreibt man beispielsweise den Zustand als eine Menge von Objekten mit den jeweiligen geometrischen, physikalischen und kinematischen Eigenschaften und Relationen, so existieren relationale Wahrscheinlichkeitsmodelle, die logische Repräsentationen und Wahrscheinlichkeiten vereinen. Sind jedoch Menschen oder andere autonome Systeme Teil der Umwelt, wird die Frage nach geeigneten Repräsentationen und Wahrscheinlichkeitsmodellen wesentlich schwieriger. Für innere Freiheitsgrade haben wir gut motivierte Priors (quadratische Regelungskosten, Gaussche Übergangsmo-delle), für externe nicht. Um interne Freiheitsgrade in einen Soll-Zustand zu überführen, ist meist kein hierarchischer Ansatz nötig – Pfadplanung und stochastische Regelungstheorie liefern meist direkte Trajektorien. Äußere Freiheitsgrade zu manipulieren verlangt oft hierarchisches und sequentielles Verhalten.

In dieser Sichtweise kann man unter „kognitiver Robotik“ die Erweiterung der theoretisch fundierten Methoden der Regelungstheorie, Pfadplanung, Zustandsschätzung, etc auf äußere Freiheitsgrade verstehen. Will man dies jedoch rigoros verfolgen, ergeben sich grundlegende Fragestellungen, die das klassische Gebiet

³ Formal: In der lokalen Linearisierung der Gesamtdynamik zeigt die Steuerbarkeitsanalyse im Sinne der linearen Regelungstheorie, dass der externe Freiheitsgrad nicht regelbar ist.

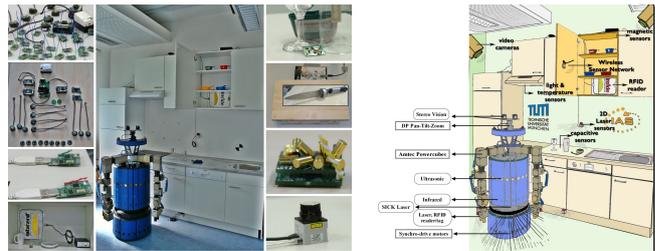


Bild 2: Forschung an integrierten, kognitiven Systemen in der Gruppe um Prof. Beetz (Beetz et al., 2008).

der Robotik verlassen, wie etwa der Umgang mit relationalen Wahrscheinlichkeitsmodellen oder die Formulierung von Wahrscheinlichkeitsmodellen, die fundamentale geometrische und physikalische Randbedingungen abbilden. In diesem Sinne führt unsere Begriffsdefinition kognitiver Systeme direkt zu der in der Einleitung diskutierten Wechselwirkung zwischen Robotik, KI und dem Maschinellen Lernen.

3 Forschungsüberblick

Im allgemeinen wird der Begriff „kognitiv“ unschärfer als oben vorgeschlagen genutzt. Dennoch zeigt sich bei existierenden Forschungsinitiativen in der kognitiven Robotik, dass der Fokus auf Aspekte intelligenter Systeme gesetzt wird, die über Regelung und Bahnplanung hinausgehen. Die Forschung beschäftigt sich mit komplexem sequentiellem Verhalten, Objektmanipulation, Mensch-Maschine Interaktion, der Kombination aus symbolischem Planen mit Regelung und Wahrnehmung, sowie schließlich der Fähigkeit, in diesen Szenarien zu lernen. Diese Aspekte der „kognitiven Robotik“ sind mit unserer Begriffsdefinition konsistent, in dem die Umwelt und ihre Struktur als zentraler Gegenstand der Forschung mit einbezogen wird.

Ein gutes Beispiel dafür, was ein kognitiver Roboter sein könnte, der effizient externe Freiheitsgrade manipuliert, findet sich in einem PR1-Video⁴. Der Roboter der Firma Willow Garage räumt hier ein Wohnzimmer vollständig auf: Er stapelt die Zeitschriften ordentlich, räumt Spielzeug in eine Box und schüttelt die Kissen auf. Leider ist die komplette Sequenz von einem Menschen teleoperiert! Der Roboter selbst nutzt keinerlei Perzeption, Bewegungs- oder Handlungsplanung, sondern ist letztlich eine Verlängerung des vom Menschen gesteuerten Joysticks. Das Video demonstriert, was heutige Roboter im Prinzip, trotz der rudimentären Mechanik, tun könnten und veranschaulicht damit das unerreichte Ziel der Forschung an autonomen Robotern. Offensichtlich ist das vornehmliche Problem der autonomen kognitiven Robotik nicht die Mechanik, sondern fehlende Methoden der Handlungsplanung, Modellierung und der Kontrolle externer Freiheitsgrade.

