Antizipierende interaktiv lernende autonome Agenten

13

Kognitive Modellansätze für eine Realisation von gegenseitiger Antizipation in der Mensch-Roboter-Kollaboration

Nele Rußwinkel

Zusammenfassung

Bisher ist die Zusammenarbeit von Menschen mit autonomen Agenten noch sehr eingeschränkt und künstlich. Um zu einer möglichst natürlichen Zusammenarbeit in der Mensch-Roboter-Kollaboration zu kommen, werden Ansätze benötigt, die auf den menschlichen Fähigkeiten der Antizipation anderer fußen. Der Frage wird nachgegangen, wie der Mensch in der Lage ist, Kollaborationspartner zu antizipieren und mit ihnen indirekt zu kommunizieren. Weitere Fragen sind, wie mögliche kritische Situationen vorhergesehen werden können und wie eine flexible Aufgabenallokation realisiert werden kann. Die Integration kognitiver Modellansätze kann einem autonomen Agenten antizipative Fähigkeiten verleihen. Vorgestellt werden hierzu Umsetzungen mentaler Modelle spezifischer Situationen (1), spezielle Eigenschaften des Kooperationspartners (2) als auch der möglichen Handlungen und deren Auswirkungen des Agenten selbst (3). Es werden Beispiele aus dem Feld der kognitiven Assistenz herangezogen, in welcher ebenfalls Kollaborationspartner antizipiert werden. Des Weiteren ist die Fähigkeit, schnellere und flexiblere Lernformen zu verwenden, gegeben. Die Integration kognitiver Systemansätze in autonome Systeme könnte einige derzeit bestehende Probleme, wie die fehlende Transparenz und Anpassung an die Nutzer, weitreichend lösen.

N. Rußwinkel (✉)
Technische Universität Berlin, Berlin, Deutschland
E-Mail: nele.russwinkel@tu-berlin.de

H.-J. Buxbaum (Hrsg.), *Mensch-Roboter-Kollaboration*,
https://doi.org/10.1007/978-3-658-28307-0_13

13.1 Einleitung

Wir sind noch weit entfernt von einem natürlichen Zusammenwirken von Mensch und Roboter. Bisher sind die Handlungsspielräume für die meisten Arbeitskontexte größtenteils deutlich voneinander getrennt. In vielen Einsatzbereichen, wie z. B. der Pflege, bewegen sich Roboter sehr langsam, um die Sicherheit der Nutzer zu gewährleisten. Arbeiten Mensch und Roboter nah beieinander, sind die Aufgaben klar zugeteilt. Auch wenn hier, aufgrund eines geteilten Arbeitsraums, schon von einer Kollaboration gesprochen wird (Onnasch et al. 2016) und gemeinsame Ziele identifiziert werden können (z. B. ein Werkstück zu fertigen), ist noch keine natürliche Zusammenarbeit erkennbar. In diesem Text wird der übergeordneten Frage nachgegangen, über welche Fähigkeiten ein Roboter verfügen muss, um in der Lage zu sein, den Nutzer auch in unbekannten Situationen sinnvoll unterstützen zu können. Diese recht umfangreiche Frage wird auf zwei Teilfragen heruntergebrochen:

1. Welche Voraussetzungen müssen erfüllt sein, um eine flexible Kollaboration von Mensch und Roboter realisieren zu können, indem die Absichten des anderen abgeleitet werden und die nächsten Handlungsschritte antizipiert werden können?
2. Über welche Fähigkeiten muss ein Roboter verfügen, um durch die Interaktion mit einer dynamischen Umwelt und der Beobachtung von Prozessen flexibel neue Aufgaben erlernen zu können? Für beide Fragen ist es notwendig zu erörtern, welche kognitiven Mechanismen diesen menschlichen Fähigkeiten zugrunde liegen, und ob diese kognitiven Mechanismen modelliert und auf autonome Systeme übertragen werden können?

Offensichtlich benötigen wir Modellansätze, die eine Form von Verstehen eigener und fremder Handlungen in einer Situation über deren Resultate ermöglichen.

Bisher verfügen autonome Agenten nicht über eine Instanz des „Verstehens“, sie sind nicht in der Lage „den Sinn von etwas zu erfassen“. Warum und mit welchem Ziel wird etwas getan und mit welchen Handlungen werden welche Veränderungen in der Umgebung verursacht? Diese Fragen müssen von einem Roboter beantwortbar sein, wenn dieser auch in neuen Situationen flexibel agieren soll. Wenn ein Mensch mit einem anderen Menschen zusammen eine Aufgabe bearbeiten möchte, müssen beide über eine Vorstellung des gemeinsamen Ziels verfügen und beide müssen ihre Aufgabe gemeinsam koordinieren – selbst dann, wenn zu Beginn noch nicht klar ist, welche Arbeitsschritte notwendig sind und wer welche Teilaufgaben übernimmt. Bisherige Verfahren, einem Roboter eine neue Aufgabe beizubringen, umfassen hauptsächlich zahlreiche Beispiele der exakten Handlungsabfolge bzw. die direkte Programmierung dieser. Diese Form des Lernens ist aufwendig und baut auf einer festen Sequenz von Handlungsanweisungen auf.

Wünschenswert wäre, dass ein Roboter lernt zu verstehen, warum welcher Handlungsschritt wie ausgeführt werden muss, wie diese Handlungsschritte variieren können und worin das eigentliche Ziel besteht.

Ziel dieses Textes ist es, aufzuzeigen, wie Robotern bzw. autonomen Agenten eine einfache Form des „Verstehens“ verliehen werden kann, welche Vorteile dies bringt und inwiefern sich die Interaktion mit autonomen Agenten durch derartige Ansätze verändern wird.

