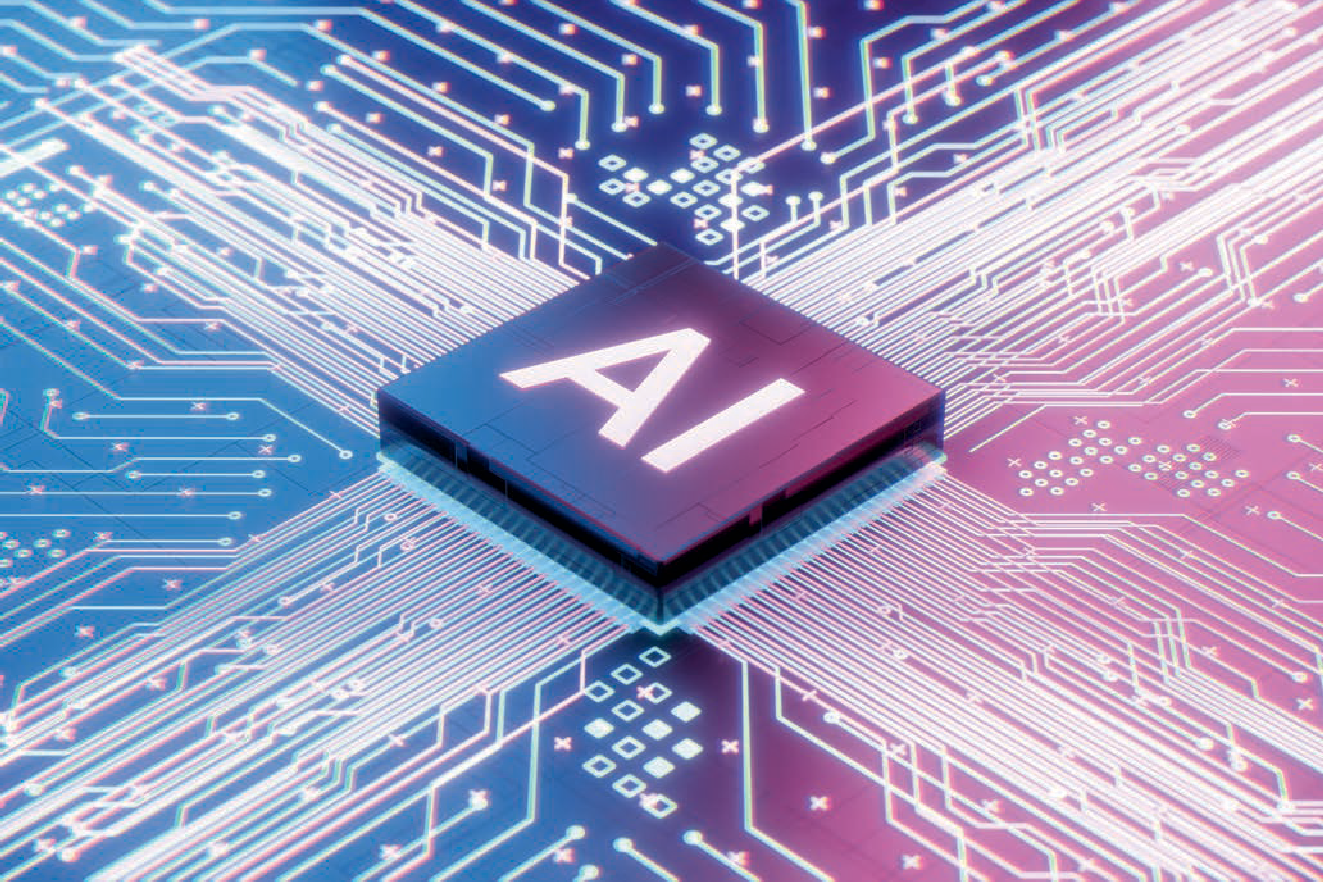
STUDIENSKRIPT



## Artificial Intelligence

**DLBDSEAIS01**



Übergeordnete Lernziele

##### Einführung **9**



In diesem Kurs erhalten Sie eine Einführung in das Gebiet der Künstlichen Intelligenz.

Die Disziplin der **Künstlichen Intelligenz** hat ihren Ursprung in verschiedenen Forschungsbereichen wie der Kognitionswissenschaft und den Neurowissenschaften. Das Studienskript beginnt mit einem Überblick über wichtige Ereignisse und Paradigmen, die das aktuelle Verständnis der Künstlichen Intelligenz geprägt haben. Darüber hinaus lernen Sie die typischen Aufgaben und Anwendungsbereiche der Künstlichen Intelligenz kennen.

Nach Abschluss dieses Studienskripts werden Sie die Konzepte des bestärkenden Lernens (engl. Reinforcement Learning) verstehen, die mit der menschlichen Art des Lernens in der realen Welt durch Informationserkundung und -ausnutzung vergleichbar sind.

Außerdem lernen Sie die Grundlagen des Natural Language Processing (Verarbeitung natürlicher Sprache, NLP) und der Computer Vision (Bildverarbeitung) kennen. Beides ist wichtig für künstliche Agenten, damit sie mit ihrer Umgebung interagieren können.



# Lektion 1

## Geschichte der KI

#### LERNZIELE

Nach der Bearbeitung dieser Lektion können Sie …

… beschreiben, wie sich die Künstliche Intelligenz als wissenschaftliche Disziplin entwickelt hat.

… die verschiedenen Paradigmen des Winters der Künstlichen Intelligenz nachvollziehen.

… die Bedeutung von Expertensystemen erklären und erläutern, wie sie zur Künstlichen Intelligenz beigetragen haben.

… über die Fortschritte der Künstlichen Intelligenz sprechen.

DL-E-DLBDSEAIS01-U01

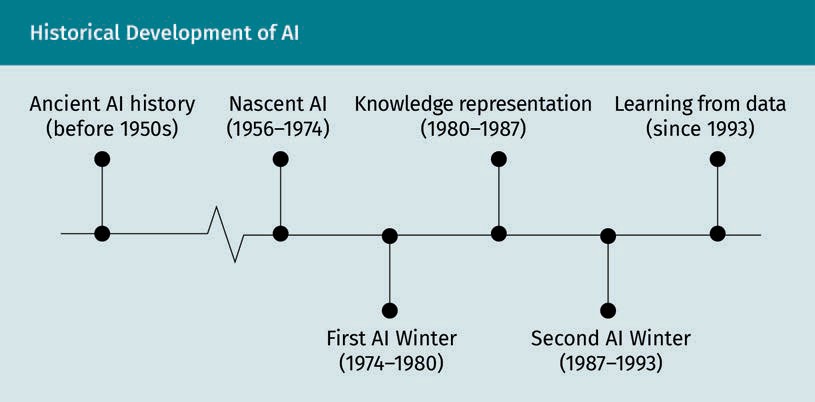
1. Geschichte der KI

### Einführung

In dieser Einheit wird die Geschichte der Künstlichen Intelligenz (KI) erläutert. Wir beginnen mit einer Übersicht über die historische Entwicklung der KI, die bis ins antike Griechenland zurückreicht.

Daran anschließend werden wir uns die jüngste Geschichte der KI ansehen und etwas über die sogenannten KI-Winter erfahren. Es gab in der Vergangenheit immer wieder Zyklen, in denen das Interesse an der KI stark schwankte, weil zu gewissen Zeitpunkten nicht alle Anforderungen an ein leistungsfähiges System erfüllt werden konnten.

Wir betrachten außerdem Expertensysteme und ihre Entwicklung. Im letzten Abschnitt widmen wir uns schließlich den bemerkenswerten Fortschritten im Bereich der Künstlichen Intelligenz, einschließlich moderner Konzepte und ihrer Anwendungsmöglichkeiten.



Die Abbildung oben veranschaulicht die Meilensteine der KI, die in den folgenden Abschnitten besprochen werden.

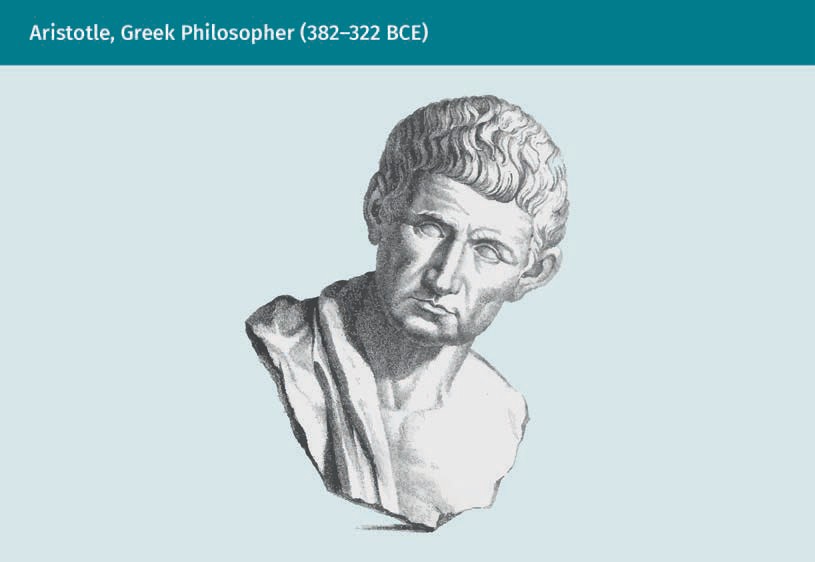
### Historische Entwicklung

Die historische Betrachtung der Künstlichen Intelligenz beginnt oft in den 1950er Jahren, als sie erstmals in der Informatik eingesetzt wurde. Allerdings gab es bereits im 4. Jahrhundert vor Christus Überlegungen zur KI. Daher werden wir zunächst einen kurzen Überblick über die Geschichte der Künstlichen Intelligenz in der Antike geben, bevor wir uns mit der jüngeren Geschichte befassen.

Geschichte der KI

###### Aristoteles, griechischer Philosoph (384–322 v. Chr.)

Aristoteles war der erste, der das menschliche Denken auf eine Weise formalisierte, dass es imitiert werden konnte. Um logische Schlussfolgerungen zu formalisieren, zählte er alle möglichen kategorischen Syllogismen vollständig auf (Giles, 2016).



Syllogismen (griechisch: *syllogismós*, „Schluss“, „Schlussfolgerung“) verwenden deduktives Denken, um aus zwei oder mehr gegebenen Sätzen sinnvolle Schlussfolgerungen abzuleiten. Logische Programmiersprachen, wie sie heute verwendet werden, basieren auf einem modernen Äquivalent von Aristoteles' Art und Weise, das Denken so zu formalisieren, wie logische Ableitungen verwendet werden. Moderne Algorithmen in der KI können so programmiert werden, dass sie ausgehend von einer gegebenen Auswahl zuvor definierter Regeln gültige logische Schlussfolgerungen ableiten.

###### Leonardo da Vinci, italienischer Universalgelehrter (1452–1519)

Leonardo da Vinci entwarf eine hypothetische Rechenmaschine auf Papier, die jedoch nie in die Praxis umgesetzt wurde. Die Maschine verfügte über 13 Register und konnte zeigen, dass eine Blackbox auf der Grundlage eines im Speicher oder in der Mechanik gespeicherten Programms Eingaben annehmen und Ausgaben erzeugen kann.

Diese frühen Überlegungen zu Rechenmaschinen sind sehr wichtig, weil Fortschritte in der Informatik eine notwendige Voraussetzung für jede Art von Entwicklung in der KI sind.

###### René Descartes, französischer Philosoph (1596–1650)

Der französische Philosoph Descartes glaubte, dass sich Rationalität und Vernunft mithilfe von Prinzipien aus der Mechanik und Mathematik definieren lassen. Die Fähigkeit, Ziele mithilfe von Gleichungen zu formulieren, ist eine wichtige Grundlage für KI, da ihre Ziele mathematisch definiert sind. Nach Descartes sind Rationalismus und Materialismus zwei Seiten der gleichen Medaille (Bracken, 1984). Dies knüpft an die in der KI verwendeten Methoden an, bei denen rationale Entscheidungen auf mathematische Weise abgeleitet werden.

###### Thomas Hobbes, britischer Philosoph (1588–1679)

Thomas Hobbes präzisierte Descartes' Theorien über Rationalität und Vernunft. In seiner Arbeit stellte er Ähnlichkeiten zwischen dem menschlichen Denken und den Berechnungen von Maschinen fest. Hobbes beschrieb, dass der Mensch bei der rationalen Entscheidungsfindung Verfahren anwendet, die der Analysis ähnlich sind, so dass sie in einer Weise formalisiert werden können, die der Mathematik entspricht (Flasiński, 2016).

###### David Hume, schottischer Philosoph (1711–1776)

Lernkurve Die Lernkurve ist eine grafische Darstellung des Verhältnisses zwischen einem Lernergebnis und der für die Lösung einer neuen Aufgabe

benötigten Zeit.

Hume leistete grundlegende Beiträge zu Fragen der logischen Induktion und zum Konzept des kausalen Denkens (Wright, 2009). So kombinierte er beispielsweise Lernprinzipien mit wiederholter Exposition, was – unter anderem – einen erheblichen Einfluss auf die **Lernkurve** hatte (Russell & Norvig, 2022).

Heutzutage basieren viele Machine-Learning-Algorithmen auf dem Prinzip der Ableitung von Mustern oder Beziehungen in Daten durch wiederholte Exposition.

###### Jüngste Geschichte der Künstlichen Intelligenz

Die jüngere Geschichte der KI begann um 1956, als die bahnbrechende Dartmouth-Konferenz stattfand. Der Begriff der Künstlichen Intelligenz wurde erstmals auf dieser Konferenz geprägt und eine Definition des Konzepts vorgeschlagen (Nilsson, 2009). Im Folgenden werden wir die wichtigsten Persönlichkeiten, Organisationen und Konzepte in der Entwicklungsgeschichte der KI diskutieren.

Wichtige Persönlichkeiten

Die jüngere Geschichte der KI beginnt üblicherweise mit der bahnbrechenden Dartmouth-Konferenz im Jahr 1956, in deren Rahmen der Begriff „Künstliche Intelligenz“ zum ersten Mal geprägt und eine Definition des Begriffs vorgeschlagen wurde.

In dem Jahrzehnt, in dem die KI ihren Anfang nahm, trugen wichtige Persönlichkeiten zur Entwicklung der Disziplin bei.

Geschichte der KI

Alan Turing war ein englischer Informatiker und Mathematiker, der rationale Denkprozesse formalisierte und mechanisierte. Im Jahr 1950 konzipierte er den bekannten Turing-Test. Bei diesem Test wird eine KI dahingehend geprüft, ob sie mit einem menschlichen Beobachter kommunizieren kann, ohne dass dieser unterscheiden kann, ob er mit einer Maschine oder einem anderen Menschen spricht. Wenn der Mensch eine KI nicht als solche identifizieren kann, wird sie als echte KI betrachtet (Turing, 1950).

Der amerikanische Wissenschaftler John McCarthy studierte Automaten. Er war es, der bei den Vorbereitungen für die Dartmouth-Konferenz erstmals den Begriff „Künstliche Intelligenz“ prägte (McCarthy *et al*., 1955). In Zusammenarbeit mit dem Massachusetts Institute of Technology (MIT) und dem Unternehmen International Business Machines (IBM) etablierte er KI als eigenständiges Forschungsgebiet. Zudem erfand er im Jahr 1958 die Programmiersprache Lisp (McCarthy, 1960). Mehr als 30 Jahre lang wurde LISP in einer Vielzahl von KI-Anwendungen eingesetzt, z. B. in der Betrugserkennung und der Robotik. In den 1960er Jahren gründete er das Stanford Artificial Intelligence Laboratory, das einen bedeutenden Einfluss auf die Forschung zur Implementierung menschlicher Fähigkeiten wie Denken, Hören und Sehen in Maschinen hatte (Feigenbaum, 2012).

Der amerikanische Forscher Marvin Minsky, der 1959 das MIT Artificial Intelligence Laboratory gründete, war ein weiterer wichtiger Teilnehmer der Dartmouth-Konferenz. Minsky kombinierte Erkenntnisse aus der KI und der Kognitionswissenschaft (Horgan, 1993).

Mit seinem Hintergrund in der Linguistik und der Philosophie ist Noam Chomsky ein weiterer Wissenschaftler, der zur Entwicklung der KI beigetragen hat. Seine Arbeiten über die formale Sprachtheorie und die Entwicklung der Chomsky-Hierarchie spielen immer noch eine wichtige Rolle in Bereichen wie dem Natural Language Processing (NLP). Außerdem ist er für seine kritischen Ansichten zu Themen wie den sozialen Medien bekannt.

