

Uwe Lorenz

Reinforcement Learning

Aktuelle Ansätze verstehen –
mit Beispielen in Java und Greenfoot

EBOOK INSIDE

 Springer Vieweg

Reinforcement Learning

Uwe Lorenz

Reinforcement Learning

Aktuelle Ansätze verstehen – mit
Beispielen in Java und Greenfoot

Dipl. Inf. Uwe Lorenz
Neckargemünd, Baden-Württemberg
Deutschland

ISBN 978-3-662-61650-5 ISBN 978-3-662-61651-2 (eBook)
<https://doi.org/10.1007/978-3-662-61651-2>

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature 2020

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag, noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Planung: Martin Börger

Springer Vieweg ist ein Imprint der eingetragenen Gesellschaft Springer-Verlag GmbH, DE und ist ein Teil von Springer Nature.

Die Anschrift der Gesellschaft ist: Heidelberger Platz 3, 14197 Berlin, Germany

Vorwort

„Man muss die Dinge so einfach wie möglich machen. Aber nicht einfacher.“ (A.Einstein).

Ziel des Buches ist es, nicht nur eine lose Auflistung gängiger Ansätze des „Reinforcement Learnings“ zu liefern, sondern auch einen inhaltlich zusammenhängenden Überblick über dieses faszinierende Gebiet der künstlichen Intelligenz zu geben. Gleichzeitig sollen die Konzepte einem möglichst großen Leserkreis aufgeschlossen und z. B. auch Impulse für den Schulunterricht ermöglicht werden. Dieser scheinbare Widerspruch löst sich auf, wenn es gelingt, einen umfassenden Grundriss zu zeichnen, der zwar die wesentlichen Ideen beinhaltet, dabei aber von komplexeren „Real-World“-Anwendungen absieht. Jeder Lehrer weiß, dass eine vereinfachte Darstellung nicht falsch ist, wenn sie innere Konsistenz und Zweckmäßigkeit aufweist. Im Sinne der „Spirale des Begreifens“, kann ein solcher Grundriss als natürlicher Ausgangspunkt für tiefere Einsichten, weitere Untersuchungen und praktische Anwendungen – auch mit Hochleistungswerkzeugen – dienen. Um dieses Buch zu verstehen, sollten die Mittel der höheren Schulmathematik ausreichen. Zudem benötigen Sie einige grundlegende Kenntnisse in der Programmiersprache Java.

Das Forschungsgebiet „Künstliche Intelligenz“ war Anfang der 2000er Jahre, also in der Zeit als ich mich damit in Berlin besonders intensiv beschäftigte, in einem fundamentalen Umbruch. Der Ansatz der „good old fashioned artificial intelligence“ (GOFAI), der im Prinzip davon ausgeht, dass Kognition durch objektive Modelle der Außenwelt und eine darauf angewandte logische Symbolverarbeitung entsteht, war in einer Krise. Gleichzeitig sorgte die Entwicklung der allgemein zugänglichen Rechenleistung bereits dafür, dass mit künstlichen neuronalen Netzen immer erstaunlichere Resultate produziert wurden. Es gab damals die Fragestellung von „Konnektionismus“ vs. „Symbolischer K.I.“, also die Frage nach der Leistungsfähigkeit von verteilten, sub-symbolischen Repräsentationen einerseits und formalisiertem Spezialwissen, aus dem logisch geschlossen wird, andererseits. Der fundamentale Unterschied zwischen den Ansätzen besteht allerdings nicht nur in der Art der Repräsentation von Wissen. Es geht vielmehr um einen Unterschied in der Stellung der Basiskonzepte „Beschreibung“ und „Verhalten“. Welches der beiden Prinzipien sollte eine dominante Position einnehmen?

Die Bedeutung des verhaltensbasierten Ansatzes wurde mir damals bei dem Versuch klar, im Zusammenhang mit dem Thema „Mustererkennung“ einen objektiven Begriff von „Ähnlichkeit“ zu bilden. Diese Bemühungen wurden zwar enttäuscht, führten aber zu einer grundlegenden Einsicht: „Ähnlichkeit“ und somit auch „Klassifikation“ und damit auch „Begriffe“ allgemein sind nichts anderes als evolutionär gewachsene Mittel, um die für das Leben notwendigen Unterscheidungen treffen zu können. Eine gänzlich „zweckfreie Beschreibung“ ist im Prinzip unmöglich. Dies führte mich schließlich zu der Schlussfolgerung, dass es bei den Bemühungen der künstlichen Intelligenz nicht zuerst darum gehen kann, logisch „denkende“, sondern praktisch „handelnde“ Maschinen zu konstruieren.

Bei verhaltensbasierten Systemen muss neues Wissen aus der Differenz von Vorhersage und Beobachtung generiert werden. Bei diesem „Trial and Error“ entstehen sehr viele Beobachtungen deren Bedeutung sich erst später mehr oder weniger zufällig herausstellt. Eine zentrale Herausforderung beim Reinforcement Learning ist es, diese Beobachtungen für die Optimierung der Agentensteuerung nutzbar zu machen. In den letzten Jahren wurden auf diesem Gebiet große Fortschritte gemacht und das gesamte Gebiet entwickelt sich weiterhin außerordentlich dynamisch.