Der Ansatz, der in diesem Text vorgestellt wird, beschränkt sich zunächst auf das Verstehen des Kooperationspartners in einer Aufgabe und auf das Verstehen beim Lernen neuer Aufgaben. Es geht nicht darum, einen Menschen nachzubauen. Wir Menschen verfügen über zahlreiche kognitive Fähigkeiten, die es uns ermöglichen, unser Gegenüber auf verschiedenen Ebenen zu verstehen und selbst aus Einzelereignissen zu lernen. Menschen sind Spezialisten darin, andere zu verstehen und zeitnah neue Aufgaben und Gefahren zu identifizieren. Die Wissenschaft beginnt gerade erst, Stück für Stück die einzelnen Mechanismen, die hier eine Rolle spielen, zu identifizieren. Aber bereits wenige einfache Mechanismen dieser komplexen Fähigkeiten würden autonomen Agenten einen sehr viel größeren und flexibleren Aktionsradius verleihen und eine völlig neue Art der Mensch-Roboter-Kollaboration ermöglichen.

Um ein enges Zusammenwirken zu realisieren, welches auf dem Verstehen des anderen basiert, müsste der autonome Agent über drei Repräsentationsformen verfügen. Der Agent sollte über (1) einfache Repräsentationen der Situation des menschlichen Partners verfügen, dies beinhaltet das aktuelle Ziel oder Teilziel, die hierfür relevanten wahrgenommenen Informationen und der nächste Handlungsschritt. Der Agent sollte über (2) eine Repräsentation der individuellen Eigenschaften des Partners verfügen. Dies beinhaltet, über welche besonderen Fähigkeiten oder Einschränkungen der Partner verfügt, und in welchen emotionalen Zustand er sich befindet (z. B. Stress, Überraschung oder Zufriedenheit). Die letzte Repräsentation (3) umfasst den Agenten selbst. Was hat der Agent für eine Vorstellung bzw. Information von der Situation (evtl. im Unterschied zum Partner). Was kann der Agent in seinem Umfeld für Handlungen ausführen, die für das Ziel relevant sind.

Zunächst soll dargelegt werden, welche Art der Zusammenarbeit von Mensch und Roboter langfristig anvisiert werden soll, bevor auf mögliche konkrete Modellformen eingegangen wird, die das Potenzial haben, dies zu realisieren. Im Anschluss wird das Thema interaktives Lernen und die hierfür notwendigen Ansätze erörtert.

13.2 Vision eines natürlichen Zusammenwirkens von Mensch und Roboter

Es herrscht eine Vorstellung von Robotern der Zukunft vor, die in der Lage sind, unkompliziert mit uns Menschen zu interagieren. Diese Interaktion würde von einem Menschen als natürlich wahrgenommen werden und benötigt kein langwieriges Training des Nutzers. Menschen sind in der Lage, sich auf unterschiedlichste natürliche Agenten einzustellen, wie z. B. Kinder und Tiere. Dies fällt uns Menschen leichter, als uns an künstliche Systeme anzupassen, was eine umfangreiche Anpassung des Nutzers an das technische System fordert. Interaktionen mit natürlichen Agenten beinhalten normalerweise ein gegensei-

tiges Anpassen, bzw. wird versucht, dieses zu erlangen. Derartigen Interaktionen liegt normalerweise die Bemühung zugrunde, ein gegenseitiges Verständnis aufzubauen. Es wäre wünschenswert, dies auch für die Interaktion mit Robotern umzusetzen. In diesem Falle versteht der Nutzer die unmittelbaren Ziele des Roboters; gleichzeitig hat der Roboter eine Vorstellung der unmittelbaren Situation und der Ziele des Nutzers. Beide sind in der Lage, sich gegenseitig zu unterstützen, ohne dass umfangreiche explizite Instruktionen notwendig sind. Ist etwas unklar, gibt es verschiedene Wege diesen Punkt hervorzuheben und eine Antwort bzw. eine Lösung zu generieren; sei dies nun in expliziter oder impliziter Form. Möchte ein Agent einen anderen auf etwas hinweisen, reicht es oft in der Handlung innezuhalten und den Blick auf den Ort des Problems zu richten.

Zum jetzigen Zeitpunkt sind wir weit entfernt davon, Systeme zu bauen, die den Menschen verstehen, Handlungen antizipieren könnten und den Menschen bei seiner Aufgabe unterstützen können. Doch was genau beinhaltet eine gute Kollaboration zwischen zwei Agenten?

13.2.1 Was macht eine gute Mensch-Roboter-Kollaboration aus?

Bisherige Taxonomien der Art der Mensch-Roboter-Interaktion (Onnasch et al. 2016) unterscheiden zwischen Ko-Existenz, Kooperation und Kollaboration bei der Zusammenarbeit von Mensch und Roboter. Ersteres beschreibt ein episodisches Zusammentreffen von Mensch und Roboter, ein gemeinsames Ziel liegt dem Treffen nicht zugrunde. Bei der Kooperation wird auf ein gemeinsames übergeordnetes Ziel hingearbeitet, aber die Handlungen sind nicht unmittelbar voneinander abhängig, da es eine klare Aufgabenteilung gibt. Die Kollaboration hingegen beschreibt eine direkte Zusammenarbeit von Mensch und Roboter mit einer gemeinsam verfolgten Zielstellung und auch mit gemeinsamen Unterzielen, d. h. auch Teilhandlungen werden gemeinsam durchgeführt und erfordern eine unmittelbare Koordination der Handlungen. Die Zuteilung von Teilaufgaben erfolgt situationsangepasst während der Zusammenarbeit. Es geht hier um die Schaffung und Nutzung von Synergien.

Nach dieser Definition von Kollaboration ist die Zusammenarbeit situationsabhängig und es existiert im Vorfeld keine klare Aufgabenteilung. Diese Voraussetzungen erfordern bereits ein gegenseitiges Verstehen, zumindest der unmittelbaren Unterziele und Erfordernisse, als auch das Antizipieren der nächsten Handlungsschritte. Wie lässt sich ein solches Ziel umfassend realisieren? Und welche Anforderungen stehen je nach Aufgabe im Vordergrund?