Wichtige Institutionen

Die einflussreichsten Institutionen, die an der Entwicklung von KI beteiligt sind, sind das Dartmouth College und das MIT. Seit der Dartmouth-Konferenz gab es mehrere wichtige Konferenzen am Dartmouth College, auf denen die neuesten Entwicklungen im Bereich der KI diskutiert wurden. Viele der ersten einflussreichen KI-Forscher haben am MIT gelehrt, weshalb das MIT bis heute einer der Schlüsselinstitution für die KI-Forschung ist. Aber auch Unternehmen wie IBM und Intel sowie staatliche Forschungsinstitute wie die Defense Advanced Research Projects (DARPA) haben durch die Finanzierung von Forschungsarbeiten zum Thema viel zur Entwicklung der KI beigetragen (Crevier, 1993).

Schlüsseldisziplinen, die zur Entwicklung der KI beitragen

Viele Forschungsbereiche haben zur Entwicklung der Künstlichen Intelligenz beigetragen. Die wichtigsten Disziplinen sind Entscheidungstheorie, Spieltheorie, Neurowissenschaften und Natural Language Processing:

* In der Entscheidungstheorie werden mathematische Wahrscheinlichkeit und wirtschaftlicher Nutzen kombiniert. Dies liefert die formalen Kriterien für die Entscheidungsfindung in der KI hinsichtlich des wirtschaftlichen Nutzens und des Umgangs mit Unsicherheit.
* Die Spieltheorie ist eine wichtige Grundlage für rationale Agenten, um Strategien zur Lösung von Spielen zu erlernen. Sie basiert auf den Forschungen des amerikanisch-ungarischen Computerwissenschaftlers John von Neuman (1903–1957) und des amerikanisch-deutschen Mathematikers und Spieltheoretikers Oskar Morgenstern (1902–1977) (Leonard, 2010).
* Die Erkenntnisse der Neurowissenschaften über die Funktionsweise des Gehirns werden zunehmend in Modellen der Künstlichen Intelligenz verwendet, zumal die Bedeutung von künstlichen neuronalen Netzen (engl. Artificial Neural Networks, ANN) zunimmt. Heutzutage gibt es in der KI viele Modelle, die versuchen, die Art und Weise nachzuahmen, wie das Gehirn Informationen speichert und Probleme löst.
* Natural Language Processing (NLP) kombiniert Linguistik und Informatik. Das Ziel von NLP ist es, nicht nur geschriebenen Text, sondern auch gesprochene Sprache zu verarbeiten.

Höhere Programmiersprachen sind wichtig für die Programmierung von KI. Sie sind der menschlichen Sprache näher als niedrige Programmiersprachen wie Maschinencode oder Assemblersprache und ermöglichen es dem Programmierer, unabhängig von den Befehlssätzen der Hardware zu arbeiten. Einige der Sprachen, die speziell für KI entwickelt wurden, sind Lisp, Prolog und Python:

* Lisp wurde von John McCarthy entwickelt und ist eine der ältesten Programmiersprachen. Der Name kommt von „list processing“, da Lisp in der Lage ist, Zeichenketten auf eine einzigartige Weise zu verarbeiten (McCarthy, 1960). Auch wenn die Sprache auf die 1960er Jahre zurückgeht, wurde sie nicht nur für die frühe KI-Programmierung verwendet, sondern ist auch heute noch relevant.
* Eine weitere frühe KI-Programmiersprache ist Prolog, die speziell dafür entwickelt wurde, Theoreme zu beweisen und logische Formeln zu lösen.
* Heutzutage ist die vielseitig einsetzbare höhere Programmiersprache Python die wichtigste Programmiersprache. Da es sich bei Python um Open-Source handelt, gibt es umfangreiche Bibliotheken, die Programmierern helfen, Anwendungen auf sehr effiziente Weise zu erstellen.

Es gibt drei wichtige Faktoren, die zu den jüngsten Fortschritten im Bereich der Künstlichen Intelligenz beigetragen haben:

* Die zunehmende Verfügbarkeit riesiger Datenmengen, die für die Entwicklung und das Training von KI-Algorithmen benötigt werden.
* Enorme Verbesserungen bei der Datenverarbeitungskapazität von Computern.
* Neue Erkenntnisse aus Mathematik, Kognitionswissenschaft, Philosophie und Machine Learning.

Diese Faktoren unterstützen die Entwicklung von Herangehensweisen, die bisher nicht möglich waren, sei es aufgrund mangelnder Verarbeitungsmöglichkeiten oder fehlender Trainingsdaten.

Geschichte der KI

### KI-Winter

Der Begriff „KI-Winter“ tauchte erstmals in den 1980er Jahren auf. Der Begriff wurde von KI-Forschern geprägt, um Zeiten zu beschreiben, in denen das Interesse, die Forschungsaktivitäten und die Finanzierung von KI-Projekten deutlich zurückgingen (Crevier, 1993). Der Begriff klingt vielleicht ein wenig dramatisch. Er spiegelt jedoch sehr gut die Kultur der KI wider, die für ihre Begeisterung und ihren Überschwang bekannt ist.

Historisch gesehen hat der Begriff seinen Ursprung in dem Ausdruck „nuklearer Winter“, der eine Nachwirkung eines hypothetischen nuklearen Weltkrieges darstellt. Er beschreibt den Zustand, in dem die Atmosphäre von Asche verdunkelt wird und die Sonne die Erdatmosphäre nicht mehr erreichen kann, was bedeutet, dass die Temperaturen extrem sinken würden und nichts mehr wachsen könnte. Wird dieser Begriff auf die KI übertragen, so markiert er Zeiten, in denen das Interesse und die Finanzierung von KI-Technologien deutlich zurückgingen, was einen Rückgang der Forschungsaktivitäten zur Folge hatte. Derartige Rückschläge beruhen in der Regel auf übertriebenen Erwartungen an die Fähigkeiten neuer Technologien, die in der Folge realistischerweise nicht erfüllt werden können.

Es hat bisher zwei KI-Winter gegeben. Der erste dauerte ungefähr von 1974 bis 1980 und der zweite von 1987 bis 1993 (Crevier, 1993).

###### Der erste KI-Winter (1974–1980)

Während des Kalten Krieges zwischen der ehemaligen Sowjetunion und den Vereinigten Staaten war die automatische Sprachübersetzung eine der wichtigsten Triebfedern für die Finanzierung von KI-Forschungsaktivitäten (Hutchins, 1997). Da es nicht genügend Übersetzer gab, um die Nachfrage zu befriedigen, waren die Erwartungen groß, diese Aufgabe zu automatisieren. Die versprochenen Ergebnisse bei der maschinellen Übersetzung konnten jedoch nicht erreicht werden. Frühe Versuche, Sprache automatisch zu übersetzen, scheiterten spektakulär. Eine der großen Aufgaben zu dieser Zeit war der Umgang mit Mehrdeutigkeiten von Wörtern. So wurde zum Beispiel der englische Satz „out of sight, out of mind“ (dt. „aus den Augen, aus dem Sinn“) ins Russische mit „unsichtbarer Idiot“ übersetzt (Hutchins, 1995).

Als das Automatic Language Processing Advisory Committee die Ergebnisse der von den USA großzügig finanzierten Forschung auswertete, kam es zu dem Schluss, dass maschinelle Übersetzungen weder so genau, noch schneller oder billiger waren als der Einsatz von Menschen (Automatic Language Processing Advisory Committee, 1966). Darüber hinaus hatten Perzeptren (sing. Perzeptron) – die damals ein beliebtes Modell der neuronal inspirierten KI waren – schwerwiegende Mängel, da selbst einfache logische Funktionen, wie z. B. Exklusiv-Oder (XOR), in diesen frühen Systemen nicht wiedergegeben werden konnten.

###### Der zweite KI-Winter (1987–1993)

Der zweite KI-Winter begann um das Jahr 1987, als die KI-Community die Entwicklungen pessimistischer betrachtete. Ein Hauptgrund dafür war der Zusammenbruch des Geschäfts mit **Lisp-Maschinen**, der zu der Auffassung führte, dass die Branche untergehen könnte (Newquist, 1994). Außerdem stellte sich heraus, dass es ab einem bestimmten Punkt nicht mehr möglich war, frühe erfolgreiche Beispiele für Expertensysteme zu entwickeln. Diese Expertensysteme waren der

Lisp-Maschine Eine Lisp-Maschine ist ein Computertyp, der die Sprache Lisp unterstützt.

Haupttreiber für das erneute Interesse an KI-Systemen nach dem ersten KI-Winter. Der Grund für die Einschränkungen war, dass das Wachstum der Faktendatenbanken nicht mehr zu bewältigen war und die Ergebnisse bei unbekannten Eingaben, d. h. bei Eingaben, auf die die Maschinen nicht trainiert worden waren, unzuverlässig waren.

Es gibt jedoch auch Argumente dafür, dass es so etwas wie KI-Winter nicht gebe und dass es sich dabei lediglich um Mythen handele, die von einigen wenigen prominenten Forschenden und Organisationen verbreitet würden, die Geld verloren haben (Kurzweil, 2014). Während das Interesse an Lisp-Maschinen und Expertensystemen abnahm, war KI auch weiterhin tief in viele andere Arten von Verfahren wie Kreditkartentransaktionen eingebettet.

###### Ursachen für die KI-Winter

Es gibt verschiedene Faktoren, die zu einem KI-Winter führen können. Die drei wichtigsten Voraussetzungen für den Erfolg von Künstlicher Intelligenz sind

* Algorithmen und Erfahrung mit ihnen,
* Rechenkapazität und
* die Verfügbarkeit von Daten.

Die vergangenen KI-Winter erwuchsen daraus, dass nicht alle diese Anforderungen erfüllt wurden.

Während des ersten KI-Winters gab es bereits leistungsstarke Algorithmen. Um erfolgreich Ergebnisse zu erzielen, ist es jedoch notwendig, eine große Datenmenge zu verarbeiten. Dies erfordert eine große Speicherkapazität und eine hohe Verarbeitungsgeschwindigkeit. Zur damaligen Zeit standen jedoch nicht genügend Daten zur Verfügung, um diese Algorithmen richtig zu trainieren. Daher konnten die Erwartungen der Beteiligten und Investoren nicht erfüllt werden. Da die finanzierte Forschung nicht die versprochenen Ergebnisse liefern konnte, wurde die Finanzierung eingestellt.

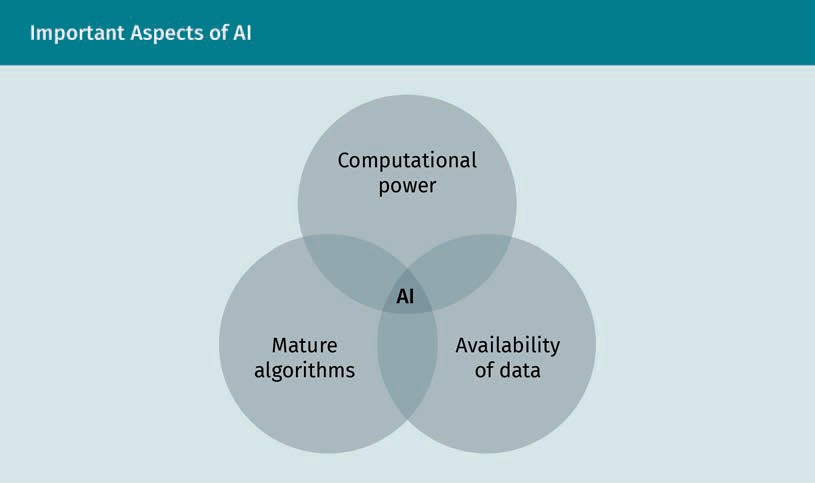
Bis in die 1980er Jahre war die Rechnerkapazität ausreichend gewachsen, um die verfügbaren Algorithmen auf kleinen Datensätzen zu trainieren. Als jedoch Ansätze des Machine Learning und des Deep Learning in den späten 1980er Jahren zum festen Bestandteil der KI wurden, stieg der Bedarf an großen Datensätzen zum Trainieren von KI-Systemen, was erneut zu einem Problem wurde. Der Mangel an gekennzeichneten Trainingsdaten führte – obwohl Rechenkapazitäten zur Verfügung gestanden hätten – dazu, dass einige der KI-Projekte als gescheitert angesehen wurden.

Wie die KI-Winter zeigen, ist es unmöglich, Fortschritte bei der Entwicklung von KI-Algorithmen zu machen, wenn nicht genügend Rechenkapazität (hinsichtlich Datenspeicherung und Verarbeitungsgeschwindigkeit) und Trainingsdaten vorhanden sind.

###### Der nächste KI-Winter

Heutzutage sind alle drei oben genannten Faktoren vollständig erfüllt. Es steht ausreichend Rechenleistung zur Verfügung, um die verfügbaren Algorithmen auf einer großen Anzahl von bestehenden Datensätzen zu trainieren. Die folgende Abbildung fasst die Voraussetzungen für den Erfolg von KI zusammen.

Geschichte der KI



Die Frage, ob es in Zukunft einen weiteren KI-Winter geben könnte, lässt sich jedoch kaum beantworten. Wenn ein hochgejubeltes Konzept viel Geld erhält, aber nicht funktioniert, kann es sein, dass es nicht mehr finanziert wird, was zu einem weiteren KI-Winter führen könnte. Nichtsdestotrotz sind KI-Technologien heute auch in viele andere Forschungsbereiche eingebettet. Wenn leistungsschwache Projekte nicht mehr finanziert werden, ist trotzdem noch Raum für neue Entwicklungen. Es steht also jedem frei zu entscheiden, ob KI-Winter nur ein Mythos sind oder ob das Konzept wirklich von Bedeutung ist.

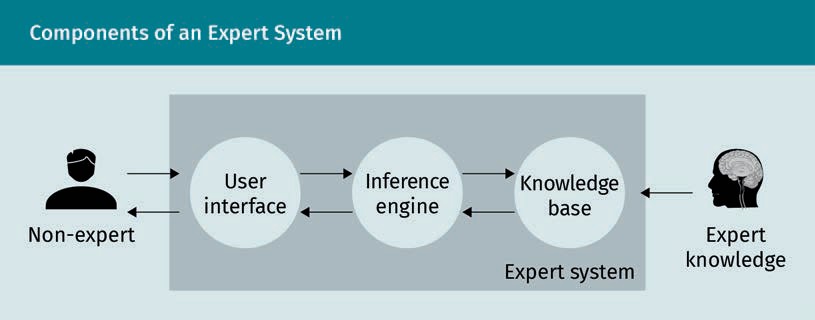
### Expertensysteme

Eines der Schlüsselkonzepte, wenn die Geschichte der Künstlichen Intelligenz betrachtet wird, sind Expertensysteme. Expertensysteme gehören zu der Gruppe der wissensbasierten Systeme. Wie der Name schon sagt, besteht das Ziel von Expertensystemen darin, den Entscheidungs- und Lösungsfindungsprozess mithilfe des domänenspezifischen Wissens eines Experten zu emulieren. Das Wort „Experte“ bezieht sich auf einen Menschen mit spezieller Erfahrung und Wissen in einem bestimmten Bereich, wie z. B. Medizin oder Mechanik. Da Probleme in einem bestimmten Definitionsbereich zwar ähnlich, aber nie ganz gleich sein können, kann die Lösung von Problemen in diesem Bereich nicht allein durch Auswendiglernen erreicht werden. Vielmehr wird das Problemlösen durch eine Methode ergänzt, bei der es darum geht, Erfahrungswissen an neue Probleme und Anwendungsszenarien anzupassen oder darauf anzuwenden.

###### Komponenten eines Expertensystems

Expertensysteme sollen nicht fachkundigen Benutzern und Benutzerinnen helfen, Entscheidungen auf der Grundlage des Wissens eines Experten zu treffen.

Die folgende Abbildung zeigt die typischen Komponenten eines Expertensystems:



Expertensysteme bestehen aus einem Korpus formalisierten Expertenwissens zu einem bestimmten Anwendungsbereich, das in der Wissensbasis gespeichert ist. Die Inferenzmaschine (engl. inference engine) nutzt die Wissensbasis, um aus den Regeln und Fakten des Wissens Schlussfolgerungen zu ziehen. Sie implementiert Regeln des logischen Denkens, um neue Fakten, Regeln und Schlussfolgerungen abzuleiten, die nicht explizit in dem gegebenen Korpus der Wissensbasis enthalten sind. Eine Benutzeroberfläche ermöglicht es dem nicht fachkundigen Benutzer, mit dem Expertensystem zu interagieren, um ein bestimmtes Problem aus dem Anwendungsbereich zu lösen.