Gegenwärtig erreichen wir einen Punkt, an dem wir erkennen, dass wissens- bzw. modellbasierte Methoden der intelligenten Verhaltenssteuerung sehr wohl auch dienlich sein können. Nämlich dann, wenn es gelingt, diese Beobachtungen so zu verarbeiten, dass sie mehrfach genutzt werden können. Der künstliche Agent „weiß“ dann z. B. schon, was passieren würde, wenn eine Aktion ausgeführt werden würde und muss nicht tatsächlich agieren. Solche aus der Erfahrung entstandenen Modelle erlauben zweckmäßige Simulationen, die Beobachtungen liefern, welche zwar nicht in dem Maße abgesichert sind wie echte Erfahrungen, allerdings sehr viel billiger und in größerer Zahl generiert werden können. Mithilfe der zu solchen Modellen kondensierten Beobachtungen kann eine Art „virtuelle Praxis“ ermöglicht werden, die dann die modellfreien Steuerungen optimiert.

Dieses Buch möchte ich meinen drei Söhnen Jonas, Elias und Paul widmen, die sich im Laufe der letzten zwei Jahre schon fast daran gewöhnt hatten, dass der Papa am Wochenende im Arbeitszimmer verschwindet. Ich verspreche, dass wir jetzt, nachdem dieses Projekt erstmal zu seinem – vielleicht vorläufigen – Abschluss gekommen ist, wieder mehr gemeinsam unternehmen werden. Zudem möchte ich natürlich meiner Frau Anja danken. Ohne ihre Bereitschaft dieses Buchprojekt zu unterstützen und mich auch von vielen alltäglichen Tätigkeiten zu entlasten, wäre dieses nebenberufliche Projekt nicht möglich gewesen. Ich hoffe, dass wir die Rollen in der kommenden Zeit auch mal wieder tauschen können.

Außerdem möchte ich mich bei der SRH Fernhochschule (<https://www.mobile-university.de>; 3.5 2020) bedanken. Am praktischen Beginn dieses Projektes stand ein Abschnitt des Studienbriefs für den Zertifikatskurs „Künstliche Intelligenz:

Anwendungen und neue Geschäftsfelder“ (<https://akademie.spiegel.de/kuenstliche-intelligenz.html>; 3.5.2020) zum Thema „Reinforcement Learning“. Die freundliche Erlaubnis Textabschnitte, die ich für den betreffenden Studienbrief geschrieben hatte, in diesem Buch zu verwenden, hat den Startschuss für dieses Projekt gegeben und die Arbeit an dem Buch deutlich erleichtert.

Meiner Schwester Ulrike alias „HiroNoUnmei“ (<https://www.patreon.com/hironounmei>; 3.5.2020) möchte ich noch ganz herzlich für die lustigen Hamster-Illustrationen an den jeweiligen Kapitelanfängen danken.

Zu den Begleitmaterialien (Java-Programme, Erklärvideos usw.) gelangen Sie über die Produktseite des Buchs unter (<https://www.springer.com/de/book/9783662616505>; 3.5.2020) und über Seite (<https://www.facebook.com/ReinforcementLearningMitJava>; 3.5.2020). Mit letzterer gelingt es vielleicht auch eine kleine Community zu bilden und einen Austausch über die Inhalte des Buchs und das Thema „Reinforcement Learning“ in Gang zu bringen. Posten Sie gerne inhaltliche Beiträge zum Thema oder interessante Ergebnisse. Dort finden Sie ggf. auch eine Möglichkeit Verständnisfragen zu stellen oder offen gebliebene Punkte anzusprechen.

Neckargemünd
Mai 2020

Uwe Lorenz

Einleitung

„Verständnis wächst mit aktiver Auseinandersetzung: Etwas zu ‚machen‘, zu beherrschen, bedeutet zugleich besseres Verstehen. Angewandt auf die Erforschung geistiger Prozesse führt das auf die Nachbildung intelligenten Verhaltens mit Maschinen.“ (H.-D. Burkhardt¹).

Zusammenfassung

In diesem Einleitungsabschnitt wird dargestellt, worum es in diesem Buch gehen soll und für wen es gedacht ist: das Thema „Reinforcement Learning“, als ein sehr spannendes Teilgebiet der Künstlichen Intelligenz bzw. des maschinellen Lernens, soll in einer Form präsentiert werden, die Einsteigern zügig die wichtigsten Ansätze und einige der zentralen Algorithmen vermittelt und auch eigene Experimente damit ermöglicht.

Es soll ein Sachbuch sein, das Lernenden oder Interessierten, die sich inhaltlich mit diesem Gebiet der künstlichen Intelligenz beschäftigen möchten (oder müssen) einen praktischen Zugang in die Theorie und die Funktionsweise der lernfähigen Agenten ermöglicht. Es ist aber auch für Techniker oder Lehrpersonen gedacht, die sich weiterbilden und eigene Experimente oder Übungen durchführen möchten, auch im Rahmen ihrer jeweiligen Lehrveranstaltungen.

Über einige „philosophische“ Fragestellungen oder Kritiken am Forschungsgebiet der KI wird kurz reflektiert.