In (Fiebich 2018) wird ein detaillierter Ansatz vorgestellt, um die Zusammenarbeit von Mensch und Roboter je nach Aufgabenanforderung einzuordnen (sie verwendet den Begriff „Kooperation“ für eine Form der Zusammenarbeit, die wir hier unter „Kollaboration“ eingeführt haben). Die Autorin stellt einen drei-dimensionalen Ansatz vor, um die jeweiligen Anforderungen einer gelungenen Mensch-Roboter-Kooperation festzulegen. Jedes kooperative Phänomen kann auf einem Kontinuum von drei Achsen beschrieben werden: der

verhaltensbasierten Achse, der kognitiven Achse und der affektiven Achse. Im Rahmen einer Aufgabe kann auf der verhaltensbasierten Achse eine koordinativ aufwendige oder weniger aufwendige Aktion erforderlich sein. Ebenso kann eine Situation kognitiv aufwendigere Verarbeitungen erfordern als eine andere. Aufwendigere kooperative Aktivitäten involvieren geteilte Intentionen, die auch kognitive Fähigkeiten wie „Theory of Mind" erfordern. Einfachere kognitive Aktivitäten erfordern eher eine „Intentional Joint Attention" (die Intention wird durch eine gemeinsame Blicklokation ausgedrückt, bzw. verstanden). Des Weiteren ist es relevant, die affektive Anforderung zu betrachten, da ein geteilter affektiver Zustand die Kooperation nachhaltig verbessert, evtl. die Intentionen und Motivation leichter geteilt werden können.

Die notwendigen Anforderungen bezüglich der drei Achsen an den Roboter werden nach diesem Ansatz in Abhängigkeit der vorliegenden Aufgabe definiert. Kognitiv komplexere Fähigkeiten werden benötigt, um aufwendigere Aufgaben gemeinsam zu bearbeiten. Genauso ist es z. B. anhand von dem Erkennen und Interpretieren von Emotionen möglich, eine feinere Abstimmung auf nonverbaler Ebene durchzuführen. Es ist wichtig zu erwähnen, dass nicht das reine Erkennen von Emotionen zu einer besseren Zusammenarbeit führen, sondern der Bezug des affektiven Zustandes zu der jeweiligen Situation und den erfolgten oder geplanten Handlungen.

Des Weiteren sind nicht für jede Aufgabe die aufwendigsten koordinativen, kognitiven und affektiven Fähigkeiten des autonomen Agenten gefordert. Jedoch sind einfache antizipative Fähigkeiten notwendig, um die Fähigkeiten der drei genannten Dimensionen umsetzen zu können.

Kinder sind z. B. erst ab einem Alter von etwa 6 Jahren in der Lage, sich in eine andere Person hineinzuversetzen und zu aufwendigen „kognitiven Prozessen" fähig, die die Theory of Mind postuliert (Fodor 1992; Mahy et al. 2014). Doch bereits kleinere Kinder und auch Tiere, wie z. B. Schimpansen (Premack und Woodruff 1978) sind in der Lage, kooperatives Verhalten zu zeigen. Diese einfacheren kognitiven Mechanismen sind demnach die besser umzusetzenden Ansätze für einen kollaborierenden autonomen Agenten, wie später im Text erläutert wird.

13.2.2 Was macht eine gute Mensch-Roboter-Interaktion der Zukunft aus?

Bei der Entwicklung einer zunehmenden Verbindung von Mensch und Roboter lassen sich nach Buxbaum und Sen (2018) bisher zwei verschiedene Arten von Unterstützungssystemen unterscheiden. Zum einen sind dies technische Systeme, die eine Person substituieren und dadurch zu einer Entlastung führen. Hierbei führt die Technik die Aufgabe für den Menschen aus. Die zweite Art technischer Systeme ersetzt nicht den Menschen, sondern unterstützt bei der Ausführung seiner Aufgabe. Hierbei behält der Mensch die Kontrolle über die Abläufe.

Doch für ein entsprechend gelungenes Zusammenwirken von Mensch und Roboter sollte es eine dritte Art von Unterstützungssystemen geben: Technische Systeme, die den Menschen ebenfalls nicht ersetzen, sondern bei der Ausführung ihrer Aufgaben unterstützen. Diese Unterstützung basiert jedoch auf einem gegenseitigen Verständnis. Die Kontrolle über die Abläufe wechselt hier flexibel zwischen Mensch und Roboter, je nachdem, wie die Situation es gerade fordert (dieser Punkt wird in einem späteren Abschnitt genauer erläutert). Der Kooperationspartner wird jeweils mitgeplant. Der Mensch sollte nach wie vor das Vorrecht haben, über den Abbruch der Aufgabenbearbeitung zu entscheiden bzw. andere wichtige Entscheidungen zu treffen.