###### Arten von Expertensystemen

Im Hinblick auf die Wissensdarstellung lassen sich drei Ansätze für Expertensysteme unterscheiden:

* Fallbasierte Systeme speichern Beispiele für konkrete Probleme zusammen mit einer erfolgreichen Lösung. Wird ein neuer, bisher unbekannter Fall vorgelegt, versucht das System, eine Lösung für einen ähnlichen Fall zu finden und diese Lösung auf den vorliegenden Fall anzuwenden. Die wichtigste Aufgabe ist die Definition eines geeigneten Ähnlichkeitsmaßes zum Vergleich von Problemstellungen.
* Regelbasierte Systeme stellen die Wissensbasis in Form von Fakten und Regeln vom Typ Wenn-A-dann-B dar, die Beziehungen zwischen den Fakten beschreiben.
* Wenn die zu lösende Problemklasse als Entscheidungsproblem kategorisiert werden kann, kann das Wissen in einem Entscheidungsbaum dargestellt werden. Letztere werden in der Regel durch die Analyse einer Reihe von Beispielen erstellt.

###### Entwicklung von Expertensystemen

Historisch gesehen sind Expertensysteme ein Auswuchs früherer Versuche, einen allgemeinen Problemlöser zu implementieren. Dieser Ansatz wird vor allem mit den Forschern Herbert A. Simon und Allen Newell in Verbindung gebracht, die in den späten 1950er Jahren eine Kombination aus Erkenntnissen der Kognitionswissenschaft und mathematischen Modellen des formalen Denkens verwendeten, um ein System zu entwickeln, das beliebige Probleme durch sukzessive Reduktion auf einfachere Probleme lösen sollte (Kuipers & Prasad, 2021). Obwohl dieser Versuch im Vergleich zu seinen hochgesteckten Zielen letztlich als Fehlschlag betrachtet wurde, hat er sich dennoch als äußerst einflussreich für die Entwicklung der Kognitionswissenschaft erwiesen.

Geschichte der KI

Eine der ersten Erkenntnisse aus dem Versuch, allgemeine Probleme zu lösen, war, dass die Konstruktion eines domänenspezifischen Problemlösers – zumindest im Prinzip – einfacher zu bewerkstelligen sein sollte. Dies führte dazu, dass über Systeme nachgedacht wurde, die domänenspezifisches Wissen mit domänenabhängigen, passenden Schlussfolgerungsmustern kombinierten. Edward Feigenbaum, der an der Stanford University, der damals führenden akademischen Einrichtung auf diesem Gebiet, arbeitete, definierte den Begriff Expertensystem und entwickelte die ersten praktischen Beispiele, als er das Heuristic Programming Project leitete (Kuipers & Prasad, 2021).

Die erste bemerkenswerte Anwendung war Dendral, ein System zur Identifizierung organischer Moleküle. Im nächsten Schritt wurden Expertensysteme entwickelt, die bei der medizinischen Diagnose von Infektionskrankheiten auf der Grundlage vorgegebener Daten und Regeln helfen sollten (Woods, 1973). Das daraus entstandene Expertensystem wurde MYCIN genannt und verfügte über eine Wissensbasis von etwa 600 Regeln. Es dauerte jedoch bis in die 1980er Jahre, bis Expertensysteme den Höhepunkt des Forschungsinteresses erreichten und zur Entwicklung von kommerziellen Anwendungen führten.

Die größte Errungenschaft der Expertensysteme war ihre Rolle als Pionier in der Idee einer formalen, aber leicht zugänglichen Darstellung von Wissen. Diese Darstellung war explizit in dem Sinne, dass sie als eine Reihe von Fakten und Regeln formuliert wurde, die sich für die Erstellung, Inspektion und Überprüfung durch einen Experten der Domäne eignen. Dieser Ansatz trennt also klar die domänenspezifische Fachlogik von der allgemeinen Logik, die für die Ausführung des Programms erforderlich ist – letztere ist in der Inferenzmaschine gekapselt. In krassem Gegensatz dazu stellen konventionelle Programmieransätze sowohl die interne Steuerung als auch die Fachlogik implizit in Form von Programmcode dar, der für Nicht-IT-Experten schwer zu lesen und zu verstehen ist. Zumindest im Prinzip ermöglichte der Ansatz der Expertensysteme auch Nicht-Programmierern die Entwicklung, Verbesserung und Wartung einer Softwarelösung. Überdies führte er die Idee des Rapid Prototyping (dt. schnelle Anwendungsentwicklung) ein, da die feste Inferenzmaschine die Erstellung von Programmen für völlig unterschiedliche Zwecke ermöglichte, indem einfach die zugrunde liegenden Regeln in der Wissensbasis geändert wurden.

Ein großer Nachteil des klassischen Expertensystem-Paradigmas, der schließlich auch zu einem starken Rückgang seiner Popularität führte, hing jedoch auch mit der Wissensbasis zusammen. Als Expertensysteme für immer neue Anwendungen entwickelt wurden, erforderten viele interessante Anwendungsfälle immer größere Wissensbasen, damit die jeweilige Domäne zufriedenstellend repräsentiert werden konnte. Diese Erkenntnis erwies sich in zweierlei Hinsicht als problematisch:

1. Erstens wächst die rechnerische Komplexität der Inferenz schneller als linear mit der Anzahl der Fakten und Regeln. Das bedeutet, dass die Antwortzeiten des Systems für viele praktische Probleme untragbar hoch waren.
2. Zweitens wird es mit zunehmender Größe einer Wissensbasis immer schwieriger, deren Konsistenz nachzuweisen, indem sichergestellt wird, dass sich die einzelnen Bestandteile nicht widersprechen.

Außerdem fehlt den regelbasierten Systemen im Allgemeinen die Fähigkeit, aus Erfahrungen zu lernen. Bestehende Regeln können vom Expertensystem selbst nicht geändert werden. Aktualisierungen der Wissensbasis können nur durch den Experten vorgenommen werden.

### Bemerkenswerte Fortschritte

Nachdem wir die Rückschläge der KI-Winter illustriert haben, ist es nun an der Zeit, den Blick auf die prosperierenden Zeiten zu richten, in denen die Künstliche Intelligenz enorme Fortschritte gemacht hat. Nach einem Überblick über die Forschungsthemen, die in den jeweiligen Epochen im Mittelpunkt standen, werden wir die wichtigsten Entwicklungen in angrenzenden Fachgebieten untersuchen und uns ansehen, wie sie mit dem Fortschritt in der Künstlichen Intelligenz zusammenhängen. Schließlich werden wir die Zukunftsaussichten der KI besprechen.

###### Das Aufkommen der Künstlichen Intelligenz (1956–1974)

In den Anfangsjahren wurde die KI-Forschung von der „symbolischen“ KI dominiert. Bei diesem Ansatz werden Regeln aus der formalen Logik verwendet, um Denkprozesse als Manipulation von symbolischen Repräsentationen von Informationen zu formalisieren. Dementsprechend beschäftigen sich die in dieser Ära entwickelten KI-Systeme mit der Implementierung des logischen Kalküls. In den meisten Fällen geschieht dies durch die Anwendung einer Suchstrategie, bei der die Lösungen Schritt für Schritt abgeleitet werden. Die Schritte in diesem Verfahren werden entweder logisch aus einem vorangegangenen Schritt abgeleitet oder systematisch durch Zurückverfolgen möglicher Alternativen abgeleitet, um Sackgassen zu vermeiden.

Die frühen Jahre waren auch die Zeit, in der die ersten Versuche der Computerlinguistik entwickelt wurden. Die ersten Ansätze der Sprachverarbeitung konzentrierten sich auf sehr begrenzte Umgebungen und Situationen. Daher war es möglich, erste Erfolge zu erzielen. Die Vereinfachung von Arbeitsumgebungen – ein „Mikrowelt“-Ansatz – führte ebenfalls zu guten Ergebnissen in den Bereichen Computer Vision und Robotersteuerung.

Parallel dazu wurden die ersten theoretischen Modelle von Neuronen entwickelt. Der Schwerpunkt der Forschung lag auf der Interaktion zwischen diesen Zellen (d. h. Recheneinheiten), um grundlegende logische Funktionen in Netzwerken zu implementieren.

###### Wissensdarstellung (1980–1987)

Der Schwerpunkt der ersten Welle der KI-Forschung lag vor allem auf dem logischen Schlussfolgern. Im Gegensatz dazu wurden die Hauptthemen der zweiten Welle von dem Versuch bestimmt, das Problem der Wissensdarstellung zu lösen. Der Grund für diese Schwerpunktverlagerung war die Einsicht, dass intelligentes Verhalten in alltäglichen Situationen nicht nur auf logischen Schlussfolgerungen beruht, sondern viel mehr auf allgemeinem Wissen über das Funktionieren der Welt. Diese wissensbasierte Sichtweise der Intelligenz war der Ursprung der frühen Expertensysteme. Das Hauptmerkmal dieser Technologien bestand darin, dass das domänenrelevante Wissen systematisch in Datenbanken gespeichert wurde. Mit Hilfe dieser Datenbanken wurde eine Reihe von Methoden entwickelt, um auf dieses Wissen auf effiziente und effektive Weise zuzugreifen.

Das aufkommende Interesse an der KI nach dem ersten KI-Winter ging auch mit einem Aufschwung der staatlichen Finanzierung zu Beginn der 1980er Jahre einher, mit Projekten wie dem Alvey-Projekt in Großbritannien und dem Fifth-Generation-Computer-Projekt der japanischen Regierung (Russell & Norvig, 2022).

Geschichte der KI

Außerdem konnten in dieser Zeit die frühen Rückschläge der neuronal inspirierten KI-Ansätze durch neue Netzwerkmodelle und die Verwendung von Backpropagation als Trainingsmethode in geschichteten Netzwerken von Recheneinheiten angegangen werden.

###### Aus Daten lernen (seit 1993)

In den 1990er Jahren gab es einige große Fortschritte der KI bei Spielanwendungen, als das erste Computersystem „Deep Blue“ in der Lage war, Garry Kasparov, den damaligen Schachweltmeister, zu schlagen.

Abgesehen von diesem bemerkenswerten, aber begrenzten Erfolg sind KI-Methoden bei der Entwicklung von Anwendungen für den praktischen Einsatz weit verbreitet. Erfolgreiche Ansätze aus den Teilbereichen der KI haben nach und nach ihren Weg in den Alltag gefunden – oft ohne explizit als KI bezeichnet zu werden. Zudem gibt es seit den frühen 1990er Jahren eine wachsende Zahl von Ideen aus der Entscheidungstheorie, der Mathematik, der Statistik und der Operations Research, die wesentlich dazu beigetragen haben, dass KI zu einer strengen und reifen wissenschaftlichen Disziplin geworden ist. Insbesondere das Paradigma der intelligenten Agenten wird immer beliebter. In diesem Zusammenhang verbindet sich das Konzept der intelligenten Agenten aus der Wirtschaftstheorie mit den Begriffen der Objekte und der Modularität der Informatik und bildet die Idee von Entitäten, die intelligent handeln können. Diese Perspektive ermöglicht es, die Sichtweise von KI als Imitation menschlicher Intelligenz auf das Studium intelligenter Agenten und ein breiteres Studium der Intelligenz im Allgemeinen zu verlagern.

Die Fortschritte in der KI seit den 1990er Jahren wurden durch eine erhebliche Steigerung der Datenspeicher- und Rechnerkapazitäten gefördert. Mit dem Aufschwung des Internets hat auch die Vielfalt, die Geschwindigkeit und das Volumen der erzeugten Daten unvergleichlich zugenommen, was den KI-Boom ebenfalls unterstützt hat.

Der jüngste Aufschwung des Interesses an der KI-Forschung begann 2012, als Deep Learning auf der Grundlage von Fortschritten bei konnektionistischen Machine-Learning-Modellen entwickelt wurde. Die Zunahme der Datenverarbeitungs- und Informationsspeicherkapazitäten in Kombination mit größeren Datenkorpora brachte theoretische Fortschritte bei den Machine-Learning-Modellen in die Praxis. Mit Deep Learning konnten neue Leistungsniveaus bei vielen Benchmark-Problemen des Machine Learnings erreicht werden. Dies führte zu einer Wiederbelebung des Interesses an bewährten Lernmodellen wie dem Reinforcement Learning und schuf Raum für neue Ideen wie das Adversarial Learning (Lernen durch absichtliche Verleitung zu Fehlklassifikationen).

###### Benachbarte Studienrichtungen

Es gibt viele Studienbereiche, die kontinuierlich zur KI-Forschung beitragen. Die einflussreichsten Gebiete werden im Folgenden beschrieben.

Linguistik

Linguistik kann im Großen und Ganzen als die Wissenschaft der natürlichen Sprache beschrieben werden. Sie befasst sich mit der Erforschung der strukturellen (grammatikalischen) und phonetischen Eigenschaften der zwischenmenschlichen Kommunikation. Um Sprache zu verstehen, muss der Kontext und der

Themenkomplex bekannt sein, in denen sie verwendet wird. In seinem Buch *Syntactic Structures* leistete Noam Chomsky (1957) einen wichtigen Beitrag zur Linguistik und damit auch zum Natural Language Processing. Da unsere Gedanken so eng mit der Sprache als einer Form der Darstellung verbunden sind, könnte man noch einen Schritt weiter gehen und Kreativität und Denken mit sprachlicher KI in Verbindung bringen. Wie ist es zum Beispiel möglich, dass ein Kind etwas sagt, was es noch nie zuvor gesagt hat? In der KI verstehen wir die natürliche Sprache als ein Kommunikationsmedium in einem bestimmten Kontext. Sprache ist also viel mehr als nur eine Darstellung von Wörtern.

Kognition

Im Zusammenhang mit KI bezieht sich der Begriff Kognition auf verschiedene Fähigkeiten wie Wahrnehmung und Kognition, logisches Denken, Intelligenz, Lernen und Verstehen sowie Denken und Verständnis. Dies spiegelt sich auch in dem Wort „Erkenntnis“ wider. Ein großer Teil unseres heutigen Verständnisses von Kognition ist eine Kombination aus Psychologie und Informatik. In der Psychologie werden Theorien und Hypothesen aus Beobachtungen an Menschen und Tieren gebildet. In der Informatik wird das Verhalten auf der Grundlage von Beobachtungen aus der Psychologie modelliert. Bei der Modellierung des Gehirns durch einen Computer gilt das gleiche Prinzip von Reiz und Reaktion wie im menschlichen Gehirn. Wenn der Computer einen Reiz empfängt, wird eine interne Darstellung dieses Reizes erstellt. Die Reaktion auf diesen Reiz kann dazu führen, dass das ursprüngliche Modell verändert wird. Sobald wir ein gut funktionierendes Computermodell für eine bestimmte Situation haben, wird der nächste Schritt darin bestehen, herauszufinden, wie Entscheidungen getroffen werden. Da Entscheidungen, die auf KI basieren, in immer mehr Bereichen unseres Lebens eine Rolle spielen, ist es wichtig, dass der Denkprozess für einen externen Beobachter transparent ist. Daher wird die Erklärbarkeit (die Fähigkeit zu erklären, wie eine Entscheidung getroffen wurde) immer wichtiger. Den auf Deep Learning basierenden Ansätzen mangelt es jedoch noch an Erklärbarkeit.

Spiele

Bei KI im Zusammenhang mit Spielen geht es um viel mehr als nur um Glücksspiele oder Computerspiele. Vielmehr geht es bei Spielen um Lernen, Wahrscheinlichkeit und Unsicherheit. Im frühen zwanzigsten Jahrhundert wurde die Spieltheorie von Oskar Morgenstern und John von Neuman als mathematisches Forschungsgebiet etabliert (Leonard, 2010). In der Spieltheorie wurde eine umfassende Taxonomie der Spiele entwickelt und in Verbindung damit einige Spielstrategien, die sich als optimale Strategien erwiesen haben.