⁴ <http://www.youtube.com/watch?v=jJ4XtyMoxIA>

Michael Beetz' Forschungsgruppe entwickelt sehr interessante, integrierte kognitive Systeme, die Aufgaben in Küchenumgebungen lösen (Beetz et al., 2008). Abb. 2 zeigt ein solches Szenario zusammen mit den benutzten Sensoren am Roboter und auf den Gegenständen. Die erreichten Manipulationssequenzen sind nahe, aber nicht ganz so flexibel wie die Handlungssequenzen aus obigem PR1-Video, werden aber autonom vom Roboter ausgeführt. Neuere Arbeiten zeigen einen Roboter, der Pfannkuchen macht und dafür symbolische „Handlungsanleitungen“ aus dem Internet nutzt, wie sie unter howto.com oder roboearth.org speziell für Roboter zu finden sind. Aus Anwendungssicht ist dieser Ansatz, bei dem der einzelne Roboter nicht mehr lernen muss, sondern vorverarbeitetes Wissen über Objekte und Handlungen aus dem Internet bezieht, besonders vielversprechend. Die Karlsruher Forschungsgruppe betrachtet ähnlich integrierte Robotiksysteme in Küchenumgebungen (Asfour et al., 2006). Am Bielefelder CITEC liegt der Schwerpunkt eher auf der Interaktion zwischen Mensch und Roboter, aber auch auf der Grundlagenforschung am Greifen (Steil et al., 2004).

Die erwähnten Forschungsvorhaben sind Beispiele für integrierte Systeme, die Perzeption, Handlungsplanung, Objektmanipulation und Bewegungssteuerung im Zusammenspiel an relevanten Anwendungsszenarien demonstrieren. Eine bisher unzureichend gelöste Problematik in solchen Systemen ist das autonome Lernen auf der Handlungsebene. Nur wenige existierende Forschungsarbeiten widmen sich dem Problem des Handlungslernens (auf symbolischer Objektebene) in voll integrierten Robotiksystemen (Beetz et al. (2008), siehe auch die Diskussionen zum Lernen auf Systemebene (Thrun, 2000; Andre and Russell, 2001)).

Im Folgenden wollen wir einige eigene Arbeiten diskutieren, die den Fokus auf sogenanntes relationales Reinforcement Lernen legen, das ein vielversprechender Ansatz für das Lernen auf Ebene der Objektmanipulation und symbolischer Handlungsplanung ist. Während diese Methoden symbolische, logische Repräsentationen mit der statistischen Lerntheorie kombinieren, ergibt sich ein klassisches Problem: Woher kommen die Symbole? Abschnitt 5 wird explizit auf Symbol-Lernen zum Zwecke der Handlungsplanung eingehen und Abschnitt 6 auf autonome Exploration.

4 Relationales Lernen und Planen

Die Künstliche Intelligenz verfolgt den Ansatz, grundlegende Zusammenhänge in der Welt mittels abstrakter Symbole zu beschreiben. Unsere Alltagswelten zeichnen sich durch eine Vielzahl an Gegenständen aus. In unseren Küchen etwa finden wir Gläser, Teller und Töpfe, in unseren Büros Blätter, Stifte und Ordner. Kennzeichnend für diese Umgebungen ist die Struktur, die durch die Eigenschaften der Gegenstände und ihre Be-

ziehungen untereinander festgelegt wird. Wir können diese Struktur mit abstrakten Symbolen abbilden: Etwa mit Symbolen für den Typ, die Größe oder die Form eines Gegenstandes oder für geometrische Muster, wie beispielsweise, dass ein Gegenstand auf einem anderen liegt oder in einem anderen enthalten ist. Die symbolische Abstraktion umfasst auch die Handlungen mit Gegenständen. Verschiedene konkrete Greifbewegungen können mittels eines einzigen Symbols für Greifen abstrahiert werden.

Selbst wenn die Welt auf diese Weise symbolisch abstrahiert wird, bleibt die Anzahl möglicher Situationen unüberschaubar groß – nämlich exponentiell in der Anzahl der repräsentierten Gegenstände. Erhält ein kognitives System beispielsweise die Aufgabe, zehn Teller übereinander zu stapeln, so kann es dies auf $10!$ verschiedene Weisen tun. Nicht jeder einzelne mögliche Stapel soll als Ziel definiert werden. Daher muss ein kognitives System in der Lage sein, in seiner symbolischen Beschreibung über Gegenstände und Situationen zu verallgemeinern. Verallgemeinerung ist generell wichtig, um aus wenigen Erfahrungen lernen und in verschiedenartigen Situationen handeln zu können. Wenn wir jede zuvor ungesehene Teetasse als einen vollkommen unbekanntem Gegenstand auffassen und jedes Greifen eines weiteren Gegenstands als eine neuartige Bewegung, sind wir nicht fähig, in unserer vielschichtigen Umgebung sinnvoll zu handeln.