Bisher wird der Mensch in Bezug auf halbautonome Systeme häufig als Störfaktor gesehen. Der Mensch macht Fehler, ist nicht perfekt vorhersehbar, darf nicht verletzt werden, hat begrenzte Muskelkraft und ermüdet schnell. Auf der anderen Seite verfügt der Mensch über Fähigkeiten, die nicht von technischen Systemen übernommen werden können. Menschen können aus Einzelereignissen lernen, abstrahieren und ihr Wissen auf neue Bereiche transferieren. Menschen können sehr gut andere Menschen antizipieren und sich in sie hineinversetzen und somit besser unterstützen bzw. passend Hilfe anbieten und frühzeitig Fehler vorhersehen. Menschen können sich auf Relevantes konzentrieren, d. h. sie filtern relevante Informationen von irrelevanten. Diese Fähigkeit kann zwar manchmal dafür sorgen, dass auch Informationen übersehen werden. Aber diese Fähigkeit hilft auch, relevante Faktoren in den Fokus zu stellen und bietet die Grundlage dazu, relativ schnell neue Aufgaben zu lernen oder Erwartungen aufzubauen und Prozesse zu antizipieren. Für ein erfolgreiches Zusammenwirken von Mensch und Roboter wäre es wichtig, diese unterschiedlichen Fähigkeiten sinnvoll zu kombinieren und die entstehende Synergie zu nutzen. Hierfür sind nach den bisherigen Überlegungen vier Voraussetzungen notwendig:

1. Das halbautonome System sollte die Eigenarten menschlicher Informationsverarbeitung in einem gewissen Ausmaß antizipieren können (d. h. zwischen Zeitdauern unterscheiden können, die für die Informationsverarbeitung notwendig sind und Zeiträumen, die auf Probleme hindeuten).
2. Auch das halbautonome System sollte von der Seite des Menschen her antizipierbar sein. Das heißt, zielführende Bewegungen sollten identifizierbar und Intentionen erkennbar sein.
3. Das halbautonome System sollte in der Lage sein zu begreifen, welche Manipulationen es in der Umwelt verursachen kann (d. h. eine Vorstellung des Active Self).
4. Das halbautonome System sollte abstraktes Wissen über wichtige Prinzipien der Welt verfügen, um aus der Umwelt relevante Zusammenhänge schnell ableiten zu können und Vorwissen anwenden zu können (Lake et al. 2017). Beispiele solcher Weltmodelle könnten sein, dass auf eine Aktion meist eine Reaktion folgt. Und wenn eine Aktion nicht zu der gewünschten Reaktion führt, man eine andere ausprobiert, oder bevorzugt das Gegenteil der zuvor gewählten Aktion versucht.

13.2.3 Beispiel einer antizipierenden Mensch-Roboter-Kollaboration

Das folgende Beispiel soll ein konkreteres Bild davon zeichnen, wie eine Mensch-Roboter-Kollaboration mit den oben genannten Fähigkeiten aussehen könnte.

Das Ziel eines solchen Systems wäre es, als eine „Dritte Hand" für den Menschen agieren zu können. Dies bedeutet, dort zu unterstützen, wo es der Kollaborationspartner gerade wünscht, ohne umständliche Erklärungen. Voraussetzung hierfür ist ein Verständnis des gemeinsamen Ziels und sich ergebender Unterziele, als auch die fehlenden Ressourcen bei einer Aufgabenbearbeitung (z. B. eine dritte Hand zusätzlich zum Halten oder ein fehlendes Auge, um etwas sehen zu können, das verdeckt ist) des anderen aus der Situation und den Zielen abzuleiten. Des Weiteren ist es notwendig, die Möglichkeiten zu reflektieren, ob und wie der autonome Agent in dieser Situation unterstützen könnte. Das bedeutet, dass man bei einem unterspezifizierten Ziel – z. B. beim Bau eines Gartenschuppens ohne Anleitung, auf ein gegenseitiges Verstehen angewiesen ist. Explizierte Kommunikation ist teilweise zu aufwendig und langwierig für derartige interaktive Aufgaben, daher geht es hier eher um eine implizite Kommunikation, die unmittelbar erfolgt und schnell verstanden werden kann. Die oben genannten vier Voraussetzungen werden nun anhand von beispielhaften Umsetzungen konkretisiert:

1. Der Roboter müsste verstehen, dass der Mensch etwas sucht, wenn der Blick hin und her wandert und das gerade erforderliche Werkzeug nicht gesehen wird (da es z. B. verdeckt ist und nur von der Roboterperspektive aus sichtbar ist). Hierfür wird eine Situation erkannt und die Intention des Partners kann direkt abgeleitet werden.
2. Wenn der Roboter sich auf das Werkzeug zubewegt, sollte zum einen die Bewegung antizipierbar sein, so dass der Mensch nicht gefährdet wird. Zusätzlich sollte der Blick des Roboters beispielsweise auf das Werkzeug gerichtet sein, um das Ziel der Bewegung zu kommunizieren. Hat der menschliche Kollaborationspartner dann ebenfalls den Blick auf das Ziel gerichtet, weiß der Roboter, dass diese Information verstanden wurde. Dies sind Formen impliziter Kommunikation und der Roboter zeigt antizipierbares Verhalten für den Kollaborationspartner.
3. Wenn Grenzen des Menschen erkannt werden (z. B. etwas ist zu schwer), sollte dies reflektiert und die Aufgabe unkompliziert übernommen werden. Hierfür ist eine Simulation der nächsten Aufgabenschritte erforderlich, unter Berücksichtigungen von physischen und kognitiven Charakteristiken.
4. Für neuartige Situationen sollte abstraktes Wissen über wichtige Prinzipen der Welt zur Verfügung stehen, insbesondere im Rahmen des anvisierten Aufgabenrahmens. So können relevante Zusammenhänge neuer Situationen und Handlungsalternativen entwickelt werden. So ist es beispielsweise meist sinnvoll, erst große Teile zusammenzubauen und dann kleinere Arbeiten daran vorzunehmen oder den Boden für Fixierungsarbeiten zur Stabilisierung zu nutzen. Diese Art Vorwissen soll hier mit Weltmodellen umschrieben werden.

Mit diesen vier Ansätzen wäre es möglich, viele bestehende Probleme aus dem Feld der Mensch-Roboter-Kollaboration direkt zu adressieren. Um Roboter mit diesen Fähigkeiten auszustatten, stellt sich die Frage, wie genau Menschen in der Lage sind, andere zu antizipieren. Welche kognitiven Mechanismen eignen sich dazu, diese Fähigkeiten zu erklären und können wir diese Mechanismen algorithmisch beschreiben?