Eine weitere mit der Spieltheorie verwandte Disziplin ist die Entscheidungstheorie. Während es in der Spieltheorie eher darum geht, wie sich die Züge eines Spielers auf die Optionen eines anderen Spielers auswirken, befasst sich die Entscheidungstheorie mit Nützlichkeit und Ungewissheit, d. h. mit Nutzwert und Wahrscheinlichkeiten. Bei beiden geht es nicht unbedingt ums Gewinnen, sondern eher darum, zu lernen, mit möglichen Optionen zu experimentieren und aufgrund von Beobachtungen herauszufinden, was funktioniert.

Spiele wie Schach, Dame und Poker werden in der Regel gespielt, um zu gewinnen oder sich gut zu unterhalten. Heute können Maschinen besser spielen als menschliche Spieler. Bis 2016 glaubte man, dass das Spiel Go aufgrund seiner kombinatorischen Komplexität eine unlösbare Aufgabe für Computer sein würde. Ziel des Spiels ist es, das größte Gebiet auf einem Spielbrett mit 19 horizontalen und vertikalen Linien abzugrenzen. Auch wenn das Regelwerk recht einfach ist, ergibt sich die Komplexität aus der Größe des Spielbretts und der daraus resultierenden Anzahl möglicher Züge. Diese Komplexität macht es unmöglich, Methoden anzuwenden, die für Spiele wie Schach und Dame verwendet wurden. Doch im Jahr

Geschichte der KI

2015 entwickelte DeepMind das System AlphaGo, das auf tiefen Netzwerken und bestärkendem Lernen basiert. Dieses System war das erste, das Lee Sedol, einen der besten Go-Spieler der Welt, schlagen konnte (Silver *et al*., 2016).

Nicht lange nach AlphaGo entwickelte DeepMind das System AlphaZero (Silver *et al*., 2018). Im Gegensatz zu AlphaGo, das aus dem Go-Wissen vergangener Spielaufzeichnungen lernte, lernt AlphaZero nur durch intensives Spielen gegen sich selbst nach dem Regelwerk. Dieses System erwies sich als noch stärker als AlphaGo. Bemerkenswert ist auch, dass AlphaZero sogar einige effektive und effiziente Strategien gefunden hat, die von Go-Experten bisher übersehen worden waren.

Das Internet der Dinge

Es ist erst ein paar Jahre her, dass der Begriff „Internet der Dinge“ (IoT) zum ersten Mal aufkam. Das IoT verbindet physische und virtuelle Geräte mithilfe von Technologien aus der Informations- und Kommunikationstechnologie. In unserem Alltag sind wir von einer Vielzahl physischer Geräte umgeben, die ständig miteinander vernetzt sind, z. B. Smartphones, Smart Home-Geräte, Autos und Wearables. Die Kommunikation zwischen diesen Geräten erzeugt eine riesige Menge an Daten, die das IoT mit der KI in Verbindung bringt. Während es beim IoT selbst nur darum geht, Geräte zu vernetzen und Daten zu sammeln, kann KI dabei helfen, die Interaktion zwischen diesen Maschinen mit intelligentem Verhalten zu ergänzen.

Die Integration intelligenter Geräte in unser tägliches Leben schafft nicht nur Chancen, sondern auch viele neue Herausforderungen. Zum Beispiel könnten Daten über die Medikation, die auf Körpermessungen eines Wearable-Geräts basieren, positiv genutzt werden, um eine Person an die Medikamenteneinnahme zu erinnern, aber auch, um über eine mögliche Erhöhung des Krankenversicherungsbeitrags zu entscheiden. Daher werden Themen wie die Ethik der Datennutzung und die Verletzung der Privatsphäre angesichts der neuen Anwendungsbereiche der KI immer wichtiger.

Quantencomputing

Das Quantencomputing basiert auf der physikalischen Theorie der Quantenmechanik. Die Quantenmechanik befasst sich mit dem Verhalten subatomarer Teilchen, die anderen Regeln folgen als in den Theorien der klassischen Physik beschrieben. In der Quantenmechanik ist es zum Beispiel möglich, dass sich ein Elektron gleichzeitig in zwei verschiedenen Zuständen befindet. Die Quantenmechanik geht davon aus, dass physikalische Systeme durch eine Wellenfunktion charakterisiert werden können, die die Wahrscheinlichkeiten beschreibt, dass sich das System in einem bestimmten Zustand befindet. Ziel ist es, diese Quanteneigenschaften zu nutzen, um Supercomputer zu bauen, in denen neue algorithmische Konzepte implementiert werden können, die es ihnen ermöglichen, klassische Rechner zu übertreffen (Giles, 2018). Die Art der Informationsverarbeitung aus dem Quantencomputing eignet sich gut für den probabilistischen Ansatz, der vielen KI-Technologien zugrunde liegt. Daher bieten Quantencomputer die Möglichkeit, KI-Anwendungen zu beschleunigen und so einen echten Vorteil bei der Verarbeitungsgeschwindigkeit zu erzielen. Da sich diese Systeme jedoch noch in einem frühen Entwicklungsstadium befinden, ist der Einsatz von Quantencomputern noch kaum erforscht.

###### Die Zukunft der KI

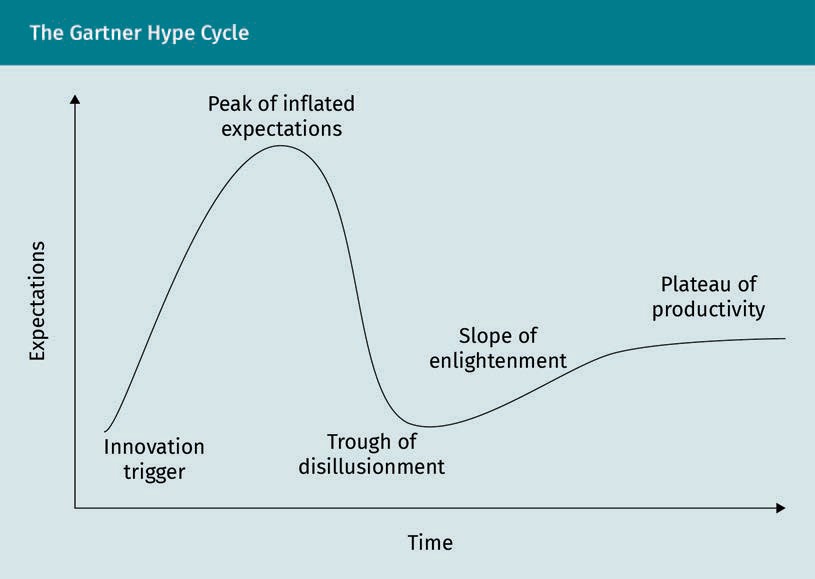
Der Versuch, die Auswirkungen eines Forschungsgebiets oder einer neuen Technologie auf die Zukunft abzuschätzen, ist immer höchst spekulativ, da die Zukunftsaussichten immer von den bisherigen Erfahrungen beeinflusst werden. Daher versuchen wir nicht, die langfristige Zukunft der KI vorherzusagen. Trotzdem wollen wir die Richtung der Entwicklungen in der KI und der unterstützenden Technologien untersuchen.

Der Hype-Zyklus von Gartner wird häufig verwendet, um das Potenzial neuer Technologien zu bewerten (Gartner, 2021). Der Hype-Zyklus wird in einem Diagramm dargestellt, bei dem die y-Achse die Erwartungen an eine neue Technologie darstellt, während die Zeit auf der x-Achse aufgetragen ist.

Die Zeitachse ist durch fünf Phasen gekennzeichnet:

1. In der Entdeckungsphase weckt ein technologischer Auslöser oder Durchbruch großes Interesse und löst die Innovation aus.
2. Der Gipfel der überzogenen Erwartungen wird in der Regel von viel Enthusiasmus begleitet. Auch wenn es möglicherweise erfolgreiche Anwendungen gibt, haben die meisten von ihnen mit frühen Problemen zu kämpfen.
3. Im Tal der Enttäuschungen zeigt sich, dass nicht alle Erwartungen erfüllt werden können.
4. Auf dem Pfad der Erleuchtung wird der Wert der Innovation erkannt. Es gibt ein Verständnis für den praktischen Nutzen und die Vorteile, aber auch für die Grenzen der neuen Technologie.
5. In der letzten Phase wird ein Plateau der Produktivität erreicht, und die neue Technologie wird zur Norm. Die endgültige Höhe dieses Plateaus hängt davon ab, ob die Technologie in einer Nische oder auf einem Massenmarkt eingeführt wird.

Geschichte der KI



Der Hype-Zyklus hat einige Ähnlichkeiten mit der umgekehrten U-Form einer Normalverteilung, mit der Ausnahme, dass das rechte Ende der Kurve in eine zunehmende Steigung mündet, die schließlich abflacht.

Im Jahr 2021 zeigt der Hype-Zyklus für Künstliche Intelligenz die folgenden Trends (Gartner, 2021):

* In der Phase des technologischen Auslösers tauchen Themen wie Composite AI (eine Kombination aus verschiedenen Ansätzen der KI) und allgemeine KI (die Fähigkeit einer Maschine, menschenähnliche intellektuelle Aufgaben auszuführen) auf. Außerdem zeigen Themen wie Human-Centered AI (KI, die sich durch Eingaben von Menschen verbessert und eine benutzerfreundliche Kommunikation zwischen KI und Mensch erlaubt) und Responsible AI (Entwicklung von sicheren, vertrauenswürdigen und ethischen KI-Lösungen), dass die Integration des Menschen für die Zukunft der KI immer wichtiger wird.
* Tiefe neuronale Netze, die in den letzten Jahrzehnten in vielen Anwendungen des maschinellen Lernens für neue Leistungsniveaus gesorgt haben, befinden sich immer noch auf dem Gipfel der überzogenen Erwartungen oder des Hypes. Außerdem tauchen in dieser Phase Themen wie Wissensgraphen und smarte Roboter auf.
* Im Tal der Enttäuschungen finden wir Themen wie autonome Fahrzeuge, bei denen die Finanzierung eingestellt wurde, weil die hohen Erwartungen in diesem Bereich nicht erfüllt werden konnten.

Bisher hat noch keines der KI-Themen das Plateau der Produktivität erreicht. Dies spiegelt die allgemeine Akzeptanz dieses Bereichs und den produktiven Einsatz der entsprechenden Technologien wider.

**Zusammenfassung**

Die Forschung im Bereich der Künstlichen Intelligenz ist schon seit langem im Gange. Die ersten theoretischen Überlegungen zur Künstlichen Intelligenz gehen auf griechische Philosophen wie Aristoteles zurück. Diese frühen Überlegungen wurden von Philosophen wie Hobbes und Descartes weitergeführt. Seit den 1950er Jahren ist sie auch zu einem wichtigen Bestandteil der Informatik geworden und hat wichtige Beiträge in Bereichen wie der Wissensdarstellung in Expertensystemen, dem maschinellen Lernen und der Modellierung neuronaler Netze geleistet.

In den vergangenen Jahrzehnten gab es in der KI-Forschung mehrere Höhen und Tiefen. Sie wurden durch einen Zyklus zwischen Innovationen, die mit hohen Erwartungen einhergingen, und der Enttäuschung, wenn diese Erwartungen nicht erfüllt werden konnten, oft aufgrund technischer Beschränkungen, verursacht.

Im Laufe der Zeit wurde die KI durch verschiedene Paradigmen aus unterschiedlichen Disziplinen geprägt. Das beliebteste Paradigma ist heutzutage das Deep Learning. Neue Anwendungsbereiche wie IoT oder Quantencomputing bieten eine Vielzahl von Möglichkeiten, wie KI eingesetzt werden kann. Es bleibt jedoch abzuwarten, wie intelligentes Verhalten in der Zukunft in Maschinen implementiert werden wird.



# Lektion 2

## Moderne KI-Systeme

#### LERNZIELE

Nach der Bearbeitung dieser Lektion können Sie …

… den Unterschied zwischen engen und allgemeinen KI-Systemen erklären.

… die wichtigsten Anwendungsbereiche für Künstliche Intelligenz nennen.

… die Bedeutung der Künstlichen Intelligenz für Unternehmensaktivitäten nachvollziehen.

DL-E-DLBDSEAIS01-U02

1. Moderne KI-Systeme

### Einführung

Künstliche Intelligenz ist zu einem festen Bestandteil unseres Alltags geworden. Es gibt verschiedene Beispiele, bei denen wir die Anwesenheit von KI gar nicht bemerken, sei es bei Google Maps oder den intelligenten Antworten in Google Mail.

Es gibt zwei Kategorien von KI, die in der folgenden Lektion erläutert werden: enge und allgemeine KI.

Organisationen wie Gartner, McKinsey oder PricewaterhouseCoopers (PwC) sagen der KI eine atemberaubende Zukunft voraus. Berichte wie der PWC-Bericht (2018) schätzen, dass KI einen Beitrag von 15,7 Billionen $ zur Weltwirtschaft leisten könnte. Nachdem wir die beiden Kategorien der KI erörtert haben, werden wir uns daher auf die wichtigsten Anwendungsbereiche der KI konzentrieren. Darüber hinaus werden wir uns ansehen, wie moderne KI-Systeme bewertet werden können.

### Enge und allgemeine KI

Neuere Forschungsthemen im Bereich der Künstlichen Intelligenz unterscheiden zwischen zwei Arten: der engen Künstlichen Intelligenz (Artificial Narrow Intelligence, ANI), auch als schwache Künstliche Intelligenz bezeichnet, und der allgemeinen Künstlichen Intelligenz (Artificial General Intelligence, AGI) oder starken Künstlichen Intelligenz. Bei ANI werden Systeme gebaut, um spezielle Funktionen in kontrollierten Umgebungen auszuführen, während AGI offene, flexible und von der Domäne unabhängige Formen der Intelligenz umfasst, wie sie von Menschen zum Ausdruck gebracht wird.

Auch wenn viele Menschen glauben, dass wir bereits über eine Art starker Künstlicher Intelligenz verfügen, sind die derzeitigen Ansätze immer noch domänenspezifisch implementiert und lassen die notwendige Flexibilität vermissen, um als AGI zu gelten. Es besteht jedoch ein breiter Konsens darüber, dass es nur eine Frage der Zeit ist, bis die Künstliche Intelligenz die menschliche Intelligenz übertreffen wird. Die Ergebnisse einer Umfrage unter 352 KI-Forschern zeigen, dass eine 50-prozentige Chance besteht, dass Algorithmen diesen Zustand bis 2060 erreichen könnten (Grace *et al*., 2017).

Im Folgenden werden wir uns die zugrundeliegenden Konzepte der schwachen und starken Künstlichen Intelligenz näher ansehen.

###### Enge Künstliche Intelligenz

Der Begriff ANI oder schwache KI steht für die aktuelle und zukünftige Künstliche Intelligenz. Auf ANI basierende Systeme können bereits heute komplexe Probleme oder Aufgaben schneller lösen als Menschen. Allerdings sind die Möglichkeiten dieser Systeme auf die Anwendungsfälle beschränkt, für die sie entwickelt wurden. Im Gegensatz zum menschlichen Gehirn können enge Systeme nicht von einer bestimmten Aufgabe auf eine Aufgabe aus einer anderen Domäne generalisieren.

Moderne KI-Systeme

Ein Gerät oder System, das zum Beispiel Schach spielen kann, wird wahrscheinlich nicht in der Lage sein, ein anderes Strategiespiel wie Go oder Shogi zu spielen, ohne dass es explizit für das Erlernen dieses Spiels programmiert wurde. Sprachassistenten wie Siri oder Alexa können als eine Art hybride Intelligenzen angesehen werden, die mehrere schwache KIs kombinieren. Diese Tools sind in der Lage, natürliche Sprache zu übersetzen und diese Wörter mit ihren Datenbanken abzugleichen, um verschiedene Aufgaben zu erfüllen. Sie sind jedoch nur in der Lage, eine begrenzte Anzahl von Problemen zu lösen, für die ihre Algorithmen geeignet sind und für die sie trainiert wurden. Zum Beispiel können sie derzeit keine Bilder analysieren oder den Verkehr optimieren.

Kurz gesagt: ANI umfasst die Anwendung von Intelligenz in Bezug auf komplexe Problemlösungen und die Anwendung von Intelligenz in Bezug auf eine einzelne Aufgabe.