Dieses Buch ist vielleicht nicht ganz ungefährlich, denn es geht in ihm um lernfähige künstliche Agenten. Auf dem Digital-Festival „South by Southwest“ im US-Bundesstaat Texas sagte Elon Musk im Jahr 2018 „Künstliche Intelligenz ist sehr viel gefährlicher als Atomwaffen“. Vielleicht ist dies etwas dick aufgetragen, allerdings ist es sicherlich

¹Prof. Dr. Hans-Dieter Burkhard war einer meiner ehemaligen Hochschullehrer an der Humboldt-Universität Berlin, - in 2004, 2005 und 2008 Weltmeister in der Four-Legged Robots League beim RoboCup mit dem „German Team“.

von Vorteil, wenn möglichst viele Menschen verstehen und beurteilen können, wie diese Technik funktioniert. Das ermöglicht nicht nur zu beurteilen, was die Technik leisten kann und sollte, sondern auch, ihren weiteren Entwicklungsweg mit zu gestalten. Der Versuch einer Einhegung von wertvollem Wissen wäre in der Welt von heute sicherlich auch ein vergebliches Unterfangen.

Eine besondere Bedeutung in diesem Zusammenhang besitzt die Thematik des „Reinforcement Learning“, denn hierbei geht es um autonome Maschinen, die ähnlich wie biologische Lebewesen in einer bestimmten Umgebung vor Probleme gestellt werden und durch aktive Prozesse ihr Verhalten optimieren. Beim „Reinforcement Learning“ handelt es sich aber auch um einen der faszinierendsten Bereiche des maschinellen Lernens, der im deutschsprachigen Raum oft noch vergleichsweise wenig behandelt wird, obwohl immer wieder spektakuläre Erfolgsmeldungen aus diesem Gebiet der Künstlichen Intelligenz nicht nur das Fachpublikum, sondern auch die breite mediale Öffentlichkeit erreichen. Einige Beispiele: Eine der am längsten untersuchten Domänen in der Geschichte der künstlichen Intelligenz ist das Schachspiel. Es gelang schon vor geraumer Zeit Programme zu schreiben, die menschliche Champions schlagen konnten, diese nutzten allerdings ausgefeilte, spezialisierte Suchtechniken und handgefertigte Auswertungsfunktionen sowie enorme Datenbanken mit archivierten Spielzügen. Die Programme der „Alpha Zero“-Reihe von „Google DeepMind“ dagegen können innerhalb weniger Stunden allein durch Lernen aus Spielen gegen sich selbst übermenschliche Leistungen erreichen ohne jegliches menschliches Vorwissen (Silver et al. 2017). Es übertrifft schließlich auch die bislang besten Programme, welche die zuvor erwähnten Methoden nutzten, deutlich. Spektakulär hierbei ist auch, dass das System nicht nur auf ein einzelnes Spiel wie Schach festgelegt ist, sondern in der Lage ist, alle möglichen Brettspiele selbsttätig zu erlernen. Darunter zählt auch das wohl in vielfacher Hinsicht komplexeste Brettspiel „Go“, welches in Asien schon seit tausenden von Jahren gespielt wird und eigentlich eine Art „intuitive“ Interpretation der Situation des Spielbretts erfordert. Alpha Zero kann sich an vielfältige hochkomplexe Strategiespiele überaus erfolgreich anpassen und benötigt dafür kein weiteres menschliches Wissen, – es reichen hierfür allein die Spielregeln. Das System wird damit zu einer Art universellem Brettspiel-Lösungssystem. Wie konnte Google deepMind zu diesen Erfolgen kommen? Bis vor kurzem bestand noch große Übereinstimmung darin, dass ein „intuitives“ Spiel mit einer derartig großen Zahl von Zuständen wie bei Go, in absehbarer Zeit nicht von seriellen Computern zu bewältigen ist. Hinzu kommt, dass die Maschine, die in der Lage ist in solchen Spielen menschliche Champions zu schlagen, völlig ohne das in Jahrtausenden angesammelte menschliche Wissen über die Spiele auskommt.

Ähnlich spektakulär sind auch die Leistungen innerhalb „dynamischer“ Umgebungen, bspw. die von „Deep Q Networks“ beim eigenständigen Erlernen beliebiger Atari-Arkade Games (Kavukcuoglu et al. 2015) oder die Ergebnisse im Bereich der Robotik, wo Systeme durch selbstständiges Ausprobieren erlernen, komplexe Bewegungen, wie Greifen, Laufen, Springen usw. erfolgreich auszuführen und Aufgaben in vorbereiteten

Arenen meistern, wie sie z. B. in den zahlreichen Robotik-Wettbewerben, die es mittlerweile für alle möglichen Anforderungsniveaus und Altersstufen gibt, gestellt werden.

Eine große Rolle spielen hierbei sogenannte „tiefe“ künstliche neuronale Netze mit besonderen Fähigkeiten bei der Generalisierung. Es zeigen sich dabei erstaunliche Blüten, wie z. B. beim automatischen Verwandeln von Fotos in täuschend „echte“ Gemälde nach dem Stil eines bestimmten Künstlers, wobei hier zwei solcher „tiefen“ Netze so interagieren, dass eines davon „produziert“ und ein anderes die Produkte „kritisiert“. Unter <https://deepart.io> (Febr. 2020) kann man sich davon einen eigenen Eindruck verschaffen.