13.3 Kognitive Mechanismen zur Antizipation Anderer

13.3.1 Mentale Modelle

Viele der oben genannten Modelle und Repräsentationen beziehen sich auf das übergeordnete Konzept von mentalen Modellen. Ein mentales Modell ist die Repräsentation eines Gegenstandes (bzw. technischen Gerätes) oder eines Prozesses. Da Lebewesen dazu neigen, die in der Welt vorhandene Information stark zu filtern, kann ein mentales Modell immer nur ein Ausschnitt der Wirklichkeit sein, und auch nur so kann es sinnvoll auf ähnliche Situationen angewandt werden. Bei „guten" mentalen Modellen bleiben die relevanten Aspekte, insbesondere ihre Struktur, erhalten.

Es gibt viele theoretische Ansätze zu mentalen Modellen, doch wie werden mentale Modelle aufgebaut, verwendet oder verändert? Ein reduzierter und gut anwendbarer Ansatz stellt der CER des Cycle-of-Model-Updation-for-Decision-Making-Ansatzes dar (Li und Maani 2011). CER steht für Conceptualization-Experimentation-Reflection. Für die Conceptualization-Phase wird ein Verständnis bzw. eine Beschreibung der aktuellen Situation (in Form einer Repräsentation) herangezogen und mental das Ergebnis einer potenziellen Entscheidung bzw. die in Bezug stehende Handlung simuliert. Während der Experimentation-Phase werden eine Entscheidung bzw. Interventionen, die aus dem mentalen Modell abgeleitet wurden, ausgewählt und getestet. In der Reflection-Phase wird das Ergebnis der Experimentation-Phase reflektiert bzw. das perzeptuelle Feedback verarbeitet. Wenn das erwartete Ergebnis der Intervention erreicht wird, wird die Entscheidung konsolidiert. Ist das Feedback unerwartet bzw. unterscheidet es sich stark von der Erwartung, wird das aktuelle mentale Modell aktualisiert bzw. mit der neuen Information angereichert.

Diese Form der Verwendung bzw. des Aufbaus und Umbaus mentaler Modelle wurde in einem kognitiven Modellierungsansatz umgesetzt. In zwei verschiedenen Aufgaben wurden die Modelldaten mit Humandaten verglichen (Prezenski et al. 2017). Es handelte sich um das Erlernen der Bedienung verschiedener Smartphone-Apps. Im Rahmen der ersten App sollten bestimmte Produkte in der App gefunden und ausgewählt werden. In der zweiten App sollte nach bestimmten Immobilien gesucht werden. In beiden Studien wurde nach der Hälfte des Versuchs der Aufbau der Menüstruktur verändert, um Softwareupgrades nachzubilden. Der Modellansatz sollte das Erlernen des Umgangs mit der App abbilden. Die resultierenden Interaktionszeiten zu Beginn des Versuches als auch die Interaktionszeiten, die sich nach dem Strukturwechsel zeigten, wurden abhängig von Modell und

Versuchsteilnehmern verglichen. Je nachdem, wie aufwendig die neue Menüstruktur in das Mentale Modell zu integrieren ist, dauert das Umlernen und die Produktsuche sowohl beim Modell als auch den Versuchsteilnehmern länger oder kürzer. Die Reaktionszeiten zeigten eine hohe Übereinstimmung der Humandaten mit den Modelldaten, sogar über unterschiedliche Apps und Menüstrukturen hinweg.

Diese Ergebnisse lassen vermuten, dass die verwendete Umsetzung mentaler Modelle für technische Systeme einen guten Ansatz bietet, um menschliche Nutzer nachzubilden. Doch für Anwendungen in der Robotik reicht es nicht aus, nur eine Interaktionsaufgabe mit einem technischen System abzubilden. Welche Arten von mentalen Modellen wären für einen autonomen Agenten notwendig und welche Voraussetzungen müssten sie erfüllen?

13.3.2 Person Model Theory

Wie oben erwähnt, können bereits Vorschulkinder kooperatives Verhalten zeigen, ohne über die kognitiv aufwändige Fähigkeit der „Theory of Mind“ zu verfügen.

Albert Newen entwickelte den Person-Model-Theory-(PMT-)Ansatz (Newen und Schlicht 2009; Newen 2015), der berücksichtigt, dass wir normalerweise in einer Interaktion mit der Umgebung involviert sind, wenn wir versuchen, andere zu verstehen. Der PMT-Ansatz basiert darauf, dass das Verstehen einer anderen Person den Aufbau von verschiedenen Arten von Modellen erfordert.

Eine Situation zu verstehen ist oft schon ausreichend dafür, ein Verständnis für die Intention und nächsten Handlungen einer anderen Person abzuleiten (z. B. in der Kantine, mit Blick auf die Kasse oder Blick auf das Menü). In diesem Zusammenhang werden insbesondere sensomotorische Fähigkeiten hervorgehoben. Diese sogenannten Situationsmodelle beinhalten, welche Bedeutung eine Situation für den Agenten hat und was die nächsten Schritte und Handlungen bzw. Ziele in einer Situation normalerweise sind.

Hier wird zunächst die Situation von der eigenen Perspektive aus betrachtet. Man versetzt sich selbst in die Situation: Was würde ich normalerweise als nächstes tun, bzw. was wäre mein Ziel?

Zusätzlich nennt der Autor „Personenmodelle“, welche stärker individuelle Emotionen oder individuelle Eigenschaften, auch spezifische Einschränkungen und ähnliches, umfassen. Je nach Verfügbarkeit können ausschließlich Situationsmodelle oder aber auch eine Wechselwirkung von Situationsmodellen und Personenmodellen für das Verständnis anderer herangezogen werden. Unser Verständnis von anderen nutzt mehrere mögliche Modelle zur Orientierung und selektiert die hilfreichsten Modelle, um eine andere Person zu verstehen.