###### Allgemeine Künstliche Intelligenz

Der Bezugspunkt, an dem AGI gemessen und beurteilt wird, sind die vielseitigen kognitiven Fähigkeiten des Menschen. Das Ziel von AGI ist nicht nur die Nachahmung der Interpretation von Sinneseindrücken, sondern auch die Nachahmung des gesamten Spektrums menschlicher kognitiver Fähigkeiten. Dazu gehören alle Fähigkeiten, die derzeit von ANI repräsentiert werden, einschließlich der Fähigkeit zur domänenunabhängigen Generalisierung. Das bedeutet, dass das Wissen über eine Aufgabe auf eine andere in einer anderen Domäne angewendet werden kann. Dazu könnten auch Motivation und Wille gehören. Einige philosophische Quellen gehen noch einen Schritt weiter und verlangen, dass AGI eine Art von Bewusstsein oder Selbsterkenntnis haben muss (Searle, 1980). Die Entwicklung einer AGI würde die folgenden Systemfähigkeiten erfordern:

* kognitive Fähigkeit, in mehreren Domänen zu funktionieren und zu lernen
* Intelligenz auf menschlichem Niveau in allen Domänen
* unabhängige Fähigkeit, Probleme zu lösen
* Problemlösungsfähigkeiten auf einem durchschnittlichen menschlichen Niveau über mehrere Domänen hinweg
* Fähigkeit, selbstständig Probleme zu lösen
* abstraktes Denkvermögen, ohne direkt auf frühere Erfahrungen zurückzugreifen
* Prüfung von Hypothesen ohne vorherige Erfahrung
* Wahrnehmung der gesamten Umgebung, in der das System agiert
* Selbstmotivation und Selbsterkenntnis

In Anbetracht des derzeitigen Stands der AGI ist es schwer vorstellbar, ein System zu entwickeln, das diese Anforderungen erfüllt. Darüber hinaus beinhalten beide Arten von KI auch das Konzept der Superintelligenz. Dieses Konzept geht sogar noch weiter als die derzeitigen Vorstellungen und beschreibt die Idee, dass ein intelligentes System eine Stufe der Erkenntnis erreichen kann, die über die menschlichen Fähigkeiten hinausgeht. Diese Selbstverbesserung könnte durch einen rekursiven Zyklus erreicht werden. Diese Stufe der KI liegt jedoch über der AGI und ist weiterhin sehr abstrakt.

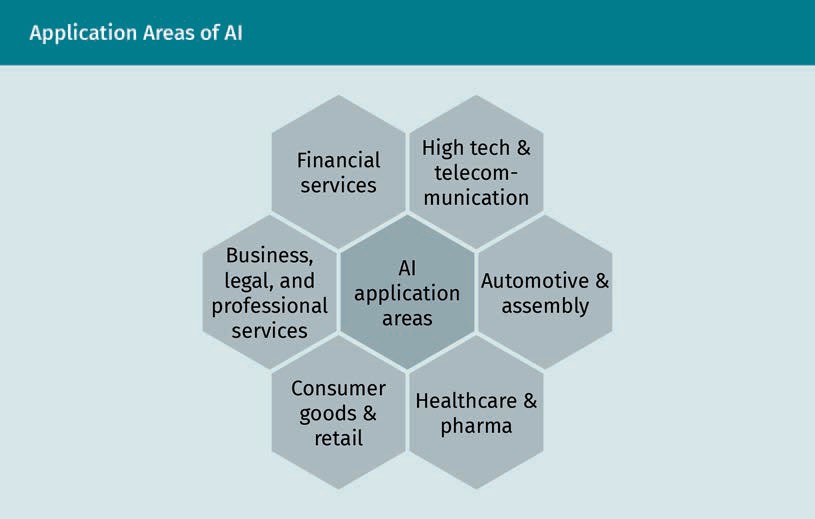
### Anwendungsbereiche

Dank der jüngsten Fortschritte bei den Rechen- und Datenspeicherkapazitäten haben die Anwendungen für KI in den letzten Jahren kontinuierlich zugenommen. Die Möglichkeiten für den Einsatz von KI sind nahezu unbegrenzt.

Das wachsende Interesse wird auch durch eine Zunahme der Forschungsaktivitäten untermauert. Laut dem jährlichen KI-Index (Zhang *et al*., 2021) ist die Zahl der Zeitschriftenveröffentlichungen über KI von 2019 bis 2020 um 34,5 Prozent gestiegen. Seit 2010 haben sich die KI-Fachartikel mehr als verzwanzigfacht. Die populärsten Forschungsthemen sind NLP und Computer Vision, die für verschiedene Anwendungsbereiche von Bedeutung sind.

KI-Einführung Die Nutzung von KI-Fähigkeiten wie Machine Learning in mindestens einer Geschäftsfunktion wird als KI-Einführung bezeichnet.

In einer weltweiten Umfrage über den Stand der KI hat McKinsey & Company (2021) die folgenden Branchen als die wichtigsten Bereiche für die **KI-Einführung** identifiziert: Hochtechnologie/Telekommunikation, Automobilbau und Fertigung, Finanzdienstleistungen, Unternehmen, Rechtsberatung und Dienstleistungen, Gesundheitswesen/Pharma und Konsumgüter/Einzelhandel. Im folgenden Abschnitt werden wir uns diese Felder genauer ansehen.



Die Abbildung oben fasst die wichtigsten Domänen zusammen, in denen KI eingesetzt wird.

###### Hochtechnologie und Telekommunikation

Aufgrund der ständigen Zunahme des weltweiten Netzwerkverkehrs und der Netzwerkausrüstung hat die KI in der Telekommunikation ein schnelles Wachstum erfahren. In diesem Bereich kann KI nicht nur zur Optimierung und Automatisierung von Netzwerken eingesetzt werden, sondern auch um sicherzustellen, dass die Netzwerke funktionsfähig und sicher bleiben.

Moderne KI-Systeme

Durch den Einsatz von KI bei der vorausschauenden Wartung können Netzwerkprobleme behoben werden, noch bevor sie auftreten. Darüber hinaus können Netzwerkanomalien bei der Verwendung von selbstoptimierenden Netzwerken genau vorhergesagt werden.

Big Data ermöglicht es zudem, Netzwerkanomalien mühelos zu erkennen und somit betrügerisches Verhalten in Netzwerken zu verhindern.

###### Automobilindustrie und Fertigung

In den letzten Jahren hat sich das autonome Fahren zu einem wichtigen Forschungsthema entwickelt. Sie wird die Automobilindustrie in den nächsten Jahrzehnten drastisch von einer stahlbasierten zu einer softwarebasierten Industrie verändern. Schon heute sind Autos mit zahlreichen Sensoren ausgestattet, um die Sicherheit des Fahrers zu gewährleisten, z. B. um in der Spur zu bleiben oder eine Notbremsung zu unterstützen.

Intelligente Sensoren können außerdem technische Probleme am Fahrzeug oder Risiken durch den Fahrer – wie Müdigkeit oder Alkoholeinfluss – erkennen und entsprechende Maßnahmen einleiten.

Wie in der Hochtechnologie und Telekommunikation kann KI auch in der Fertigung für die vorausschauende Wartung und die Behebung von Ineffizienzen in der Fertigungslinie eingesetzt werden. Außerdem ist es mithilfe von Computer Vision bereits möglich, Produktfehler schneller und genauer zu erkennen als dies für Menschen möglich ist.

###### Finanzdienstleistungen

Finanzdienstleistungen bieten zahlreiche Anwendungsmöglichkeiten für Künstliche Intelligenz. Intelligente Algorithmen ermöglichen es Finanzinstituten, betrügerische Transaktionen und Geldwäsche wesentlich früher zu erkennen und zu verhindern, als dies bisher möglich war. Computer-Vision-Algorithmen können verwendet werden, um gefälschte Unterschriften präzise zu identifizieren, indem sie mit Scans der in einer Datenbank gespeicherten Originale verglichen werden.

Darüber hinaus nutzen viele Banken und Makler bereits Robo-Advising. Auf der Grundlage des Anlageprofils eines Nutzers können genaue Empfehlungen für zukünftige Investitionen gegeben werden (D'Acunto *et al*., 2019). Portfolios können ebenso auf der Grundlage von KI-Anwendungen optimiert werden.

###### Geschäftswesen, Rechtswesen und professionelle Dienstleistungen

Besonders in Branchen, in denen Bürokratie und sich wiederholende Aufgaben eine wichtige Rolle spielen, kann KI helfen, Prozesse schneller und effizienter zu gestalten.

Wesentliche Teile von Routineabläufen werden derzeit mithilfe von **Robotic Process Automation** (RPA) automatisiert, was die Verwaltungskosten drastisch senken kann. RPA-Systeme müssen nicht unbedingt mit intelligenten KI-Funktionen ausgestattet sein. Methoden wie NLP und Computer Vision können jedoch dazu beitragen, diese Prozesse durch eine intelligentere Geschäftslogik zu verbessern.

Robotic Process Automation

Die automatisierte Ausführung von sich wiederholenden, manuellen, zeitaufwendigen oder fehleranfälligen Aufgaben durch Software-Bots wird als Robotic Process Automation bezeichnet.

Die laufenden Entwicklungen im Bereich der Big-Data-Technologien können Unternehmen dabei helfen, mehr Informationen aus ihren Daten zu gewinnen. Mithilfe von Predictive Analytics können aktuelle und zukünftige Trends in den Märkten, in denen ein Unternehmen tätig ist, erkannt und entsprechend reagiert werden.

Ein weiterer wichtiger Anwendungsfall ist die Verringerung von Risiken und Betrug, insbesondere in Rechts-, Buchhaltungs- und Beratungspraxen. Intelligente Agenten können dabei helfen, potenziell betrügerische Muster zu erkennen, was eine frühere Reaktion ermöglicht.

###### Gesundheitswesen und Pharmaindustrie

In den letzten Jahren waren das Gesundheitswesen und die Pharmaindustrie die am schnellsten wachsenden Bereiche, in denen KI eingesetzt wurde.

KI-basierte Systeme können helfen, Krankheiten anhand der Symptome zu erkennen. So konnten KI-basierte Systeme in jüngsten Studien eingesetzt werden, um COVID-19 anhand von Hustenaufzeichnungen zu erkennen (Laguarta *et al*., 2020).

Nicht nur in der Diagnostik kann KI viele Vorteile bieten. Intelligente Agenten können eingesetzt werden, um Patienten entsprechend ihren Bedürfnissen zu überwachen. Außerdem kann KI bei der optimalen Kombination von Medikamenten helfen, Nebenwirkungen zu vermeiden.

Mit Wearables – also am Körper tragbaren Geräten wie z. B. Herzfrequenz- oder Körpertemperaturmessgeräten – können die Vitalparameter einer Person ständig überwacht werden. Anhand dieser Daten kann ein Agent Empfehlungen zum Gesundheitszustand des Trägers geben. Falls kritische Anomalien festgestellt werden, kann ein Notruf abgesetzt werden.

###### Konsumgüter und Einzelhandel

In der Konsumgüterindustrie steht wie im Einzelhandel die Vorhersage des Kundenverhaltens im Fokus. Websites verfolgen anhand der Anzahl der Besuche, wie sich das Profil eines Kunden verändert. Dies ermöglicht persönliche Kaufvorhersagen für jeden Kunden. Diese Daten können nicht nur verwendet werden, um personalisierte Einkaufsempfehlungen abzugeben, sondern auch, um die gesamte Lieferkette zu optimieren und über künftige Forschung zu informieren.

Die Marktsegmentierung basiert heutzutage nicht mehr auf geografischen Regionen wie dem Bundesland oder Land. Moderne Technologien ermöglichen es, das Verhalten der Kunden auf Straßenbasis zu segmentieren. Diese Informationen können zur Feinabstimmung der Betriebsabläufe und zur Entscheidung über die Erhaltung oder Schließung von Filialen verwendet werden.

Außerdem wird die jüngste Verbesserung der NLP-Technologien zunehmend für Chatbots und Konversationsschnittstellen genutzt. Wenn es um Kundenbindung und Kundenservice geht, ist ein gut entwickelter künstlicher Agent der Schlüssel zur Kundenzufriedenheit.

Moderne KI-Systeme

###### Beurteilung von KI-Systemen

Wie die oben erwähnten Beispiele zeigen, sind die Anwendungsbereiche für moderne KI-Systeme nahezu unbegrenzt. Es ist wichtig, dass alle Datensätze unabhängig voneinander sind und einer ähnlichen Wahrscheinlichkeitsverteilung folgen.

Um geeignete Modelle für KI-Anwendungen zu entwickeln, werden die verfügbaren Daten in drei Datensätze aufgeteilt:

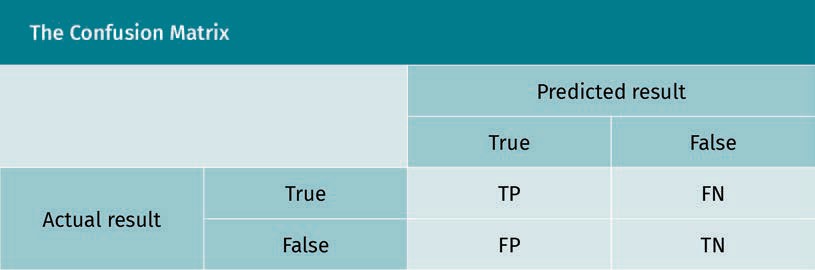
1. Trainingsdatensatz: Wie der Name schon sagt, wird dieser Datensatz verwendet, um die Parameter eines Algorithmus während des Trainingsprozesses anzupassen.
2. Entwicklungsdatensatz: Dieser Datensatz wird oft auch als Validierungssatz bezeichnet. Er wird verwendet, um die Leistung des mit dem Trainingsdatensatz entwickelten Modells zu bewerten sowie für weitere Optimierungen. Es ist wichtig, dass der Entwicklungsdatensatz Daten enthält, die nicht in den Trainingsdaten enthalten sind.
3. Testdatensatz: Sobald das Modell mithilfe des Trainings- und des Entwicklungsdatensatzes fertiggestellt ist, kann der Testdatensatz für eine endgültige Beurteilung des Modells verwendet werden. Wie beim Entwicklungsdatensatz ist es wichtig, dass die Daten im Testdatensatz noch nicht verwendet wurden. Der Testdatensatz wird nur einmal verwendet, um das Modell zu validieren und um sicherzustellen, dass es kein Overfitting gibt.

Bei der Entwicklung und Abstimmung von Algorithmen sollten Kennzahlen vorhanden sein, um zu beurteilen, wie gut sie unabhängig und im Vergleich zu anderen Systemen funktionieren. Bei einer binären Klassifizierungsaufgabe sind Genauigkeit, Relevanz, Trefferquote und F-Maß Metriken, die üblicherweise für diesen Zweck verwendet werden.

Zum Beispiel verwenden Finanzdienstleister eine binäre Klassifizierungsaufgabe bei der Betrugserkennung. Eine finanzielle Transaktion kann entweder als Betrug eingestuft werden oder nicht. Auf dieser Grundlage erhalten wir vier Kategorien von Klassifizierungsergebnissen:

1. Richtig Positive (True Positives, TP): Stichproben, die korrekt als positiv klassifiziert wurden, d. h. bei denen es sich um betrügerische Transaktionen handelt
2. Falsch Positive (False Positives, FP): alle Ergebnisse, die eine Stichprobe fälschlicherweise als positiv einstuft, obwohl sie negativ ist, d. h. eine nicht betrügerische Transaktion wird als Betrug eingestuft
3. Richtig Negative (True Negatives, TN): Klassifizierungsergebnisse, die korrekt als negativ eingestuft wurden, d. h. nicht betrügerische Transaktionen, die auch als solche eingestuft wurden
4. Falsch Negative (False Negatives, FN): Klassifizierungsergebnisse, die fälschlicherweise als negativ eingestuft wurden, obwohl sie positiv hätten sein müssen, d. h. betrügerische Transaktionen, die als nicht betrügerisch eingestuft wurden

Die Klassifizierungsergebnisse können in einer Konfusionsmatrix angezeigt werden, die auch als Fehlermatrix bezeichnet wird. Dies sehen wir in der folgenden Tabelle.



Anhand dieser Kategorien können die oben erwähnten Metriken berechnet werden.