So genannte relationale Repräsentationen (Fikes and Nilsson, 1971) bieten die Möglichkeit, erlerntes Wissen zu verallgemeinern. Relationale Repräsentationen beschreiben eine Situation durch die Menge der Eigenschaften und Beziehungen (Relationen) von Gegenständen in Form von logischen Prädikaten und Funktionen. Zum Beispiel drückt das Prädikat $on(a, b)$ aus, dass der Gegenstand a auf b liegt. Grundlegend für die Fähigkeit zu verallgemeinern ist die Annahme, dass die Wirkung von Handlungen nur von den Eigenschaften und Typen der Gegenstände abhängt (beschrieben durch die Prädikate), nicht aber von den jeweiligen Identitäten der Gegenstände (dass es sich um Gegenstand a handelt und nicht b). Indem man Variablen für Gegenstände einführt, können strukturelle Äquivalenzklassen von Situationen beschrieben werden. Zum Beispiel kann das Symbol $on(a, X)$ dazu benutzt werden, alle möglichen Weltsituationen zusammenzufassen, in denen sich der Gegenstand mit dem Bezeichner a auf irgendeinem anderen, nicht weiter spezifizierten Gegenstand X befindet. Abstrakte relationale Beschreibungen ermöglichen dadurch, kompakte, verallgemeinernde Modelle über die Wirkungsweise von Handlungen zu definieren.

Um relationale Modelle tatsächlich auch in den komplexen Szenarien der echten Welt einsetzen zu können, ist es von entscheidender Bedeutung, in ihnen die Unsicherheit von Wissen auszudrücken. Die Wirkung von Handlungsketten ist selbst in einfachen Alltagsumge-

bungen nicht immer eindeutig vorherzusagen; das heißt, wir können nicht eindeutig abschätzen, ob im Folgezustand einer Handlung ein Symbol wahr oder falsch sein wird. Beim Einschenken in ein Glas kann Wasser verschüttet werden, Teller fallen auf den Boden, Busse haben Verspätung. Diese Unsicherheit über die Wirkung hat vielfältige Gründe. Zunächst mag die Welt inhärent stochastisch sein. Doch wichtiger als Stochastizität ist, dass wir Vorhersagen nie mit völliger Sicherheit treffen können, weil wir stets nur unvollständige Informationen über die Welt zur Verfügung haben. Teller mögen feuchter und rutschiger sein, als wir annehmen, wir wissen nicht, dass ein Bus eine Panne hat. Wir können unsere Welt stets nur teilweise und ungenau wahrnehmen. Darüber hinaus ist es gerade der Zweck einer abstrakten Beschreibung, Einzelheiten außer Acht zu lassen und damit nur teilweise Information zu repräsentieren. Vor allem aber ist Wissen zwangsläufig unsicher, wenn es aus Erfahrung gelernt und generalisiert wird. In Bayesianischen Modellen ist die Wechselwirkung zwischen Generalisierung, Prior und Unsicherheit des Modells explizit. In frequentistischen Modellen wird die inhärente Unsicherheit eines verallgemeinernden Modells über die Regularisierung geregelt. In beiden Fällen lässt sich die Unsicherheit einer verallgemeinernden Prädiktion nicht vermeiden (*bias-variance tradeoff*).

Stochastische relationale Modelle ermöglichen die Modellierung von Unsicherheit, und damit sowohl das Erlernen kompakter und verallgemeinernder Modelle aus Erfahrung als auch das Planen mit erlernten Modellen. Ein Beispiel für stochastische relationale Modelle bilden probabilistische relationale Regeln wie die folgende:

$$\begin{aligned} \text{greife}(X) : & \text{ klotz}(X), \text{ ball}(Y), \text{ auf}(Y, X) \\ \rightarrow & \begin{cases} 70\% : \text{ inhand}(X), \neg \text{ auf}(Y, X) \\ 20\% : \neg \text{ auf}(Y, X) \\ 10\% : \text{ noise} \end{cases} \end{aligned}$$

Diese Regel beschreibt auf abstrakte Weise eine Situation, in der ein System versucht, einen Klotz zu greifen, auf dem ein Ball liegt. Diese Regel dient der Vorhersage in jeder Situation, in der das System einen sonst nicht näher spezifizierten Klotz, auf dem ein Ball liegt, greifen möchte. Die relationale Beschreibung fasst hierbei eine exponentielle Anzahl an möglichen Situationen in einer kurzen probabilistischen Regel zusammen.