Zusätzlich werden in der PMT auch „Selfmodelle“ genannt. Hier wird davon ausgegangen, dass diese als Grundlage dienen, um bei fehlender ergänzender Information zunächst davon auszugehen, wie man selbst vorgehen würde.

Für einen Modellansatz von Mensch-Roboter-Kollaboration ist es nicht nur wichtig nachzuvollziehen, was jemand als nächstes tun möchte und warum diese Person etwas Bestimmtes tut. Sondern es ist auch relevant, ob ich als Mensch oder Roboter den Partner dabei unterstützen kann. Analog zu dem Situations- und Personenmodell sollte der Roboter ein Modell davon haben, was er als Agent in der aktuellen Situation als nächsten Handlungsschritt bewirken kann, um das gemeinsame Ziel weiter voran zu bringen. Diese drei Modellarten – Situationsmodell, Personenmodell und Selfmodell – sollen hier vorgestellt und anhand von zwei Beispielen gezeigt werden, wie dies umsetzbar wäre.

13.4 Kognitiver Modellierungsansatz von Situationsmodell, Personenmodell und Selfmodell

13.4.1 Voraussetzungen der Modellierungsmethode

Für einen entsprechenden Modellansatz ist es notwendig, eine Methode zu wählen, die gewisse Voraussetzungen erfüllt. Die Modellierungsmethode muss über symbolische Repräsentationen verfügen, um nachvollziehbar zu sein und Repräsentationen miteinander vergleichen zu können. Darüber hinaus sollte die Modellierungsmethode kognitive Verarbeitungscharakteristiken berücksichtigen, um nachvollziehen zu können, ob eine Person aufgrund von begrenzten kognitiven Ressourcen (z. B. visuelle Aufmerksamkeit, Arbeitsgedächtnis, o. ä.) vorliegende Informationen überhaupt in vollem Umfang verarbeiten konnte. Da die Dauer von Handlungen eine wichtige Komponente bei der Kollaboration ausmacht, sollten zeitliche Anhängigkeiten bei kognitiver Verarbeitung berücksichtigt werden. Die Modellierungsmethode sollte ferner über verschiedene menschliche Lernmechanismen (Instanzenlernen aber auch Optimalisierungslernen, z. B. Utility Learning) und Gedächtnisfunktionen verfügen (Deklaratives Gedächtnis, Prozedurales Gedächtnis, Arbeitsgedächtnis). Darüber hinaus ist es wichtig, eine Methode zu wählen, die Echtzeitfähigkeit aufweist und flexibel auf neue Situationen und Umstände reagieren kann. Vielversprechende Modellierungsmethoden hierfür sind beispielsweise kognitive Architekturen, wie z. B. ACT-R (Anderson et al. 2004).

Das Ziel ist es, einem autonomen Agenten die Fähigkeit zu verleihen, andere antizipieren zu können. Insbesondere bedeutet dies, Vorhersagen über die nächste Aktion zu generieren, Handlungen zu erklären und gemeinsame Handlungen synchronisieren zu können. Zusätzlich ist es notwendig, individuelle Eigenschaften des Partners zu erlernen, um die Antizipation zu verbessern. So ist es möglich, nächste Handlungsschritte des anderen vorherzusagen und genug Zeit zur Verfügung zu haben, eigene Handlung zu planen und zu koordinieren oder den anderen zu unterstützen, z. B. aus dem Weg zu fahren, etwas zu halten oder etwas zu befestigen.

13.4.2 Beispiele für antizipierende Assistenzsysteme

Zwar gibt es bisher kaum Beispiele aus dem Bereich der Robotik, aber das Forschungsfeld der kognitiven Assistenz ist interessant für die vorliegenden Fragen. In einem Projekt zur kognitiven Assistenz war das Ziel, Piloten in kritischen Situationen zu antizipieren. Hierfür wurde ein Situationsmodell entwickelt, welches teilweise mit Aspekten des Personenmodells des Piloten ergänzt wurde (Klaproth et al. 2019). In diesem Projekt wurde auf die kognitive Architektur ACT-R zurückgegriffen. Situationen und der jeweilige nächste mögliche Handlungsschritt werden mithilfe von flexiblen Produktionen umgesetzt. Nach jedem Verarbeitungszyklus (ca. 50 ms) wird erneut evaluiert, welche Produktionen auf die aktuelle Situation passen und die jeweils beste ausgewählt. Dem Modell werden die Informationen der Flugzeugsensoren als auch die Handlungsaktionen des Piloten zugeführt. Zusätzlich können zu bestimmten Ereignissen Informationen über die individuellen Reaktionen des Piloten abgerufen werden. Der Fokus liegt hier zunächst auf der Erhebung von ERPs (Event Related Potenzials), die eine Überraschungsreaktion des Piloten wiedergeben. So konnte evaluiert werden, ob beispielsweise eine Warnung von dem Piloten überhaupt wahrgenommen oder aktiv ignoriert wurde. Das adaptive Modell kann so bereits sehr verlässlich das Verhalten bzw. die Entscheidungen des Piloten vorhersagen und kritische Situationen erkennen.

Weitere kognitive Assistenzsysteme werden derzeit entwickelt, wie im halbautonomen Fahren, wo anhand von Blickbewegungsmessung und Sensordaten des Fahrzeuges ein Situationsmodell und ein Personenmodell des Fahrers antizipiert werden (Scharfe und Russwinkel 2019). Hier geht es insbesondere um die Antizipation des Situationsbewusstseins (Endsley 1995), nachdem eine Übernahme des Fahrzeuges durch den Fahrer initiiert wurde. Wie schnell lässt sich eine ausreichende Repräsentation der aktuellen Situation aufbauen und weisen die Blickdaten die erwarteten Muster auf? Zusätzlich kann so auch die Wahrnehmung der Komplexität einer Situation antizipiert werden.