Die Genauigkeit ist ein Indikator dafür, wie viele Stichproben korrekt klassifiziert wurden. Sie lässt sich wie folgt berechnen:

Genauigkeit = TP + TN

TP + TN + FP + FN

Es wird gemessen, wie viel Prozent der gesamten Vorhersage korrekt war. Die Relevanz (engl. Precision) gibt die Anzahl der positiven Stichproben an, die im Verhältnis zu allen in dieser Klasse vorhergesagten Stichproben korrekt klassifiziert wurden:

Relevanz = TP

TP + FP

Die Trefferquote (engl. Recall) gibt an, wie viele der positiv erkannten Stichproben im Verhältnis zur Gesamtzahl der Stichproben, die als solche hätten identifiziert werden müssen, korrekt identifiziert wurden:

Trefferquote = TP

TP + FN

Das F-Maß schließlich kombiniert Relevanz und Trefferquote in einem Ergebnis:

F = 2 · Relevanz · Trefferquote

Relevanz + Trefferquote

Bei Klassifizierungsaufgaben mit mehr als zwei Klassen können die Metriken für jede Klasse berechnet werden. Am Ende kann der Durchschnitt der Werte zu einer Metrik für alle Klassen zusammengefasst werden.

Auch wenn in den aktuellen Anwendungsbereichen der KI noch enge KI-Algorithmen zum Einsatz kommen, gibt es viele mögliche Anwendungsfälle in verschiedenen Branchen. Immer mehr Unternehmen schaffen es, ihre Geschäftsmodelle mit KI zu unterstützen oder sogar ganz neue zu schaffen.

Moderne KI-Systeme

**Zusammenfassung**

Es gibt zwei Arten von KI: enge und allgemeine. Aktuelle KI-Systeme gehören alle in die Kategorie der engen KI (ANI). ANI kann komplexe Probleme schneller lösen als Menschen. Ihre Fähigkeiten sind jedoch auf die Domäne beschränkt, für die sie programmiert wurde. Auch wenn der Begriff ANI eine Einschränkung suggerieren mag, ist das Prinzip in vielen Bereichen unseres Lebens fest verankert. Im Gegensatz dazu bleibt AGI (KI, die über die kognitiven Fähigkeiten verfügt, Wissen auf andere Anwendungsbereiche zu übertragen) ein theoretisches Konstrukt, ist aber weiterhin ein wichtiges Forschungsthema.

Die Anwendungsbereiche für KI sind nahezu unbegrenzt. KI hat einen erheblichen Einfluss auf die heutige Unternehmenslandschaft. Anwendungsfälle wie die Optimierung von Serviceabläufen, die Verbesserung von Produkten auf der Grundlage von KI und die Automatisierung manueller Prozesse können Unternehmen dabei helfen, ihre Geschäftsfunktionen zu optimieren. Diese Anwendungsfälle erstrecken sich über eine Vielzahl von Branchen, sei es die Automobil- und Fertigungsindustrie, Finanzdienstleistungen, Gesundheitswesen und Pharmazie, Konsumgüter und viele mehr.



# Lektion 3

## Bestärkendes Lernen

#### LERNZIELE

Nach der Bearbeitung dieser Lektion können Sie …

… die Grundprinzipien des bestärkenden Lernens erklären.

… Markow-Entscheidungsprozesse nachvollziehen.

… den Q-Lernen-Algorithmus verwenden.

DL-E-DLBDSEAIS01-U03

1. Bestärkendes Lernen

### Einführung

Stellen wir uns vor, wir hätten uns in einem Labyrinth verirrt und müssten den Weg hinaus finden. Da wir zum ersten Mal dort sind, wissen wir nicht, welchen Weg wir wählen sollen, um zur Tür zu gelangen und das Labyrinth zu verlassen. Außerdem gibt es im Labyrinth gefährliche Felder, die wir nicht betreten dürfen.

Wir können vier Aktionen ausführen: uns nach oben, nach unten, nach links oder nach rechts bewegen. Da wir das Labyrinth nicht kennen, können wir den Weg hinaus nur durch zufällige Aktionen finden. Im Laufe des Lernprozesses werden wir herausfinden, dass es im Labyrinth Felder gibt, die uns belohnen, indem wir aus dem Labyrinth entkommen können. Es gibt jedoch auch Felder, für die wir eine Strafe erhalten, da es gefährlich ist, sie zu betreten. Nach einiger Zeit werden wir es schaffen, den Weg nach draußen zu finden, ohne auf die gefährlichen Felder zu treten, die wir beim Herumlaufen entdeckt haben. Dieser Prozess des Lernens durch Belohnung wird als bestärkendes Lernen (engl. Reinforcement Learning) bezeichnet.



In dieser Lektion erfahren wir mehr über die grundlegenden Ideen des bestärkenden Lernens und die zugrunde liegenden Prinzipien. Außerdem lernen wir Algorithmen kennen, wie z. B. das Q-Lernen, mit denen wir den Lernvorgang optimieren können.

### Was ist bestärkendes Lernen?

Im Allgemeinen gibt es beim maschinellen Lernen drei Techniken, um ein bestimmtes Lernmodell zu trainieren: überwachtes, unüberwachtes und bestärkendes Lernen.

Bestärkendes Lernen

Beim überwachten Lernen lernt eine Maschine auf der Grundlage eines zuvor markierten Datensatzes, wie sie ein Problem lösen kann. Typische Anwendungsbereiche für überwachtes Lernen sind Regressions- und Klassifizierungsprobleme wie die Einschätzung von Kreditrisiken oder die Erkennung von Spam. Das Training dieser Art von Algorithmen ist sehr mühsam, da es eine große Menge an vormarkierten Trainingsdaten erfordert.

Beim unüberwachten Lernen wird das Training mit unmarkierten Daten durchgeführt und die Maschine erkennt selbst Muster in den Daten. Auf der Grundlage der Eingabedaten werden Cluster identifiziert, die später zur Klassifizierung verwendet werden können. Dieser Ansatz wird häufig verwendet, um riesige Mengen unstrukturierter Daten, wie z. B. das Kundenverhalten, zu organisieren und relevante Vergleichsgruppen zu identifizieren.

Die Techniken des bestärkenden Lernens verfolgen einen eher explorativen Ansatz. Algorithmen, die auf diesem Ansatz basieren, verbessern sich selbst, indem sie mit der Umgebung interagieren. Im Gegensatz zum überwachten und unüberwachten Lernen sind keine vordefinierten Daten erforderlich. Ein Agent lernt aus einer unbekannten Menge von Daten auf der Grundlage der Belohnung, die er von seiner Umgebung erhält. Die folgende Tabelle fasst die Grundbegriffe des bestärkenden Lernens zusammen.

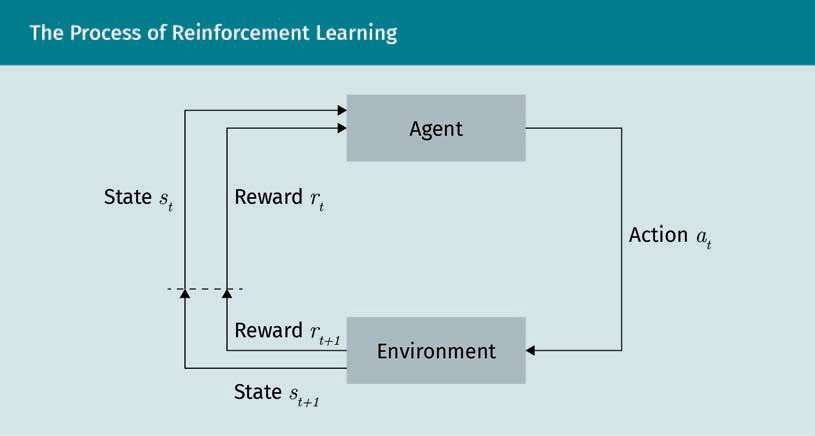
|  |  |
| --- | --- |
| Grundbegriffe des bestärkenden Lernens | |
| Agent | Führt Aktionen in einer Umgebung aus und erhält dafür eine Belohnung |
| Aktion (Action, A) | Die Menge aller möglichen Aktionen, die der Agent ausführen kann |
| Umwelt (Environment, E) | Das Szenario, das der Agent erkunden muss |
| Zustand (State, S) | Der aktuelle Zustand des Agenten in der gegebenen Umgebung |
| Belohnung (Reward, R) | Unmittelbares Feedback aus der Umgebung, um die Aktion eines Agenten zu belohnen |
| Policy (π) | Die Policy (dt. auch Strategie) ist das Verfahren, das der Agent anwendet, um die nächste Aktion auf der Grundlage des aktuellen Zustands zu bestimmen |
| Wert (Value, V ) | Der langfristige Wert des aktuellen Zustands S unter Verwendung der Policy π |

Innerhalb des Prozesses des bestärkenden Lernens beginnt der Agent in einem bestimmten Zustand st ∈ S und wendet eine Aktion at ∈ A st auf die Umgebung E an, wobei A st die Menge der im Zustand st verfügbaren Aktionen ist. Die Umwelt reagiert, indem sie einen neuen Zustand st + 1 und eine Belohnung rt + 1 an den Agenten zurückgibt. Im nächsten Schritt wendet der Agent die nächste Aktion at + 1 auf die Umgebung an, die wiederum einen neuen Zustand und eine Belohnung ergibt.

In dem einleitenden Beispiel agieren wir als Agent in der Labyrinthumgebung. Die Aktionen, die wir ausführen können, sind nach oben, unten, links oder rechts zu gehen. Nach jedem Zug erreichen wir einen anderen Zustand, indem wir auf ein anderes Feld im Labyrinth ziehen. Jedes Mal, wenn wir eine Aktion durchführen, erhalten wir eine Belohnung aus der Umgebung. Diese wird positiv sein, wenn wir die Tür erreichen, oder negativ, wenn wir auf ein gefährliches Feld treten. Von unserer neuen Position aus beginnt der ganze Lernzyklus von neuem. Unser Ziel ist es, unsere Belohnung zu maximieren. Der Prozess des Erhaltens einer Belohnung als Funktion eines Zustand-Aktion-Paares lässt sich wie folgt formalisieren:

f st, at = rt + 1

Der gesamte Prozess der Interaktion zwischen Agent und Umgebung ist in der folgenden Abbildung dargestellt.



Der Prozess, bei dem eine Aktion aus einem bestimmten Zustand ausgewählt wird, in einen neuen Zustand übergeht und eine Belohnung erhält, wiederholt sich. Für eine Folge von diskreten Zeitschritten t = 0, 1, 2, … , die mit dem Zustand s0 ∈ S beginnen, führt die Agent-Umgebung-Interaktion zu einer Folge:

s0, a0, r1, s1, a1, r2, s2, a2, r3, s3, …

Das Ziel des Agenten ist es, die Belohnung zu maximieren, die er während des Lernprozesses erhält. Der Zyklus wird fortgesetzt, bis der Agent in einem Endzustand endet. Die Gesamtbelohnung R nach einer Zeit T kann als die Summe der zu diesem Zeitpunkt erhaltenen Belohnungen berechnet werden:

Rt = rt + 1 + rt + 2 + … + rT

Diese Belohnung wird auch als der Wert V π s im Status s bei Verwendung der Strategie π bezeichnet. In unserem Beispiel erhalten wir die maximale Belohnung, sobald wir den Ausgang des Labyrinths erreicht haben. Wir werden uns die Wertfunktion im nächsten Abschnitt genauer ansehen.

Bestärkendes Lernen

### Markow-Entscheidungsprozess und Wertfunktion

Damit wir verschiedene Wege im Labyrinth bewerten können, brauchen wir einen geeigneten Ansatz zum Vergleich von Interaktionsfolgen. Eine Methode zur Formalisierung der sequentiellen Entscheidungsfindung sind Markow-Entscheidungsprozesse (engl. Markov Decision Process, MDP). Im Folgenden werden wir erörtern, wie MDPs funktionieren.

###### Der Markow-Entscheidungsprozess

MDPs werden verwendet, um die Wahrscheinlichkeit eines zukünftigen Ereignisses auf der Grundlage einer Folge möglicher Ereignisse abzuschätzen. Wenn ein gegenwärtiger Zustand alle relevanten Informationen über vergangene Aktionen enthält, wird davon gesprochen, dass er über die „Markow-Eigenschaft“ verfüge. Beim bestärkenden Lernen ist die Markow-Eigenschaft von entscheidender Bedeutung, da alle Entscheidungen und Werte Funktionen des aktuellen Zustands sind (Sutton & Barto, 2018), d. h. Entscheidungen werden in Abhängigkeit vom Zustand der Umgebung getroffen.

Wenn eine Aufgabe beim bestärkenden Lernen die Markow-Eigenschaft erfüllt, wird diese Aufgabe als Markow-Entscheidungsprozess bezeichnet. Wenn die Markow-Eigenschaft erfüllt ist, hängt in jedem Zustand einer Markow-Kette die Wahrscheinlichkeit, dass ein anderer Zustand erreicht wird, nur von zwei Faktoren ab: der Übergangswahrscheinlichkeit, den nächsten Zustand zu erreichen, und dem aktuellen Zustand. Die Wahrscheinlichkeit, einen anderen Zustand zu erreichen, hängt ausschließlich vom aktuellen Zustand sowie von der Übergangswahrscheinlichkeit ab, den nächsten Zustand zu erreichen. MDPs bestehen aus den folgenden Komponenten:

* + - Zuständen S
    - Aktionen A
    - Belohnungen für eine Aktion in einem bestimmten Zustand ra = R s, a, s′
    - Übergangswahrscheinlichkeiten der Aktionen für den Übergang von einem Zustand in den nächsten Zustand

Ta s, s′

Aufgrund der Markow-Eigenschaft hängt die Übergangsfunktion ausschließlich vom aktuellen Zustand ab:

P st + 1 st, at, st − 1, at − 1, … = P st + 1 st, at = Tat s, s′

Welche Aktion in einem bestimmten Zustand gewählt wird, wird durch die Policy π beschrieben:

π s, a = p at = a st = s

In unserem Beispiel des Labyrinths gibt die Position, an der wir stehen, keine Auskunft über die Abfolge der Zustände, die wir genommen haben, um dorthin zu gelangen. Unsere Position im Labyrinth stellt jedoch alle Informationen dar, die für die Entscheidung über unseren nächsten Zustand erforderlich sind, d. h. sie hat die Markow-Eigenschaft.

###### Die Wertfunktion

Zusätzlich zu den zuvor erläuterten Konzepten verwenden die Algorithmen des bestärkenden Lernens auch Wertfunktionen. Wertfunktionen geben eine Einschätzung darüber ab, wie gut es für einen Agenten ist, sich in diesem Zustand zu befinden und eine bestimmte Aktion in diesem Zustand durchzuführen (Sutton & Barto, 2018).

Zuvor haben wir gelernt, dass der Wert eines Zustands als Summe der während des Lernprozesses erhaltenen Belohnungen berechnet werden kann. Zusätzlich kann ein Diskontierungsfaktor verwendet werden, um die Vorteile zukünftiger Aktionen zum gegenwärtigen Zeitpunkt zu bewerten. Der Diskontierungsfaktor gibt die Wahrscheinlichkeit an, einen Belohnungszustand in der Zukunft zu erreichen. Dies hilft dem Agenten, seine Aktionen genauer nach der erwarteten Belohnung auszuwählen. Eine Aktion at + 1 wird dann so gewählt, dass die erwartete diskontierte Rendite maximiert wird:

V π s =Eπ rt + 1 + γr t + 2 + … + γ T − 1 rT

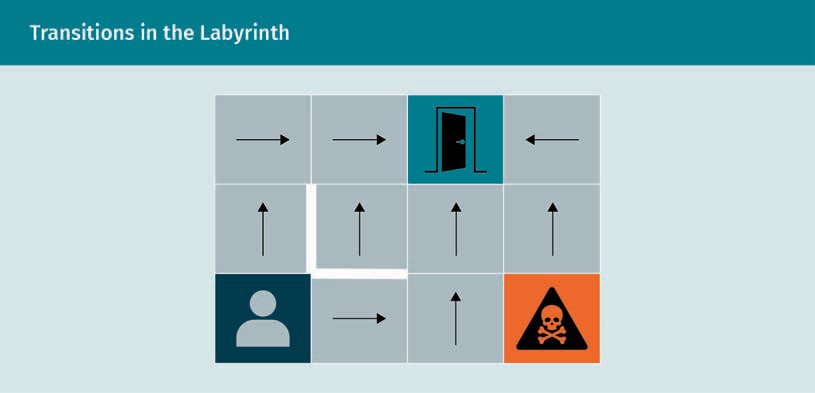
∞

=Eπ ∑ γkrt + k + 1 st = s

k = 0

wobei γ der Diskontierungsfaktor ist, mit 0 ≤ γ ≤ 1, der die Sicherheit der erwarteten Rendite bezeichnet. Je näher der Wert von γ bei 1 liegt, desto höher ist die Wahrscheinlichkeit für zukünftige Belohnungen. Vor allem in Szenarien, in denen die Dauer des Prozesses nicht im Voraus bekannt ist, ist es wichtig, γ < 1 zu setzen, da die Wertfunktion sonst nicht konvergiert.