Pasula et al. (2007) beschreiben einen Algorithmus zum Lernen von Mengen solcher Regeln aus einer Erfahrungsmenge. Der Algorithmus folgt fundamental der klassischen Lerntheorie: Er versucht alle Erfahrungen mit der Regelmenge zu erklären (*likelihood*-Maximierung), aber gleichzeitig möglichst wenige und kompakte Regeln zu nutzen (Regularisierung in Form einer minimalen Beschreibungslänge). Bei wenig Erfahrung entstehen wenige, kompakte und stark (über-)verallgemeinernde Regeln; bei mehr Erfahrung werden die Regeln mehr, expliziter und genauer. Dieser gewünschte Effekt basiert fundamental auf

der Möglichkeit, Wahrscheinlichkeiten mit diesen Regeln auszudrücken. Das Inkaufnehmen von Modellungenauigkeiten ist also entscheidend für Lernen, Generalisierungsfähigkeit und Kompaktheit.

Das Nutzen erlernter probabilistischer Regeln zur Handlungsplanung ist Gegenstand aktueller Forschung. Die Herausforderung liegt einerseits in dem exponentiell großen Zustandsraum, der mit relationalen Regeln beschrieben wird, andererseits in der Stochastizität des Modells, was die direkte Anwendung klassischer deterministischer Logik verhindert. Ein möglicher Ansatz ist es, die Regeln zunächst in eine andere Repräsentation – faktorisierte Dynamische Bayesnetze – zu übersetzen und dann probabilistische Inferenzmethoden zur Planung zu nutzen. Lang and Toussaint (2010) beschreiben den ersten Algorithmus, der effizient mit erlernten probabilistischen Regeln planen kann.

5 Symbole lernen

Bei der Diskussion relationalen Lernens und Planens nahmen wir an, dass bereits abstrakte Symbole existieren. Ein autonomes System nimmt die Welt jedoch über Sensoren wahr, die im Wesentlichen geometrische Informationen liefern. Symbole müssen daher entweder von einem menschlichen Experten vorgegeben oder, wie wir vorschlagen, aus Erfahrung gelernt werden. Was sind geeignete Kriterien und Gütemaße für Symbole, die eine Grundlage für das Lernen bieten können? Im Folgenden diskutieren wir kurz Beispielsätze.

Wellens et al. (2008) modelliert Lernen von Symbolen als Kommunikationproblem: Agenten müssen die Bedeutungen von unbekanntem Wörtern so lernen und für sich selbst definieren, dass kohärente Kommunikation möglich ist. Die Autoren testen dieses Modell in Szenarien, in denen Roboter Objekte bewegen und dies einem anderen Roboter kommunizieren müssen. Die Roboter lernen dadurch eine Abbildung von visuellen Merkmalen der Objekte auf symbolische Wörter. Lernen ist erfolgreich, falls die Agenten einander zeigen können, welche Objekte sie demnächst manipulieren wollen.

Kollar et al. (2010) lernt die Bedeutung von Wörtern, die in Wegbeschreibungen vorkommen, aus Beispielen von Bewegungstrajektorien und deren Beschreibung in natürlicher Sprache von einem Lehrer. Man kann auf Grund dieser Daten die geometrische Bedeutung vieler Wörter (z.B. „links abbiegen“) lernen. Es wird aber ein Lehrer benötigt, der große Datenmengen erzeugt.

Feldman (2012) argumentiert als Kognitionswissenschaftler, dass eine symbolische Beschreibung alle wichtigen Eigenschaften einer kontinuierlichen Welt beinhalten kann. Die Formalisierung mittels üblicher Mixtur-Modelle in einem Vektorraum scheint jedoch naiv, wenn man an natürliche Welten denkt. Zudem wird in keiner Weise der Nutzen von Symbolen für die Organisation

von Verhalten, z.B. für das Lernen und Planen in Betracht gezogen.

Unserer Meinung nach sollten sich gute Symbole für kognitive Systeme gerade dadurch auszeichnen, dass sie die für die Handlungsplanung wesentlichen Aspekte der Umwelt repräsentieren und effizientes, generalisierendes Lernen ermöglichen. Dies steht im Gegensatz zum rein unüberwachten Lernen, das Symbole oder Unterdimensionen sucht, die die Varianz in Daten erklärt, unabhängig von deren eigentlichen Nutzen.