In letzter Zeit wurden aufwendige Machine Learning Frameworks wie TensorFlow von einschlägigen US-Playern teilweise kostenlos zur Verfügung gestellt. Warum schenken Konzerne wie Google usw. solche aufwendigen Produkte der Allgemeinheit? Es ist anzunehmen, dass es darum geht Standards zu definieren und Abhängigkeiten zu schaffen. Philanthropische Großzügigkeit ist bei solchen Kapitalgesellschaften sicherlich nicht als primärer Beweggrund anzunehmen. Das Buch möchte auch Mut machen, das „Räderwerk“, das sich hinter den „Frameworks“ verbirgt, genauer anzuschauen und zu verstehen.

Konkrete Implementationen von Reinforcement Learning wirken häufig recht kompliziert. Die „Lernvorgänge“ sind von vielen Parametern und praktischen Umständen abhängig. Sie nehmen viel Rechenzeit in Anspruch und gleichzeitig ist der Erfolg ungewiss. Die Ideen der Algorithmen, die hinter den lernfähigen Agenten stecken, sind jedoch meist sehr anschaulich und gut zu begreifen, zudem werden wir Live-Visualisierungen einsetzen, mit denen sich der Lernfortschritt und der gegenwärtig erreichte Lernstand on-line beobachten lässt.

Das Thema „Reinforcement Learning“ soll in einer Form präsentiert werden, die Einsteigern zügig die wichtigsten Ansätze und zentrale Algorithmen vermittelt sowie eigene interessante Experimente ermöglicht. Dabei kommen Werkzeuge, wie sie bspw. in Einsteigerkursen oder im Programmierunterricht verwendet werden zur Anwendung. Sie werden im Buch auch Anregungen für Unterricht und Lehre finden. Es soll in erster Linie nicht um die Bedienung einer Hochleistungsblackbox gehen, sondern um das Verstehen, Begreifen, Beurteilen und vielleicht auch das innovative Weiterentwickeln der Algorithmen in eigenen Versuchen. Ein Auto zu fahren ist das eine, zu verstehen, wie der Benzinmotor eines Autos funktioniert, ist eine andere Sache. Obwohl beides auch eng verknüpft ist: zum einen erfordert das Fahren auch gewisse Kenntnisse in der Funktionsweise eines Autos und umgekehrt bestimmt der Konstrukteur eines Autos auch, wie das Fahrzeug gefahren wird. Im übertragenen Sinne werden wir jeweils nach einigen theoretischen Vorüberlegungen einige „Motoren“ und „Seifenkisten“ selbst bauen und ausprobieren, um zu begreifen, wie die Technik funktioniert.

„Schachbrettwelten“, sogenannte Gridworlds, wie Abb. 1 spielen in Einführungskursen in die Programmierung eine große Rolle. Hierbei handelt es sich jedoch nicht um Brettspiele, sondern um zweidimensionale Raster, in denen sich diverse Objekte, also „Schätze“, „Fallen“, „Mauern“ und ähnliches, sowie bewegliche Figuren befinden.

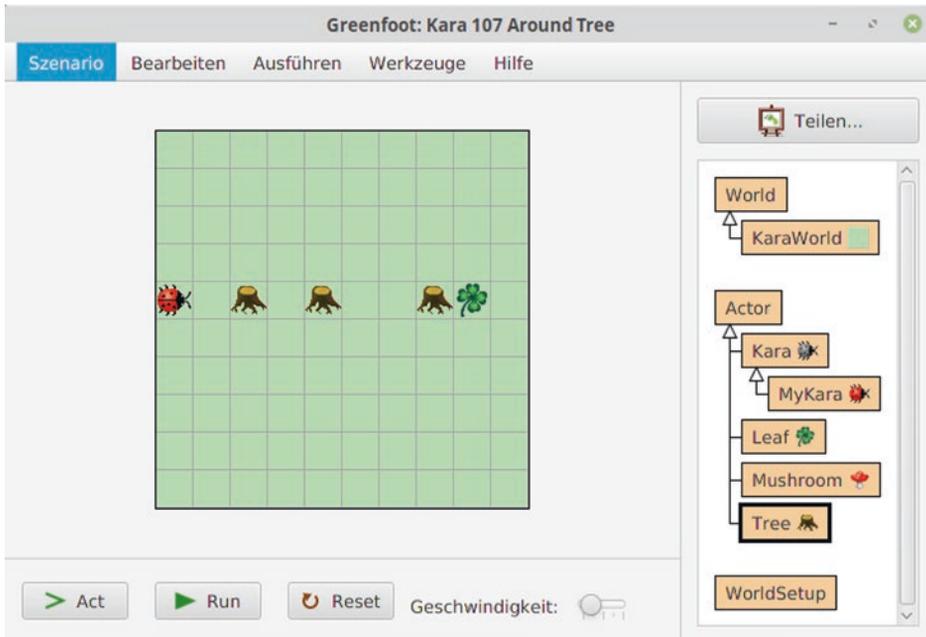


Abb. 1 Gridworld „Kara“

Weiterhin ist in der Lehre robotische Hardware, die in der Regel mit einem Differenzialantrieb und einigen einfachen Sensoren, wie z. B. Berührungs- oder Farbsensoren, ausgestattet ist, weit verbreitet, bspw. mit Bausätzen des Spielzeuherstellers LEGO, wie in Abb. 2 zu sehen, Fischertechnik oder OpenSource-Projekten wie „Makeblock“.