Die nächsten Forschungsschritte in diesem Feld werden unter anderem darauf abzielen, alternative mentale Modelle zu entwickeln und das jeweils beste auswählen. Dieser Prozess würde nicht nur eine bessere Antizipation des Kooperationspartners ermöglichen, sondern auch alternative Erklärungen für unerwartetes Verhalten generieren können.

13.5 Flexible Task Allocation

Die Möglichkeit, den anderen zu antizipieren und antizipiert zu werden, entlastet den Menschen davon, ständig die Kontrolle über alle Abläufe aufrecht zu erhalten. Manche Aufgaben kann der autonome Agent möglicherweise besser erfüllen als der Mensch, wie z. B. über einen längeren Zeitraum hohe Gewichte zu halten. Hier wäre es sinnvoll, wenn die Maschine die Kontrolle übernimmt und der Mensch möglicherweise feinmotorische Aufgaben steuert und sich an die Vorgaben des Agenten anpasst. In anderen Situationen übernimmt der Mensch die Kontrolle, aber beide Akteure berücksichtigen mögliche

Probleme, die in der jeweiligen Situation entstehen könnten (beispielsweise, dass der Mensch einen Balken übersehen hat und mit dem Kopf daran stoßen könnte, bzw. dass die Maschine mit einer neuen Situation nicht umgehen kann).

Gäbe es ein gegenseitiges Verständnis einer Situation und des Gegenübers, würden derartige Probleme antizipiert und frühzeitig unkompliziert gelöst werden können.

Bei der Mensch-Mensch-Zusammenarbeit kennen wir diese Art der Kooperation. Auch hier übernimmt nicht ausschließlich eine Person die Kontrolle, sondern die Kontrolle wechselt zwischen beiden Partnern unter Berücksichtigung einer geteilten Zielvorstellung. Wenn einem Lehrling eine neue Aufgabe beigebracht wird, überlassen wir dem Lehrling auch zu bestimmten Zeitpunkten die Kontrolle über die Aufgabe. Je nach Lernstand kann die Kontrolle über größere Zeiträume komplett dem Lehrling überlassen werden, aber der Lehrling wird weiterhin beobachtet werden bis sich ein mögliches Problem abzeichnet und der Meister wieder übernimmt. Aber beide Partner antizipieren jeweils den anderen, es gibt also die Unterscheidung zwischen „Kontrolle übernehmen“ und „den andern zu antizipieren“. Ein flexibler Wechsel der Kontrolle, bzw. ein Kontinuum an Kontrolle von einer weniger bis stärker ausgeprägten Form, ist uns Menschen wohlbekannt und mit geringen Aufwand zu realisieren. Dies bedeutet, dass wir eine gute Vorstellung davon haben, ob der andere seinen Teil der Aufgabe übernehmen kann und auch, wie stark diese Teilaufgabe weiter beobachtet werden muss. Hervorzuheben ist hier, dass Kontrolle nicht als etwas Absolutes gesehen wird, sondern als dynamischer Zustand, der für den Nutzer antizipierbar und akzeptabel ist. Es darf bei dem menschlichen Nutzer nicht der Eindruck entstehen, bevormundet zu werden. Die Übernahme von Kontrolle muss transparent kommuniziert und begründet werden. Eine solche Art der Zusammenarbeit von Menschen und autonomen Agenten würde einen Paradigmenwechsel der bisherigen Mensch-Technik-Interaktion mit sich bringen.

13.6 Interactive Task Learning

Die genannten Modellformen, über die ein Agent verfügen sollte, um einen Partner antizipieren zu können, liefern auch die Grundlagen für eine flexiblere und schnellere Form des Lernens neuer Aufgaben und Situationen. Wir Menschen lernen durch die Interaktion mit unserer Umgebung. Wir lernen, welche Änderungen durch unser Handeln in der Umgebung verursacht werden. Für einen autonomen Agenten ist es daher naheliegend, dass ein Verständnis dafür erlangt wird, welche Veränderungen der Agent selbst verursacht, welche Veränderungen durch einen anderen Agenten bewirkt wurde und welche Veränderungen durch Prozesse in der Umgebung initiiert wurden. Diese Art der Zuschreibung von Veränderungen bereitet die Basis, auf welcher wirkliches Lernen von Zusammenhängen ansetzt und damit über pures Pattern Matching hinausgeht.

In der interdisziplinären Forschungslandschaft wird der Frage nachgegangen, welches die effektivsten und natürlichsten Methoden des Lernens für Mensch, Roboter und AI-Agenten sind (Thomaz et al. 2019). Die effektivste Methode ist abhängig von der Art

der Aufgabe. Unter anderem werden verschiedene Arten des Lernens genannt: Self-Exploration, Structured Discovery, Apprenticeship (Lernen durch Imitation) und Explicit Instruction (explizite Kommunikation zwischen Lehrer und Schüler mit dem Ziel, eine neue Aufgabe zu erlernen). Alle diese Formen des Lernens erfordern den Aufbau bzw. das Heranziehen von mentalen Modellen über Situationen und Aufgaben und von Modellen über Personen.

13.7 Diskussion

Es wurde diskutiert, über welche Fähigkeiten ein Roboter verfügen muss, um in der Lage zu sein, den Nutzer auch in unbekannten Situationen sinnvoll unterstützen zu können. Hierfür sind unterschiedliche mentale Modelle notwendig, die sowohl Situation und Aufgabe, individuelle Personeneigenschaften und ein Modell des Selbst einschließen.

Um eine flexible Kollaboration von Mensch und Roboter zu realisieren, werden diese Modelle verwendet, um das Verhalten des anderen zu interpretieren und nächste Handlungsschritte zu prädizieren. Erste Ansätze einer Umsetzung in dem Feld der kognitiven Assistenz wurden kurz adressiert.