Unser Ansatz versucht dies in konkreten Robotik-Szenarien zu realisieren (Jetchev et al., 2013). Wir betrachten einen Agenten, der aus einer Menge motorischer Fähigkeiten (Primitive) wählen kann, um eine Bewegungssequenz zu erzeugen und damit die Objekte in der Welt zu manipulieren. Der Agent muss aus beobachteten kontinuierlichen Objektmerkmalen und -relationen $y \in \mathbb{R}^d$ abstrakte Symbole lernen. Diese Symbole sollen idealerweise die wichtigen Muster repräsentieren, zum Beispiel, ob gewisse Fähigkeiten in der momentanen Situation anwendbar sind und welche Wirkung sie erzielen. Die Bedeutung von „wichtig“ kann sich aber in verschiedenen Situationen unterscheiden. Wenn ein Affe etwas Essbares in einem Wald sucht, sind vielleicht Eigenschaften wie „rot“, „weich“ und „giftig“ wichtig. Wenn derselbe Affe eine Waffe sucht, sind „hart“ und „scharf“ wichtige Symbole. Erfolgreiche Agenten, die ihre Ziele erreichen können, haben Symbole gelernt, die zu Belohnung führen (hier: Ziel erreicht). Die Qualität der gelernten Symbole wird indirekt gemessen, mittels der Fähigkeit des Agenten, zielgerichtetes Verhalten mit diesen Symbolen zu erzeugen.

Wir definieren ein Symbol als relationales Prädikat ohne immanente Semantik. Zu diesem Symbol gehört eine Abbildung (Klassifikator) von geometrischen Merkmalen auf binäre Wahrheitswerte des Prädikats – dem *Grounding* des Symbols. Symbol-Lernen reduziert sich dann auf das Lernen geeigneter Klassifikatoren. Entscheidend ist nun die Wahl der Gütefunktion dieser Klassifikatoren. Unser Ansatz kombiniert drei Kriterien: Erstens sollen die Symbole gleichzeitig *prädikativ und diskriminativ* bzgl. der Wirkung von Aktionen sein. Symbole sollen also den Vor- und Nach-Zustand einer Aktion unterscheiden können und gleichzeitig hinreichend Informationen liefern, dass ein auf diesen Symbolen gelerntes Modell den Nach-Zustand vorhersagen kann. Zweitens sollen Symbole *Belohnungs-diskriminativ* sein, d.h. sie sollen hinreichend Informationen liefern, so dass ein auf diesen Symbolen gelerntes Modell die Belohnung vorhersagen kann. Drittens sollen Symbole einfach sein, d.h. möglichst wenige und selten wechselnde Symbole sollen gesucht werden.

Diese drei Kriterien lassen sich formal fassen. Das Erlernen geeigneter Symbole wird dadurch zu einem Optimierungsproblem der entsprechenden Klassifikatoren (Jetchev et al., 2013). Ein anschauliches Beispiel betrachtet die klassische Welt aus Blöcken und Kugeln, die von ei-

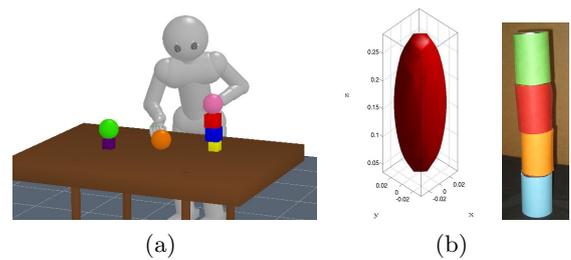


Bild 3: Beispielwelt für das Manipulieren von Objekten und gelernte Symbole für $stehtauf(X, Y)$: es ist wahr, wenn zwei Blöcke aufeinander gestapelt sind.

nem Agenten bewegt werden können (siehe Abbildung 3(a)). Unser Agent beobachtet Sequenzen zufälliger Aktionen. Die Belohnung ist abhängig von der Höhe der gebauten Türme. Ohne zuvor definierte Symbole besteht die Beobachtung des Agenten ausschliesslich aus den geometrischen Effekten der Aktionen: Wie sich (relative) Koordinaten und geometrische Beziehungen zwischen den Objekten ändern. Durch den oben genannten Optimierungsprozess lernt der Agent aus diesen Daten Prädikate, die wir z.B. mit $istinhand(X)$, $stehtauf(X, Y)$ benennen würden – aus Sicht des Agenten besteht die Semantik dieser gelernten Symbole aus nichts anderem als deren Rolle in den prädiktiven Modellen. Auf Grund der Optimierungskriterien erlauben es die Symbole effektiv – auf Basis der im vorigen Abschnitt diskutierten Methoden – Aktionssequenzen zu planen, die dem Agenten hohe Belohnung verschaffen. Zum Beispiel wird erlernt, dass manche Gegenstände, etwa Kugeln, nicht gestapelt werden können. In unserem Experiment wurde ein entsprechendes Symbol gelernt (wir könnten es $istkugel(X)$ benennen), da es zu einem besseren prädiktiven Modell führt. Viele andere denkbaren Symbole, wie z.B. $linksvon(X, Y)$ wurden nicht gelernt, da sie für die Belohnung oder Transitionsvorhersage in der konkreten Testumgebung irrelevant sind.