Mit solchen Hilfsmitteln werden algorithmische Grundstrukturen, wie Sequenzen, bedingte Anweisungen oder Wiederholungsschleifen gelehrt. Allerdings ist die Funktionsweise der elementaren algorithmischen Strukturen damit recht schnell erkannt. Die lustigen Figuren in den Schachbrettwelten, wie der Kara-Käfer oder der Java-Hamster oder auch die Roboter-Kreationen wecken die Motivation „wirklich“ intelligentes Verhalten zu programmieren, allerdings bleibt auf der Ebene der algorithmischen Grundstrukturen das Verhalten solcher „Roboter“ i. d. R. sehr mechanisch und kaum flexibel, – intelligentes Verhalten sieht anders aus.

Interessanterweise ist die akademische Standardliteratur zum Reinforcement Learning ebenfalls voll mit solchen „Gridworlds“ und einfachen Robotern. Denn diese bieten klare Vorteile: zum einen sind sie komplex genug für interessante Experimente und sehr anschaulich, zum anderen sind sie aber wegen ihrer Einfachheit gut durchschaubar und erlauben eine mathematische Durchdringung. In diesem einführenden Sach- und Experimentierbuch werden uns diese einfachen „Welten“ zunächst einen anschaulichen

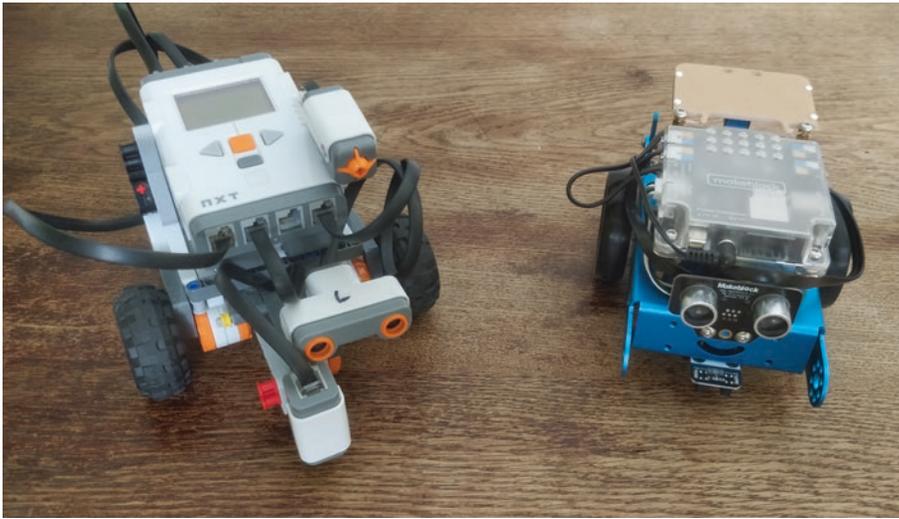


Abb. 2 Roboter mit Differenzialantrieb und diversen Sensoren (LEGO™ und Makeblock)

Zugang zu den Algorithmen der lernfähigen Agenten ermöglichen, in späteren Kapiteln werden wir uns dann auch mit „kontinuierlichen“ und dynamischen Szenarien beschäftigen. Das Buch richtet sich an Lernende oder Interessierte, die sich mit diesem Gebiet der künstlichen Intelligenz beschäftigen möchten (oder müssen), darüber hinaus ist es auch für Lehrpersonen oder Techniker gedacht, die sich weiterbilden, anschauliche Übungen mit ihren Schülern oder Studierenden oder eigene Experimente durchführen möchten.

Dieses Buch hat die Besonderheit, dass die Algorithmen nicht in der Programmiersprache Python, sondern in der allgemein bei Softwareentwicklern und auch in der Lehre weit verbreiteten Sprache Java vorgestellt werden.

Die Schlagwörter „Künstliche Intelligenz“ und „Maschinelles Lernen“ sind derzeit in aller Munde. „Künstliche Intelligenz“ ist hierbei der wesentlich umfassendere Begriff. Hierunter werden z. B. auch die regelbasierten Systeme der GOFAI („Good Old-Fashioned Artificial Intelligence“) gefasst, die nicht nur (altmodische) Schachprogramme, sondern zudem auch bestimmte Sprachassistenten, ChatBots und Ähnliches hervorgebracht haben. In solchen old-fashioned „Expertensystemen“ ist „Wissen“ symbolisch repräsentiert und wird mit Produktionsregeln, ähnlich von Wenn-Dann-Anweisungen, miteinander verbunden. Lernen wird hierbei vor allem als eine Aufnahme und Verarbeitung von symbolischem „Wissen“ verstanden. „Maschinelles Lernen“ dagegen wird aktuell eher mit dem überwachten Training von Mustererkennern insbesondere mit künstlichen neuronalen Netzen assoziiert. Diese Technologien besitzen eine große praktische Bedeutung und werden insbesondere

von US-Unternehmen schon seit Jahrzehnten mit gewaltigem wirtschaftlichem Erfolg angewendet.

„Reinforcement Learning“ kann dagegen eigentlich keinem der beiden genannten Felder zugeordnet werden. Es hat eine deutlich erweiterte und vor allem ganzheitlichere Perspektive auf lernende Systeme, bei der z. B. auch die Einbettung des Lernsystems in seine Umwelt viel stärker mit berücksichtigt wird. Beim Reinforcement Learning geht es im Kern darum, aktiv lernende, d. h. autonom agierende Agenten zu bauen, die sich in ihrer Umgebung zunehmend erfolgreich verhalten, wobei sie sich durch Versuch und Irrtum verbessern. Die Herausforderung besteht darin, während des selbsttätigen Erkundungsprozesses innere Strukturen aufzubauen, die das Agentenverhalten immer zweckmäßiger steuern. Hierfür müssen auch Erkenntnisse aus verschiedenen Gebieten der Künstlichen Intelligenz und des maschinellen Lernens miteinander kombiniert werden.