Die folgende Abbildung zeigt, welche Aktion der Agent in den jeweiligen Zuständen des Labyrinths optimalerweise ausführen sollte, um die Belohnung zu maximieren, d. h. zu versuchen, den Ausgang zu erreichen und das gefährliche Feld zu meiden.



Bestärkendes Lernen

### Temporales Differenzlernen und Q-Lernen

Bisher haben wir über modellbasiertes bestärkendes Lernen gesprochen. Das bedeutet, dass ein Agent versucht, das Modell der Umgebung zu verstehen. Alle Entscheidungen beruhen auf einer Wertfunktion. Diese Wertfunktion basiert auf dem aktuellen Zustand und dem zukünftigen Zustand, den der Agent erreichen wird.

Im Gegensatz dazu analysieren modellfreie Konzepte die Qualität einer Aktion, um ihre Aktionen zu bewerten. Q-Lernen ist ein sehr bekannter Algorithmus für modellfreies bestärkendes Lernen und basiert auf dem Konzept des temporalen Differenzlernens. Im Folgenden werden wir die zugrunde liegenden Konzepte des temporalen Differenz- und des Q-Lernens erklären.

###### Temporales Differenzlernen

Da das temporale Differenzlernen (engl. Temporal Difference Learning, auch TD-Lernen) ein modellfreier Ansatz ist, wird hier kein Modell der Lernumgebung benötigt. Stattdessen erfolgt das Lernen direkt aus der Erfahrung in einem teilweise unbekannten System. Wie der Name schon sagt, trifft das TD-Lernen Vorhersagen auf der Grundlage der Tatsache, dass es oft eine Korrelation zwischen aufeinander folgenden Vorhersagen gibt. Das prominenteste Beispiel zur Veranschaulichung des Prinzips des TD-Lernens von Sutton (1988) betrifft die Vorhersage des Wetters. Nehmen wir an, wir wollen das Wetter an einem Montag vorhersagen. Bei einem Ansatz des überwachten Lernens würde man die Vorhersage eines jeden Tages heranziehen und sie mit dem tatsächlichen Ergebnis vergleichen. Das Modell würde aktualisiert, sobald es Montag ist. Im Gegensatz dazu vergleicht ein TD-Ansatz die Vorhersage eines jeden Tages mit der Vorhersage des folgenden Tages, d. h. er berücksichtigt die zeitliche Differenz zwischen den aufeinanderfolgenden Tagen und aktualisiert die Vorhersagen eines Tages auf der Grundlage des Ergebnisses des vorherigen Tages. Daher macht TD-Lernen im Laufe der Zeit besseren Gebrauch von der Erfahrung.

###### Q-Lernen

Ein bekannter Algorithmus, der auf dem TD-Lernen basiert, ist das Q-Lernen. Nach der Initialisierung führt der Agent zufällige Aktionen aus, die dann ausgewertet werden. Auf der Grundlage des Ergebnisses einer Aktion passt der Agent sein Verhalten für die nachfolgenden Aktionen an.

Das Ziel des Q-Lernalgorithmus ist die Maximierung der Qualitätsfunktion Q s, a. Das Ziel ist es, die kumulative Belohnung zu maximieren, während man sich in einem bestimmten Zustand s befindet, indem man die beste Aktion a vorhersagt (van Otterlo & Wiering, 2012). Während des Lernprozesses wird Q s, a anhand der **Bellman-Gleichung** iterativ aktualisiert:

Q s, a = r + γmaxa′Q s′, a′

Alle Q-Werte, die während des Lernprozesses berechnet werden, werden in der Q-Matrix gespeichert. In jeder Iteration wird die Matrix verwendet, um die bestmögliche Aktion zu finden. Wenn der Agent eine neue Aktion ausführen muss, sucht er nach dem maximalen Q-Wert des Zustand-Aktion-Paares.

Bellman-Gleichung Die Bellman-Gleichung berechnet die erwartete Belohnung in einem MDP, wenn eine Aktion in einem bestimmten Zustand ausgeführt wird. Die Belohnung ist unterteilt in die

sofortige und die gesamte zukünftig erwartete

Belohnung.

Der Q-Lernalgorithmus

Im Folgenden werden wir den Q-Lernalgorithmus näher erläutern. Der Algorithmus besteht aus einer Initialisierungs- und einer Iterationsphase. In der Initialisierungsphase werden alle Werte in der Q-Tabelle auf 0 gesetzt. In der Iterationsphase führt der Agent die folgenden Schritte aus:

1. Wähle eine Aktion für den aktuellen Zustand. In dieser Phase gibt es zwei verschiedene Strategien, die verfolgt werden können:
   * Erkundung: Führe zufällige Aktionen durch, um mehr Informationen über die Umgebung zu erhalten
   * Ausnutzung: Führe Aktionen auf der Grundlage der Informationen durch, die bereits über die Umgebung anhand der Q-Tabelle bekannt sind. Das Ziel ist die Maximierung der Rendite
2. Führe die gewählte Aktion aus
3. Bewerte das Ergebnis und ermittle den Wert der Belohnung. Auf der Grundlage des Ergebnisses wird die Q-Tabelle aktualisiert.

**Zusammenfassung**

Beim bestärkenden Lernen geht es darum, die beste Strategie dafür zu finden, wie sich ein Agent in einer Umgebung verhalten sollte, um ein bestimmtes Ziel zu erreichen. Der Lernprozess dieses Agenten erfolgt auf der Grundlage eines Belohnungssystems, das den Agenten entweder für gute Entscheidungen belohnt oder für schlechte Entscheidungen bestraft.

Um den Prozess der Bewegung des Agenten in der Umgebung zu modellieren, können Markow-Entscheidungsprozesse verwendet werden. Eine Wertfunktion kann auf das System angewendet werden, um die Qualität zukünftiger Entscheidungen besser zu bewerten.

Der Q-Lernalgorithmus ist ein modellfreier Ansatz des temporalen Differenzlernens, bei dem der Agent Informationen über die Umgebung auf der Grundlage von Erkundung und Ausnutzung sammelt.

Insgesamt ist der Prozess des bestärkenden Lernens dem Lernen durch Versuch und Irrtum im echten Leben sehr ähnlich.



# Lektion 4

## Natural Language Processing (NLP)

#### LERNZIELE

Nach der Bearbeitung dieser Lektion können Sie …

… den historischen Hintergrund der Computerlinguistik erklären.

… die wichtigsten Anwendungsbereiche benennen.

… zwischen statistischen und regelbasierten Techniken der Computerlinguistik unterscheiden.

… nachvollziehen, wie Daten vektorisiert werden.

DL-E-DLBDSEAIS01-U04

1. Natural Language Processing (NLP)

### Einführung

Natural Language Processing (NLP), also die Verarbeitung natürlicher Sprache durch Computer, ist eine der wichtigsten Domänen der Künstlichen Intelligenz.

Die Computerlinguistik kann in drei Teilbereiche unterteilt werden: Spracherkennung, Sprachverständnis und Spracherzeugung. Jedes dieser Themen wird in den folgenden Abschnitten behandelt. Nach einer Einführung in NLP und seine Anwendungsbereiche erfahren wird mehr über die grundlegenden NLP-Techniken und wie die Datenvektorisierung funktioniert.

### Einführung in NLP und Anwendungsbereiche

NLP ist ein interdisziplinäres Gebiet mit Wurzeln in der Informatik (insbesondere im Bereich der Künstlichen Intelligenz), der Kognitionswissenschaft und der Linguistik. NLP befasst sich mit der Verarbeitung, dem Verständnis und der Erzeugung natürlicher Sprache (Kaddari *et al*., 2021). Bei der Interaktion zwischen Mensch und Computer spielt NLP eine Schlüsselrolle, wenn es darum geht, die Interaktion auf natürlichere Weise zu gestalten. Das Ziel von NLP ist es daher, Sprache auf einer ähnlichen Ebene wie Menschen zu verwenden und zu interpretieren. Es gibt viele interessante Anwendungsfälle, die von der automatischen maschinellen Übersetzung bis zur Generierung von Textauszügen oder sogar ganzen Literaturwerken reichen. Wie bereits erwähnt, gibt es drei Teilbereiche im NLP:

1. Spracherkennung: identifiziert Wörter in gesprochener Sprache und umfasst die Sprache-zu-Text-Umwandlung
2. Verständnis natürlicher Sprache: extrahiert die Bedeutung von Wörtern und Sätzen sowie Leseverstehen
3. Erzeugung natürlicher Sprache: ist die Fähigkeit, sinnvolle Sätze und Texte zu erzeugen.

Alle diese Teilbereiche bauen auf Methoden der Künstlichen Intelligenz auf und bilden die Grundlage für die Anwendungsbereiche von NLP.

###### Historische Entwicklung

Die frühe Forschung zur Verarbeitung natürlicher Sprache geht auf das siebzehnte Jahrhundert zurück, als Descartes und Leibnitz einige frühe theoretische Untersuchungen zum Themenbereich des NLP durchführten (Schwartz, 2019). Mitte der 1950er Jahre wurde daraus eine technische Disziplin. Die geopolitischen Spannungen zwischen der ehemaligen Sowjetunion und den Vereinigten Staaten führten zu einer steigenden Nachfrage nach Englisch-Russisch-Übersetzerinnen und -Übersetzern. Deshalb versuchte man, die Übersetzung an Maschinen auszulagern. Obwohl die ersten Ergebnisse vielversprechend waren, erwies sich die maschinelle Übersetzung als wesentlich komplexer als ursprünglich angenommen, zumal keine nennenswerten Fortschritte erzielt werden

Natural Language Processing (NLP)

konnten. Im Jahr 1964 stufte das Automatic Language Processing Advisory Committee die NLP-Technologie als „hoffnungslos“ ein und beschloss, die Forschungsförderung in diesem Bereich vorübergehend einzustellen. Dies wurde als der Beginn des NLP-Winters betrachtet.

Fast 20 Jahre nach Beginn des NLP-Winters begann NLP wieder das Interesse auf sich zu ziehen. Dies ist auf die folgenden drei Entwicklungen zurückzuführen:

1. Erhöhung der Rechenleistung: Die Rechenleistung stieg deutlich an und ermöglichte dem Mooreschen Gesetz folgend rechenintensivere Algorithmen.
2. Paradigmenwechsel: Frühe Sprachmodelle basierten auf einem grammatikalischen Ansatz, der versuchte, komplexe regelbasierte Systeme zu implementieren, die der Komplexität der Alltagssprache gerecht werden sollten. Die neuere Forschung hat sich auf Modelle verlagert, die auf statistischen und entscheidungstheoretischen Grundlagen beruhen, wie z. B. Entscheidungsbäume.
3. Part-of-Speech-Tagging (POS): Bei dieser Technik wird ein Text in kleinere Einheiten aufgeteilt, d. h. in einzelne Sätze, Wörter oder Unterwörter. Mithilfe von POS-Tagging werden grammatikalische Wortfunktionen und Kategorien zu einem gegebenen Text hinzugefügt. Dies ermöglicht es, Sprache mithilfe von **Markow-Modellen** zu beschreiben. Im Gegensatz zu Ansätzen, die die gesamte Geschichte berücksichtigen, bedeutet dies eine erhebliche Reduzierung der Komplexität.

Zusammengenommen hat der Wechsel zu statistischen, entscheidungstheoretischen und Machine-Learning-Modellen die Robustheit von NLP erhöht, insbesondere was die Fähigkeit betrifft, mit unbekannten Konstellationen umzugehen. Darüber hinaus ermöglichte die verbesserte Rechenleistung die Verarbeitung einer viel größeren Menge von Trainingsdaten, die nun zudem aufgrund der wachsenden Menge an elektronischer Literatur zur Verfügung standen. Dies eröffnete den verfügbaren Algorithmen enorme Möglichkeiten, zu lernen und sich zu verbessern.

###### NLP und der Turing-Test

Einer der frühen Pioniere der KI war der Mathematiker und Informatiker Alan Mathison Turing. Mit seinen Forschungen schuf er die theoretische Grundlage für den späteren Turing-Test (Turing, 1950). Bei diesem Test befragt eine menschliche Testperson über einen Chat zwei Chatpartner: einen anderen Menschen und einen Chatbot. Beide versuchen, die Testperson davon zu überzeugen, dass sie menschlich sind. Wenn die Testperson nicht erkennen kann, welcher ihrer Gesprächspartner ein Mensch und welcher eine Maschine ist, ist der Test erfolgreich bestanden. Laut Turing erlaubt das Bestehen des Tests die Annahme, dass die intellektuellen Fähigkeiten eines Computers auf dem gleichen Niveau liegen wie die des menschlichen Gehirns.

Der Turing-Test befasst sich in erster Linie mit den Fähigkeiten einer Maschine, natürliche Sprache zu verarbeiten. Daher wurde der Turing-Test oft kritisiert, weil er sich zu sehr auf die Funktionalität und nicht auf das Bewusstsein konzentriert. Ein früher Versuch, den Turing-Test zu bestehen, wurde von Joseph Weizenbaum unternommen, der ein Computerprogramm zur Simulation eines Gesprächs mit einem Psychotherapeuten entwickelte (Weizenbaum, 1966). Sein Computerprogramm ELIZA war eine der ersten dialogfähigen KIs. Um den vom Benutzer eingegebenen Satz zu verarbeiten, verwendet ELIZA einen regelbasierten Mustervergleich in Kombination mit einem Wortschatzspeicher. Die Veröffentlichung zog einige bemerkenswerte Reaktionen aus der Forschungsgemeinde nach sich. Die Einfachheit dieses Ansatzes wurde jedoch bald erkannt und entsprechend den Erwartungen der Forschungsgemeinde hat ELIZA den Turing-Test nicht bestanden.

Markow-Modelle

In einem Markow-Modell wird der nächste Zustand auf der Grundlage des aktuellen Zustands und einer Reihe von Übergangswahrscheinlichkeiten definiert.

Im Jahr 2014 war der Chatbot „Eugene Goostman“ die allererste KI, die den Turing-Test bestanden zu haben schien. Der Chatbot gab vor, ein 13-jähriger Junge aus der Ukraine zu sein, der kein englischer Muttersprachler war. Dieser Trick wurde verwendet, um zu erklären, dass der Bot nicht alles weiß und manchmal sprachliche Fehler macht. Dieser Trick war jedoch auch der Grund, warum die Gültigkeit des Experiments später in Frage gestellt wurde (Masnick, 2014).

###### Anwendungsbereiche von NLP

Wir werden nun kurz die wichtigsten Anwendungsbereiche von NLP beschreiben.

Themenerkennung

Wie der Name schon sagt, beschäftigt sich die Themenerkennung mit der Aufgabe, die Themen eines gegebenen Textes automatisch zu erkennen (May *et al*., 2015). Dies kann entweder auf überwachte oder auf unüberwachte Weise geschehen. Bei der überwachten Themenerkennung kann ein Modell beispielsweise mit Zeitungsartikeln trainiert werden, die mit Themen wie Politik, Sport oder Kultur gekennzeichnet wurden. In einer unüberwachten Umgebung sind die Themen dagegen nicht im Voraus bekannt. In diesem Fall muss sich der Algorithmus mit Themenmodellierung oder Themenfindung beschäftigen, um Cluster mit ähnlichen Themen zu finden.

Beliebte Anwendungsfälle für die Themenerkennung sind z. B. die Überwachung von sozialen Medien und Marken, der Kundensupport und die Marktforschung. Die Themenerkennung kann dabei helfen, herauszufinden, was die Leute über eine Marke oder ein Produkt denken. Soziale Medien liefern eine enorme Menge an Textdaten, die für diese Anwendungsfälle analysiert werden können. Die Kunden können nach ihren Interessen gruppiert werden, und die Reaktionen auf bestimmte Anzeigen oder Marketingkampagnen können mühelos analysiert werden. In der Marktforschung kann die Themenerkennung bei der Analyse offener Antworten in Fragebögen hilfreich sein. Wenn diese Antworten vorab klassifiziert werden, kann dies den Aufwand für die Analyse offener Antworten verringern.