6 Autonome Exploration

Kognitive Systeme sollen selbstständig und zielgerichtet in ihrer Umgebung handeln. Aus ihren Erfahrungen müssen sie verstehen lernen, auf welche Weise sie auf ihre Umgebung einwirken können. Nur selten steht ihnen dabei ein Lehrer zur Seite. Stattdessen müssen sie ihren derzeitigen Wissensstand überdenken und daraufhin selbst aussuchen, was sich lohnt zu explorieren.

Wie kann man ein kognitives System bauen, das eine neue Umgebung in möglichst kurzer Zeit erkundet und dadurch ein Modell der eigenen Handlungen erlernt? Die in der jüngeren KI- und Robotik-Forschung entwickelten Explorationsmethoden realisieren eine Form von „Neugier“, die Lernoptimalität oder -effizienz in dem ein oder anderen Sinne garantiert. Zentrale Paradigmen sind etwa das sog. *Bayesische Reinforcement-Lernen* (Poupart

et al., 2006) und auch *R-max* (Brafman and Tenenholz, 2002). Im Wesentlichen erhält das System eine formale Belohnung jedes Mal, wenn es etwas Neues lernt. Ein wesentliches Problem ist dabei das Abwägen zwischen dem möglichen Wert neuer Erfahrungen und dem Ausnutzen bereits erlernten Wissens (der *exploration-exploitation tradeoff*). Die genannten Methoden geben mögliche optimale Antworten auf dieses Problem.

Bei der Anwendung solcher Explorationsmethoden in Alltagsumgebungen ergeben sich interessante Herausforderungen: Die Anzahl möglicher Handlungen ist unüberschaubar, der Raum möglicher Weltzustände ist wie oben dargelegt exponentiell in der Anzahl der modellierten Gegenstände. In einem dermaßen großen Suchraum kommt es so gut wie nie vor, dass exakt dieselbe Situation zweimal auftritt. Vielmehr ist jede Situation neu. Daher kann Neuheit allein keine effiziente Explorationsstrategie sein. Von entscheidender Bedeutung ist die Verallgemeinerung des eigenen Wissens auf neue, aber strukturell äquivalente Situationen. Wenn ein bisher erlerntes Modell scheinbar gut auf eine neue Situation verallgemeinert werden kann, so ist diese weniger interessant für die Exploration. Die oben beschriebenen stochastischen relationalen Modelle implizieren also nicht nur eine bestimmte Form der Verallgemeinerung beim Lernen, sondern auch eine bestimmte Explorationsstrategie. Nehmen wir als Beispiel einen Roboter, der einige Zeit mit blauen und roten Blöcken Erfahrung gesammelt hat. Danach wird ihm ein grüner Klotz und ein grüner Ball gegeben. Beide Gegenstände hat der Roboter noch nie gesehen. Ein simpel gestricktes System würde beide daher als gleich interessant einstufen. Ein Roboter, der relationale Modelle lernt, kann hingegen ausnutzen, dass er bereits zuvor anhand der blauen und roten Klötze gelernt hat. Er weiß, was er mit Klötzen anfangen und dass er sie zum Beispiel aufeinander stapeln kann. Insbesondere hat er erkannt, dass die Farbe eines Klotzes keine Auswirkung darauf hat, was mit ihm gebaut werden kann. Ein solcher Roboter wird daher den grünen Ball zur Exploration bevorzugen, auf den sein zuvor erlerntes Wissen nicht verallgemeinert. Aktuelle Systeme, die in der stochastischen relationalen KI erforscht werden, zeigen in Experimenten genau dieses Verhalten (Lang et al., 2010, 2012).