Dies ist auch kein Wunder, wenn man bedenkt, dass das Szenario des Reinforcement Learnings – erfolgreiches Agieren innerhalb eines Umweltsystems – viel mehr den biologischen Wurzeln der Kognition entspricht, wo ja alle kognitiven Fähigkeiten den natürlichen Anforderungen entsprechend kombiniert angewendet und permanent entwickelt werden. Derzeit wird in der KI-Forschung allgemein der biologische Ursprung des kognitiven Apparates stärker zur Kenntnis genommen als früher. Begriffe wie „Situiertheit“ („situated AI approach“) und in einem erweiterten Sinne auch „Embodiement“, also die Beschaffenheit des „Körpers“ und der sensorischen und motorischen Fähigkeiten, spielen eine größere Rolle. Zum einen versucht man z. B. in den Neurowissenschaften die Arbeitsweise biologischer kognitiver Systeme besser zu verstehen oder die Dynamik lebendiger Systeme künstlich nachzubilden und zu simulieren, bis hin zu Experimenten mit „künstlichem Leben“.

„Künstliche Intelligenz“-Forschung begann mindestens schon im mechanischen Zeitalter, also lange vor Turings Zeiten, z. B. mit den mechanischen Rechenmaschinen von Leibniz oder Pascal und auch auf dem Pfad der erwähnten „aktiven Auseinandersetzung“ mit geistigen Prozessen und der entsprechenden maschinellen Nachbildungsversuche wurden manche kuriose Konstruktionen hervorgebracht vgl. Abb. 3, so manche Irrwege beschränkt, aber auch zahlreiche wichtige Erkenntnisse gesammelt, nicht nur in technischer, sondern auch in erkenntnistheoretischer Hinsicht.

Im „Reinforcement Learning“ können sich klassische, modellbasierte KI-Ansätze, die z. B. diverse Suchalgorithmen hervorgebracht haben, mit „situierbaren“, „modellfreien“ oder „konnektionistischen“ Ansätzen, wo z. B. dann auch neuronale Deep-Learning Netze eine Rolle spielen, produktiv verbinden, wie wir z. B. beim „Alpha Go Zero“-Algorithmus sehen werden.

Das Thema künstliche Intelligenz weckt auch Ängste, nicht nur wegen der Gefahren, die die neue Technologie mit sich bringt. Manche haben auch grundsätzliche Kritik, wie z. B. Julian Nida-Rümelin, der mit dem Aufkommen eines „Maschinenmenschen“

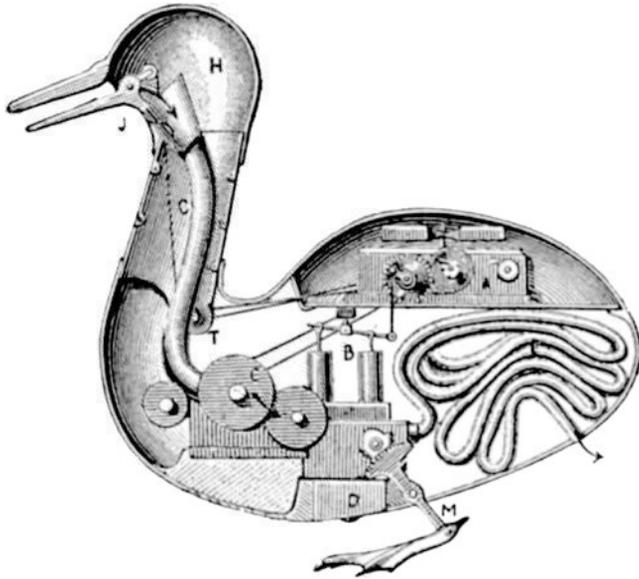


Abb. 3 Die mechanische Ente von Vaucanson (1738) konnte mit den Flügeln flattern, schnattern und Wasser trinken. Sie hatte sogar einen künstlichen Verdauungsapparat: Körner, die von ihr aufgepickt wurden, „verdaute“ sie in einer chemischen Reaktion in einem künstlichen Darm und schied sie daraufhin in naturgetreuer Konsistenz aus. (Quelle: Wikipedia)

das Ende des aufgeklärten Humanismus kommen sieht. Oder Weizenbaums klassische Kritik „Die Macht der Computer und die Ohnmacht der Vernunft“, wo er den simplen Intelligenzbegriff der KI-Forscher kritisiert und die Idee von „Künstlicher Intelligenz“ als „perverse, grandiose Phantasie“ bezeichnet. Auf der anderen Seite finden wir „KI-Propheten“ die manches mageres Ergebnis überhöhen, mystifizieren und ihr „Geheimwissen“ preisen, um sich und ihre Zunft zu beweihräuchern.