Die mentalen Modelle bieten darüber hinaus die Möglichkeit, einen Roboter durch die Interaktion mit der Umgebung lernen zu lassen. Die entsprechenden erworbenen mentalen Modelle kann er anschließend in neuen Situationen einsetzen.

Das umrissene Ziel für Roboter, die als Kollaborationspartner eingesetzt werden können, formt sich erst langsam mit dem schrittweisen Zuwachs an verliehenen intelligenten Fähigkeiten. Doch können schon erste Schritte in diese Richtung einen deutlichen Mehrgewinn für eine neue Generation von technischen Systemen und einer neuen Form der Mensch-Roboter-Interaktion bedeuten. Viele der derzeit bestehenden Herausforderungen aus der Mensch-Roboter-Kollaboration lassen sich mit derartigen Ansätzen elegant adressieren.

Die Sicherheitstechnik könnte flexibilisiert werden, da der autonome Agent für den Menschen antizipierbar wäre und die Handlungen des Menschen prädiziert werden würden. Eine Umsetzung, wie mit fehlbarer Automation umzugehen ist, wird auch mit diesem Ansatz adressiert. Ein künstliches kognitives System, das versucht, Zusammenhänge zu verstehen und flexibel auf Situationen zu reagieren, beinhaltet bereits die Möglichkeit, dass es mehrere mögliche Lösungen eines Problems gibt und nicht ein Agent die einzig richtige Lösung bereithält.

So gesehen, würde ein kognitiver Systemansatz auf der einen Seite große Vorteile bringen. Auf der anderen Seite müssen neue Herausforderungen angenommen werden, beispielsweise um eine gewisse Verhaltenspermanenz zu gewährleisten und gewisse wichtige Systemeigenschaften sicher aufrecht zu erhalten.

Literatur

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036–1060.

Buxbaum, H., & Sen, S. (2018). Kollaborierende Roboter in der Pflege – Sicherheit in der Mensch-Maschine-Schnittstelle. In O. Bendel (Hrsg.), *Pflegeroboter*. Wiesbaden: Springer Gabler.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32–64.

Fiebich, A. (2018). Three dimensions in human-robot cooperation. In M. Coeckelbergh, M. Funck, J. Seibt & M. Norskov (Hrsg.), *Envisioning robots in society – Power, politics, and public space, proceedings of robophilosophy 2018/TRANSOR 2018* (Frontiers in artificial intelligence and applications, S. 147–155). Amsterdam: IOS Press.

Fodor, J. A. (1992). A theory of the child's theory of mind. *Cognition, 44*(3), 283–296.

Klaproth, O., Halbrügge, M., & Russwinkel, N. (2019). ACT-R model for cognitive assistance in handling flight deck alerts. In *Proceedings of the 17th international conference on cognitive modeling*, Montreal.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences, 40*, e253. https://doi.org/10.1017/S0140525X16001837.

Li, A., & Maani, K. (2011). Dynamic decision-making, learning and mental models. In *Proceedings of the 29th international conference of the system dynamics society* (S. 1–21). Washington, DC: System Dynamics Society.

Mahy, C. E., Moses, L. J., & Pfeifer, J. H. (2014). How and where: Theory-of-mind in the brain. *Developmental Cognitive Neuroscience, 9*, 68–81.

Newen, A. (2015). Understanding others: The person model theory. In T. Metzinger & J. Windt (Hrsg.), *Open-Mind* (Bd. 26, S. 1–28). www.open-mind.net. https://doi.org/10.15502/9783958570320.

Newen, A., & Schlicht, T. (2009). Understanding other minds: A criticism of Goldman's simulation theory and outline of the person model theory. *Grazer Philosophische Studien, 79*, 209–242.

Onnasch, L., Maier, X., & Jürgensohn, T. (2016). *Mensch-Roboter-Interaktion – Eine Taxonomie für alle Anwendungsfälle* (1. Aufl., S. 1–12). baua: Fokus, Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. https://doi.org/10.21934/baua:fokus20160630.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515–526.

Prezenski, S., Brechmann, A., Wolff, S., & Russwinkel, N. (2017). A cognitive modeling approach to strategy formation in dynamic decision making. *Frontiers in Psychology, 8*(1335), 1–18.

Scharfe, M., & Russwinkel, N. (2019). A cognitive model for understanding the takeover in highly automated driving depending on the objective complexity of non-driving related tasks and the traffic environment. In *Proceedings of the 41th annual cognitive science society meeting*.

Thomaz, A. L., Lieven, E., Cakmak, M., et al. (2019). Interaction for task instruction and learning. In K. A. Gluck & J. E. Laird (Hrsg.), *Interactive task learning: Humans, robots, and agents acquiring new tasks through natural interactions* (Strüngmann Forum Reports, Bd. 26, J. R. Lupp, series editor, S. 91–110). Cambridge, MA: MIT Press.

Nele Rußwinkel studierte Cognitive Science an der Universität Osnabrück und an der Middle East Technical University in Ankara. Ihre Masterarbeit über visuelle Aufmerksamkeit schloss sie an der Charité und der Humboldt-Universität zu Berlin ab. Nele Rußwinkel begann ihre wissenschaftliche

Karriere bei der VW-Nachwuchsgruppe ModyS und promovierte an der DFG-geförderten Graduiertenschule Prometei über quantitative Modelle der Zeitschätzung. Sie war 2012–2016 Mitglied des Vorstandes der Gesellschaft für Kognitionswissenschaften und ist seit 2012 Mitglied des internationalen Steering Board of Cognitive Modeling. Seit 2013 leitet Sie das Fachgebiet für „Kognitive Modellierung in dynamischen Mensch-Maschine-Systemen“ an der Technischen Universität Berlin.