7 Zusammenfassung

In den 1960er Jahren stellte das MIT *Shakey* vor, einen mobilen Roboter, der Objekte verschieben, greifen und plazieren konnte (Nilsson, 1984). Im Grunde hat die hier diskutierte Forschung, und generell die Forschung an kognitiven Systemen, immer noch sehr ähnliche Ziele. Hat die Robotikforschung also in den letzten 40 Jahren keinen Fortschritt gemacht? Tatsächlich ist das, was mit Robotern im Kontext der autonomen Objektmanipulation und Handlungsplanung demonstriert wird,

in *oberflächlicher* Hinsicht vergleichbar geblieben. Dennoch gab es hinter diesen Demonstrationen wichtige Fortschritte in der Robotik, KI und im Maschinellen Lernen: Die Leistungen von *Shakey* – genauso wie die im telegesteuerten PR1-Video – spiegeln direkt die Intelligenz der Ingenieure wieder, nicht die des Roboters. Die Forschung versucht seitdem, mehr Autonomie und Intelligenz im System selbst zu verwirklichen, insbesondere die Fähigkeit des Lernens. Im Zuge dieser Forschung schärfte sich vor allem unser Verständnis der fundamentalen Probleme, die in der Anfangszeit der Künstlichen Intelligenz und Robotik nicht wahrgenommen wurden.

Eines dieser Probleme ist Unsicherheit: In der klassischen KI wurde oft von fehlerfreien symbolischen Modellen der Umwelt ausgegangen – woher die notwendigen symbolischen Abstraktionen und das Wissen selbst kommt, wurde vernachlässigt. Erlerntes Wissen ist aber zwangsläufig mit Unsicherheit behaftet. Für das effiziente Lernen, die Exploration und Planung mit unsicherem Wissen sind erst in jüngerer Zeit vielversprechende Methoden entwickelt worden. Das Erlernen geeigneter symbolischer Abstraktionen bleibt weiterhin ein wichtiges Forschungsfeld. Die hier diskutierten Methoden sind nur ein erster Schritt.

Es werden auch Lücken in unseren theoretischen Methoden erkennbar. Um die Schätzung, Prädiktion und Manipulation externer Freiheitsgrade zu ermöglichen, brauchen wir geeignete Repräsentationen, strukturierte Wahrscheinlichkeitsmodelle. Die hier diskutierten relationalen Modelle sind nur ein Beispiel für geeignete Repräsentationen, die die Struktur der Umwelt wieder spiegeln – in diesem Fall ihre Komposition aus Objekten. Für viele andere Charakteristika der natürlichen Umwelt haben wir bisher noch keine mathematischen Formalismen, um sie auszudrücken und entsprechend generalisierende Lern- und Planungsmethoden abzuleiten.

Wenn man also unter kognitiver Robotik, wie hier vorgeschlagen, die Erweiterung von Lernen, Planen, Inferenz und Regelung auf äußere Freiheitsgrade verstehen will, so wird es zum zentralen Problem, die Struktur dieser äußeren Welt in geeignete Priors, Repräsentationen und Modellannahmen zu übersetzen. Mit den hier diskutierten Ansätzen zum relationalen Lernen und Symbol-Lernen sind erste Beispiele gegeben. Die existierenden Forschungsrichtungen zur Integration von Logik, Geometrie und Wahrscheinlichkeiten (etwa (Kaelbling and Lozano-Pérez, 2011)) gehen in dieselbe Richtung. Dennoch wird die Forschung vermutlich noch einige Zeit mit ähnlichen Szenarien beschäftigt sein, wie sie oberflächlich schon vor 50 Jahren mit *Shakey* begonnen wurden. Anders als damals sind das Ziel jedoch Systeme, die durch intelligente Exploration und Verallgemeinerung autonom lernen. Mit selbst erlernten Modellen und Symbolen sollen sie in gewisser Weise „begreifen“, was sie tun, und dies eigenständig zur Verhaltensorganisation nutzen. Die Fortschritte in der Lerntheorie und Robotik erlauben uns es heute besser zu verstehen, welche

fundamentalen wissenschaftlichen Probleme sich hinter diesem Ziel verbergen.

Danksagung

Die Autoren bedanken sich bei der Deutschen Forschungsgemeinschaft für die Einrichtung des Schwerpunktprogramms SPP 1527. Diese Forschung wurde im Rahmen des Projekts TO 409/7-1 gefördert.