Das Buch möchte dazu beitragen, verschiedene Aspekte dieser Technologie besser zu verstehen, die Grenzen, aber auch die gewaltigen Potenziale, realistischer einzuschätzen und mystifizierende Aussagen, aber auch kritische Anmerkungen besser zu beurteilen. Auf den letzten Seiten des Buches soll auch nochmal auf eher philosophische Fragen und Perspektiven eingegangen werden. In der Regel berühren solche kritischen Aussagen die Grundlagen der KI-Forschung nicht so fundamental, wie es auf den ersten Blick erscheint, die produktive Berücksichtigung der Kritiken führt oft auch zu interessanten und wichtigen Weiterentwicklungen.

Inhaltsverzeichnis

1 Bestärkendes Lernen als Teilgebiet des Maschinellen Lernens	1
1.1 Maschinelles Lernen als automatische Verarbeitung von Feedback aus der Umwelt	2
1.2 Verfahren des maschinellen Lernens	3
1.3 Reinforcement Learning mit Java	8
Literatur	11
2 Grundbegriffe des Bestärkenden Lernens	13
2.1 Agenten	14
2.2 Die Steuerung des Agentensystems („Policy“)	16
2.3 Die Bewertung von Zuständen und Aktionen (Q-Funktion, Bellman-Gleichung)	17
Literatur	20
3 Optimal entscheiden in einer bekannten Umwelt	21
3.1 Wertiteration	23
3.1.1 Zielorientierte Zustandsbewertung („backward induction“)	23
3.1.2 Taktikbasierte Zustandsbewertung (Belohnungsvorhersage)	32
3.2 Iterative Taktiksuche	36
3.2.1 Unmittelbare Taktikoptimierung	36
3.2.2 Wechselseitige Verbesserung von Taktik und Bewertung	38
3.3 Optimale Taktik in einem Brettspiel-Szenario	43
3.4 Zusammenfassung	46
Literatur	49
4 Entscheiden und Lernen in einer unbekanntem Umwelt	51
4.1 Exploration vs. Exploitation	52
4.1.1 Methoden für probabilistische Aktionsauswahl	53
4.1.2 Künstliche Neugier	54
4.2 Rückwirkende Verarbeitung von Erfahrungen („Modellfreies Reinforcement Learning“)	55

4.2.1	Nachträgliche Auswertung von Episoden („Monte-Carlo“-Verfahren)	55
4.2.2	Unmittelbare Auswertung der temporalen Differenz (Q- und SARSA Algorithmus)	59
4.2.3	Berücksichtigung der Aktionshistorie (eligibility traces)	69
4.3	Taktikoptimierung	72
4.3.1	Eigenschaften biologischer Steuerungen	72
4.3.2	Monte-Carlo Taktiksuche	75
4.3.3	Policy Gradient und „Actor-Critic“-Architektur	79
4.4	Erkunden mit vorausschauenden Simulationen („Modellbasiertes Reinforcement Learning“)	91
4.4.1	Dyna-Q	92
4.4.2	Monte-Carlo Rollout	97
4.4.3	Monte-Carlo-Baumsuche (MCTS)	104
4.4.4	Bemerkungen zum Intelligenzbegriff	113
4.5	Systematik der Lernverfahren	115
	Literatur	116
5	Schätzer für Zustandsbewertung und Aktionsauswahl	117
5.1	Künstliche neuronale Netze	119
5.1.1	Mustererkennung mit dem Perzeptron	122
5.1.2	Die Anpassungsfähigkeit von künstlichen Neuronalen Netzen	125
5.1.3	Backpropagation-Lernen	140
5.1.4	Regression mit Multilayer Perzeptrons	143
5.2	Generalisierende Zustandsbewertung	145
5.3	Neuronale Schätzer für die Aktionsauswahl	157
	Literatur	159
6	Leitbilder in der Künstlichen Intelligenz	161
6.1	Grundvorstellungen im Wandel	162
6.2	Über das Verhältnis von Mensch und Künstlicher Intelligenz	166
	Literatur	170

Bestärkendes Lernen als Teilgebiet des Maschinellen Lernens

1

In der Evolution gilt das klug sein nichts, wenn es nicht zu klugem Handeln führt.

Tomasello 2014, S. 21 (Michael Tomasello)



Zusammenfassung

In diesem Kapitel geht es um einen verhaltensorientierten Begriff des maschinellen Lernens und die Einordnung des Reinforcement Learnings in das Gebiet des maschinellen Lernens allgemein. Es wird ein grober Überblick über die verschiedenen Prinzipien des Maschinellen Lernens gegeben und erklärt wodurch sie sich vom technischen Ansatz her unterscheiden. Im Anschluss wird auf die Implementierung von Reinforcement Learning Algorithmen mit der Programmiersprache Java eingegangen.