Literaturverzeichnis

- D. Andre and S. Russell. Programmable reinforcement learning agents. In *Proceedings of the 13th Conference on Neural Information Processing Systems (NIPS 2001)*, pages 1019–1025, 2001.
- T. Asfour, K. Regenstein, P. Azad, J. Schroder, A. Bierbaum, N. Vahrenkamp, and R. Dillmann. Armar-iii: An integrated humanoid platform for sensory-motor control. In *Humanoid Robots, 2006 6th IEEE-RAS International Conference on*, pages 169–175. Ieee, 2006.
- M. Beetz, F. Stulp, B. Radig, J. Bandouch, N. Blodow, M. Dolha, A. Fedrizzi, D. Jain, U. Klank, I. Kresse, et al. The assistive kitchen—a demonstration scenario for cognitive technical systems. In *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, pages 1–8. IEEE, 2008.
- R. I. Brafman and M. Tennenholtz. R-max - a general polynomial time algorithm for near-optimal reinforcement learning. *Journal of Machine Learning Research (JMLR)*, 3:213–231, 2002.
- J. Feldman. Symbolic representations of probabilistic worlds. *Cognition*, 123:61–83, 2012.
- R. Fikes and N. Nilsson. STRIPS: a new approach to the application of theorem proving to problem solving. *Artificial Intelligence Journal*, 2:189–208, 1971.
- N. Jetchev, T. Lang, and M. Toussaint. Learning grounded relational symbols from continuous data for abstract reasoning. Submitted to *ICRA 2013*, 2013.
- L. Kaelbling and T. Lozano-Pérez. Hierarchical task and motion planning in the now. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1470–1477. IEEE, 2011.
- W. Köhler. *Intelligenzprüfungen an Menschenaffen*. Springer, Berlin (3rd edition, 1973), 1917. English version: Wolfgang Köhler (1925): *The Mentality of Apes*. Harcourt & Brace, New York.
- T. Kollar, S. Tellex, D. Roy, and N. Roy. Toward understanding natural language directions. In *HRI*, pages 259–266, 2010.
- T. Lang and M. Toussaint. Planning with noisy probabilistic relational rules. *Journal of Artificial Intelligence Research (JAIR)*, 39:1–49, 2010.
- T. Lang, M. Toussaint, and K. Kersting. Exploration in relational domains for model-based reinforcement learning. In *Proc. of the European Conf. on Machine Learning (ECML)*, 2010.
- T. Lang, M. Toussaint, and K. Kersting. Exploration in relational domains for model-based reinforcement learning. *Journal of Machine Learning Research*, 13:3691–3734, 2012.
- N. Nilsson. Shakey the robot. Technical Note 323. AI Center, SRI International 323, 1984.
- Y. Niv, J. D., and D. P. A normative perspective on motivation. *Trends in Cognitive Sciences (TICS)*, 10:375–381, 2006.
- H. M. Pasula, L. S. Zettlemoyer, and L. P. Kaelbling. Learning symbolic models of stochastic domains. *Journal of Artificial Intelligence Research (JAIR)*, 29:309–352, 2007.
- J. Pearl. *Probabilistic Reasoning In Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.
- P. Poupart, N. Vlassis, J. Hoey, and K. Regan. An analytic solution to discrete Bayesian reinforcement learning. In *Proc. of the Int. Conf. on Machine Learning (ICML)*, pages 697–704, 2006.
- J. Steil, F. Röthling, R. Haschke, and H. Ritter. Situated robot learning for multi-modal instruction and imitation of grasping. *Robotics and Autonomous Systems*, 47(2):129–141, 2004.
- S. Thrun. Towards programming tools for robots that integrate probabilistic computation and learning. In *Proceedings of the IEEE Int. Conf. on Robotics and Automation (ICRA 2000)*, 2000.
- S. Vijayakumar, A. D’Souza, and S. Schaal. Incremental online learning in high dimensions. *Neural Computation*, 17(12):2602–2634, 2005.
- P. Wellens, M. Loetzsch, and L. Steels. Flexible word meaning in embodied agents. *Connection Science*, 20(2-3):173–191, 2008.

Prof. Dr.rer.nat. Marc Toussaint ist Professor und Leiter des *Machine Learning and Robotics Lab* an der Universität Stuttgart. Hauptarbeitsgebiete: Maschinelles Lernen, Reinforcement Lernen, Robotik.

Adresse: Universität Stuttgart, Universitätsstraße 38, 70569 Stuttgart, marc.toussaint@informatik.uni-stuttgart.de

Dr.rer.nat. Tobias Lang ist Wissenschaftler im *Machine Learning and Robotics Lab* der FU Berlin und erforscht autonome Exploration und Schlussfolgerung kognitiver Systeme.

Adresse: tobias.lang@fu-berlin.de

Dr.rer.nat. Nikolay Jetchev ist Wissenschaftler im *Machine Learning and Robotics Lab* der FU Berlin und erforscht, wie Roboter aus Demonstration lernen und ihre Umwelt in symbolischen Beschreibungen abstrahieren können.

Adresse: nikolay.jetchev@fu-berlin.de