1.1 Maschinelles Lernen als automatische Verarbeitung von Feedback aus der Umwelt

Sehr viel Medieninteresse hat die komplexe Anpassungsfähigkeit von künstlichen neuronalen Netzen (kNNs) erregt. „Tiefe“ kNNs können z. B. erlernen, Bilder von Hunden und Katzen zu unterscheiden. Das klingt trivial, stellte allerdings mit der bislang verfügbaren Hard- und Software über Jahrzehnte eine schier unlösbare Aufgabe dar. Mit der technischen Lösung dieses Mustererkennungsproblems wurden viele völlig neuartige Anwendungsmöglichkeiten für Computersysteme geschaffen, beispielsweise in der medizinischen Diagnostik, der Industrieproduktion, der wissenschaftlichen Auswertung von Daten, im Marketing, im Finanzwesen, im Bereich der Militär- bzw. Securitytechnik u.v.m. Diese Neuerungen sind gewaltig und im Star-Trek-Jargon gesprochen so ist es doch höchst erstaunlich und faszinierend, dass wir in einem Zeitalter leben, in dem im großen Maßstab Dinge getan und Werke geschaffen werden, die nie zuvor in der doch nun schon einige zehntausend Jahre zählenden Geschichte der Menschheit getan worden sind. Bei der Mustererkennung handelt es sich allerdings nur um einen Teilbereich des maschinellen Lernens und zwar um das Gebiet des sogenannten „überwachten Lernens“, speziell um den Teil davon, der mit verteilten inneren Repräsentationen arbeitet. Auch wenn später auch nochmal auf das Trainieren von künstlichen neuronalen Netzen eingegangen wird, soll es in diesem Buch jedoch im Wesentlichen nicht um den Bereich der Mustererkennung gehen.

Der Turing-Award Preisträger von 2011 Judea Pearl sagte in der Novemberausgabe von „Spektrum der Wissenschaft“ 2018: „Jede beeindruckende Errungenschaft des ‚deep learning‘ läuft darauf hinaus, eine Kurve an Daten anzupassen. Aus mathematischer Sicht ist es egal, wie geschickt man das tut – es bleibt eine Kurvenanpassung, wenn auch komplex und keinesfalls trivial.“

Funktionsanpassungen an eine Datenmenge stellen in diesem Sinne nur einen Teilaspekt von Systemverhalten dar, was wir gemeinhin „intelligent“ nennen würden. Mithilfe einer an eine gegebene Inputdatenmenge gut angepassten Kurve können wir zwar die Funktionswerte, z. B. „Katze“ oder „Hund“, für vorher nie gesehene hochdimensional vorliegende „Argumente“ interpolieren, intelligentes Systemverhalten geht

darüber allerdings deutlich hinaus. Insbesondere wenn wir von lernfähiger, künstlicher Intelligenz sprechen wollen, möchten wir darunter auch Aktivitäten fassen, wie die sinnvolle Steuerung eines Staubsaugers, das Öffnen einer Tür durch einen Roboterarm oder kompetente Handlungsempfehlungen, z. B. an der Wertpapierbörse oder auch spannend agierende Gegner bei Brettspielen wie Schach und Go bzw. im Gamingbereich allgemein.

Hierbei muss KI-Software nicht nur vielfältige, teilweise voneinander abhängige Zustände bewerten, sondern muss auch weitblickend agieren. Das Verhalten eines Muster-Klassifikators beschränkt sich eigentlich nur auf die Einordnung in bestimmte Kategorien. Das Training eines solchen Systems erfolgt durch eine unmittelbare Rückmeldung durch einen wissenden Lehrer, „ja – richtig gemacht“ oder „nein – falsch. Bitte verbessern.“. Für Szenarien wie die oben genannten reicht das nicht aus, hier erhalten wir die meiste Zeit überhaupt keine Rückmeldung darüber, ob wir uns auf einem zielführenden Pfad befinden oder ob wir uns an einer Stelle besser anders verhalten hätten. Wir finden darüber hinaus am Ende einer Episode mitunter nicht einmal eine wissende Rückmeldung darüber, was die richtige Aktion gewesen wäre, sondern wir kassieren nur mehr oder weniger große Belohnungen oder „Strafen“, ja schlimmer noch, das Ende einer „Episode“ ist teilweise nicht einmal klar feststellbar, wie z. B. beim Erlernen des Laufens oder beim Verfassen eines Buchs über Reinforcement Learning. Wie lässt sich „intelligentes“ Systemverhalten in diesem allgemeineren Sinne automatisch erzeugen bzw. optimieren?

1.2 Verfahren des maschinellen Lernens

Zunächst müssen wir an dieser Stelle unseren Begriff von „Intelligenz“ dahin gehend anpassen, dass wir unter „Intelligenz“ im Wesentlichen „intelligentes Verhalten“ verstehen wollen und entsprechend „Lernen“ als eine Optimierung dieses Verhaltens. Allgemein kann man sagen, dass in den Verfahren des maschinellen Lernens eine kontinuierliche Verbesserung von artifiziellem Systemverhalten mithilfe von Rückmeldungen aus der Umwelt automatisch erzeugt werden soll. Die künstlichen Lernverfahren versuchen durch den Lernprozess die Ausgaben zu verbessern, die ein künstliches System bestimmten Systemeingaben zuordnet. Hierfür werden durch die Lernverfahren auf unterschiedliche Weise innere Repräsentationen gebildet, die das Systemverhalten mit Blick auf die zu erfüllenden Aufgaben zunehmend gut steuern sollen.

Für die Einschätzung der Möglichkeiten und Grenzen der diversen Lernverfahren ist es sinnvoll, diese zunächst entlang der Art des Feedbacks einzuteilen, welches sie aus der Umwelt erhalten. Hierbei lassen sich allgemein drei Arten des maschinellen Lernens unterscheiden. Sie unterscheiden sich im Wesentlichen darin, auf welche Weise die „Kritik“ präsentiert wird, durch die sich das Verhalten des künstlichen Systems verbessern soll. In der ersten Variante korrigiert ein „wissender“ Lehrer die Systemausgaben durch Präsentation der korrekten Ausgabe, in der zweiten erfolgt eine Bewertung