Auch beim Kundensupport kann die Themenerkennung von Vorteil sein, indem die Anfragen der Kunden nach Themen kategorisiert werden. Die automatische Weiterleitung von Anfragen an spezialisierte Angestellte kann nicht nur die Kosten senken, sondern auch die Kundenzufriedenheit erhöhen.

Textzusammenfassung

Die Textzusammenfassung befasst sich mit Methoden zur automatischen Erstellung von Zusammenfassungen eines gegebenen Textes, die die wichtigsten Informationen dieses Textes enthalten. Algorithmen zur Textzusammenfassung basieren auf extraktiven und abstraktiven Techniken. Extraktive Algorithmen erstellen eine Zusammenfassung eines gegebenen Textes, indem sie die wichtigsten Wortfolgen extrahieren. Abstraktive Techniken hingegen erstellen Zusammenfassungen, indem sie einen neuen Text erstellen und den Inhalt des Originaldokuments umschreiben.

Eine gängige Technik zur Textzusammenfassung, die auf unüberwachte, extraktive Weise arbeitet, ist TextRank (Mihalcea & Tarau, 2004). Dieser Algorithmus vergleicht jeden Satz eines gegebenen Textes mit allen anderen Sätzen. Dazu wird für jedes Satzpaar ein Ähnlichkeitswert berechnet. Liegt der Wert näher bei 1, weist das auf eine größere Ähnlichkeit zwischen einem Satz und einem anderen Satz hin, der den Inhalt gut wiedergibt. Für jeden Satz werden die Werte zusammengefasst, um eine Rangliste der Sätze zu erhalten. Nachdem die Sätze nach ihrem

Natural Language Processing (NLP)

Rang sortiert wurden, ist es einfach, die Wichtigkeit der einzelnen Sätze auszuwerten und eine Zusammenfassung aus einer vordefinierten Zahl mit dem höchsten Rang zu erstellen. Bei der überwachten extraktiven Textzusammenfassung gibt es zwei große Herausforderungen, da für das Training eine große Menge an mit manuellen Anmerkungen versehenen Textdaten benötigt wird. Diese sind:

1. Es ist notwendig, dass die Anmerkungen die Wörter enthalten, die in der Zusammenfassung enthalten sein müssen. Wenn Menschen Texte zusammenfassen, neigen sie dazu, dies auf eine abstrakte Weise zu tun. Daher ist es schwierig, Trainingsdaten im erforderlichen Format zu finden.
2. Die Entscheidung darüber, welche Informationen in die Zusammenfassung aufgenommen werden sollen, ist subjektiv und hängt vom jeweiligen Fokus einer Aufgabe ab. Während sich eine Produktbeschreibung eher auf die technischen Aspekte eines Textes konzentriert, legt eine Zusammenfassung des geschäftlichen Mehrwerts eines Produkts den Schwerpunkt auf ganz andere Aspekte.

Ein typischer Anwendungsfall für die Textzusammenfassung ist die Darstellung einer Vorschau auf den Inhalt von Suchergebnissen oder Artikeln. Dies erleichtert die schnelle Analyse einer großen Informationsmenge. Darüber hinaus können bei der Beantwortung von Fragen Techniken zur Textzusammenfassung verwendet werden, um einem Benutzer zu helfen, Antworten auf bestimmte Fragen in einem Dokument zu finden.

Stimmungsanalyse

Die Stimmungsanalyse erfasst subjektive Aspekte von Texten (Nasukawa & Yi, 2003), wie z. B. die Analyse der Stimmung des Autors eines Tweets auf Twitter. Wie bei der Themenerkennung geht es bei der Stimmungsanalyse um die Klassifizierung von Texten. Der Hauptunterschied zwischen der Themenerkennung und der Textklassifizierung besteht darin, dass sich die Themenerkennung auf objektive Aspekte des Textes konzentriert, während sich die Stimmungsanalyse auf subjektive Merkmale wie Stimmungen und Emotionen konzentriert.

Die Anwendungsbereiche für die Stimmungsanalyse sind vielfältig. Die Analyse der Kundenstimmung (engl. Customer Sentiment Analysis) hat in letzter Zeit als Forschungsgebiet stark an Bedeutung gewonnen. Die Fähigkeit, die Stimmung der Kunden im Zeitverlauf zu verfolgen, kann zum Beispiel wichtige Erkenntnisse darüber liefern, wie Kunden auf Änderungen eines Produkts oder einer Dienstleistung reagieren oder wie externe Faktoren wie globale Krisen die Wahrnehmung der Kunden beeinflussen. Soziale Netzwerke wie Facebook, Instagram und Twitter liefern riesige Mengen an Daten darüber, wie Kunden über ein Produkt denken. Wenn man die Bedürfnisse von Kunden besser versteht, können die eigenen Geschäftsprozesse entsprechend angepasst und optimiert werden. Die Erkennung von Emotionen in nutzergenerierten Inhalten geht mit großen Herausforderungen einher, wenn es um Ironie/Sarkasmus, Negation und Multipolarität geht.

In nutzergenerierten Inhalten gibt es immer wieder Sarkasmus, besonders in den sozialen Medien. Selbst für Menschen kann es manchmal schwierig sein, Sarkasmus zu erkennen, für eine Maschine ist es damit nur noch schwieriger. Betrachten wir zum Beispiel folgenden Satz:

„Wow, euer Handy hat einen internen Speicher von 1 Gigabyte?“

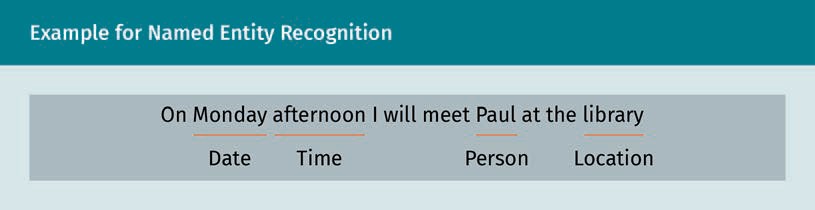
Noch vor ein paar Jahren wäre dies ein völlig klarer Satz gewesen. Wenn dies über ein modernes Smartphone gesagt wird, ist es für einen Menschen leicht zu erkennen, dass diese Aussage sarkastisch ist. Auch wenn es in jüngster Zeit einige Erfolge bei der Erkennung von Sarkasmus mit Deep-Learning-Methoden gegeben hat (Ghosh & Veale, 2016), bleibt der Umgang mit Sarkasmus weiter eine anspruchsvolle Herausforderung.

Die Negation – also die Verneinung – ist eine weitere Herausforderung, wenn wir versuchen, die Stimmung einer Aussage zu erkennen. Die Negation kann explizit oder implizit sein und ist auch mit der Morphologie eines Wortes verbunden, die durch Präfixe wie „de-“ und „ent-“ oder Suffixe wie „-los“ gekennzeichnet ist. Die doppelte Negation ist ein weiteres Sprachkonstrukt, das leicht missverstanden werden kann. In den meisten Fällen heben sich doppelte Verneinungen gegenseitig auf, in manchen Kontexten können sie die Verneinung aber auch bestärken. Die Berücksichtigung der Negation in dem für die Stimmungsanalyse verwendeten Modell kann dazu beitragen, die Genauigkeit deutlich zu erhöhen (Sharif *et al*., 2016).

Eine zusätzliche Herausforderung bei der Stimmungsanalyse kann die Multipolarität sein; das bedeutet, einige Teile des Textes sind positiv und andere negativ. Der Satz „Das Display meines neuen Handys ist großartig, aber die Audioqualität ist wirklich schlecht“ enthält eine positive Stimmung hinsichtlich des Displays, während sie für die Lautsprecher negativ ist. Einfach den Durchschnitt der Stimmungen zu berechnen, kann zu Informationsverlusten führen. Daher wäre es besser, den Satz in zwei Teile aufzuteilen: einen für die positive Bewertung des Displays und einen für das negative Feedback über die Lautsprecher.

Named Entity Recognition

Named Entity Recognition (NER, dt. Eigennamenerkennung) befasst sich mit der Aufgabe, benannte Entitäten in einem unstrukturierten Text zu finden und zu klassifizieren. Diese Entitäten können dann Kategorien wie Namen, Orten, Zeit- und Datumsausdrücken, Organisationen, Mengen und vielen anderen zugeordnet werden. NER spielt eine wichtige Rolle beim Verständnis des Inhalts eines Textes. Insbesondere für die Textanalyse und die Datenorganisation ist die NER ein guter Ausgangspunkt für weitere Analysen. Die folgende Abbildung zeigt ein Beispiel dafür, wie Entitäten anhand eines Satzes identifiziert werden können.



NER kann in allen Domänen eingesetzt werden, in denen die Kategorisierung von Text von Vorteil sein kann. So können beispielsweise Tickets im Kundensupport nach Themen kategorisiert werden. Tickets können dann automatisch an einen Spezialisten weitergeleitet werden. Auch wenn Daten aufgrund von Datenschutzbestimmungen anonymisiert werden müssen, kann NER helfen, Kosten zu sparen. Sie kann personenbezogene Daten identifizieren und automatisch entfernen. Abhängig von der Qualität der zugrunde liegenden Daten ist eine manuelle Bereinigung nicht mehr notwendig. Ein weiterer Anwendungsfall ist die Extraktion von Informationen aus Lebensläufen von Bewerbern im Bewerbungsprozess. Das kann die Arbeitsbelastung der Personalabteilung erheblich verringern, insbesondere wenn es viele Bewerber gibt (Zimmermann *et al*., 2016).

Die größte Herausforderung bei der NER besteht darin, dass zum Trainieren eines Modells eine große Menge an annotierten Daten zur Verfügung stehen muss. Das Modell wird sich später immer auf die spezifischen Aufgaben bzw. die spezifische Teilmenge von Entitäten konzentrieren, für die es trainiert wurde.

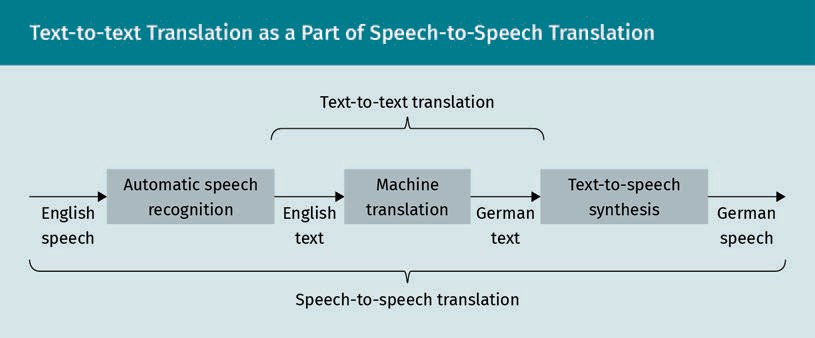
Natural Language Processing (NLP)

Übersetzung

Die maschinelle Übersetzung (MÜ; auch engl. Machine Translation, MT) ist ein Teilbereich des NLP, der mehrere Disziplinen vereint. Mithilfe von Methoden aus der Künstlichen Intelligenz, der Informatik, der Informationstheorie und der Statistik werden bei der die MÜ Texte oder gesprochene Sprache automatisch von einer Sprache in eine andere übersetzt.

In den letzten Jahrzehnten hat sich die Qualität der MÜ deutlich verbessert. In den meisten Fällen ist die Qualität der maschinellen Übersetzungen allerdings immer noch nicht so hoch wie die von professionellen Übersetzern. Die Kombination von MÜ und manueller Nachbearbeitung ist heutzutage jedoch oft schneller als die rein manuelle Übersetzung. Wie in jedem anderen Bereich des NLP hängt die Qualität der Ausgabe maßgeblich von der Qualität der Trainingsdaten ab. Daher werden oft domänenspezifische Daten verwendet. Während in der Vergangenheit die **statistische maschinelle Übersetzung** (SMT) die am häufigsten verwendete Methode war, ist die **neuronale maschinelle Übersetzung** (NMT) heute vorherrschend (Koehn & Knowles, 2017).

MÜ kann sowohl für Text-zu-Text-Übersetzungen als auch für Sprache-zu-Sprache-Übersetzungen verwendet werden. MÜ für Texte kann dazu beitragen, Textdokumente oder Websites schneller zu übersetzen, professionelle Übersetzer beim Beschleunigen des Übersetzungsprozesses zu unterstützen oder kann als Teil eines Sprache-zu-Sprache-Übersetzungssystems dienen. Mit der fortschreitenden Globalisierung gewinnt die MÜ von Tag zu Tag an Bedeutung. Im Jahr 2016 übersetzte Google über 100 Milliarden Wörter pro Tag in mehr als 100 Sprachen. (Turovsky, 2016) Die folgende Abbildung zeigt, wie die Text-zu-Text-Übersetzung mit der Sprache-zu-Sprache-Übersetzung verknüpft ist.



Text-zu-Text-Übersetzung und Sprache-zu-Sprache-Übersetzung werden immer wichtiger. Dieser Prozess wurde durch die enorme Zunahme von Videochats und -konferenzen in den letzten Jahren noch beschleunigt. Sprachübersetzung kann helfen, Sprachbarrieren zu überbrücken, indem Sie Anwendungen wie den Skype-Übersetzer verwenden.

Wie die Abbildung zeigt, ist das Herzstück eines Sprache-zu-Sprache-Übersetzungssystems ein Text-zu-Text-Übersetzungssystem. Bevor die Übersetzung beginnt, muss die Sprache mit Methoden der automatischen Spracherkennung (ASR) in Text umgewandelt werden. Nach der Übersetzung wird die Text-to-Speech-Synthese (TTS) verwendet, um aus dem übersetzten Text gesprochene Sprache zu erzeugen. Daher hängt die Qualität der Ausgabe nicht nur von der Qualität der MÜ-Komponente ab, sondern auch von der Qualität der ASR- und TTS-Komponenten, was die Sprache-zu-Sprache-Übersetzung zu einer echten Herausforderung macht.

Statistische maschinelle Übersetzung

Bei der statistischen maschinellen Übersetzung werden Übersetzungen mithilfe von statistischen Modellen erstellt, die auf der Analyse von zweisprachigen Textkorpora basieren.

Neuronale maschinelle Übersetzung

Bei der neuronalen maschinellen Übersetzung wird ein künstliches neuronales Netzwerk verwendet, um ein statistisches Modell für die MÜ zu lernen.

Die beiden größten Herausforderungen in der MÜ sind heutzutage die Nichtübereinstimmung der Domäne und die unterversorgten Sprachen. Nichtübereinstimmung der Domäne bedeutet, dass Wörter und Sätze je nach Domäne unterschiedliche Übersetzungen haben können. Daher ist es wichtig, die Domäne anzupassen, wenn Sie ein MÜ-System für einen speziellen Anwendungsfall entwickeln.

Für einige Sprachkombinationen in der MÜ gibt es keine zweisprachigen Textkorpora für Ausgangs- und Zielsprache. Ein Ansatz zur Lösung des Problems der unterversorgten Sprachen ist die Verwendung von Pivot-MT. Bei der Pivot-MT werden die Ausgangs- und die Zielsprache durch eine dritte Sprache überbrückt (Kim *et al*., 2019). Wenn Sie zum Beispiel von Khmer (Kambodscha) nach Zulu (Südafrika) übersetzen, wird ein Text zunächst von Khmer nach Englisch und anschließend von Englisch nach Zulu übersetzt.

Chatbots

Chatbots sind textbasierte Dialogsysteme. Sie ermöglichen die Interaktion mit einem Computer auf der Grundlage von Text in natürlicher Sprache. Basierend auf den Eingaben antwortet das System in natürlicher Sprache. Manchmal werden Chatbots in Kombination mit einem Avatar verwendet, der einen Charakter oder eine Persönlichkeit simuliert. Einer der beliebtesten Chatbots war ELIZA, der einen Psychotherapeuten imitierte. Chatbots werden häufig in Messenger-Apps, wie z. B. Facebook, oder in Website-Chats eingesetzt. Außerdem bilden sie die Grundlage für digitale Assistenten wie Alexa, Siri oder Google Assistant.

Chatbots können nach ihrem Intelligenzgrad kategorisiert werden:

* Benachrichtigungsassistenten (Stufe 1): Diese Chatbots interagieren nur unidirektional mit dem Benutzer. Sie können für Benachrichtigungen über Ereignisse oder Aktualisierungen (d. h. Push-Benachrichtigungen) verwendet werden.
* Assistenten für häufig gestellte Fragen (Ebene 2): Diese Bots können bidirektional mit einem Benutzer interagieren. Sie können die Anfrage des Benutzers interpretieren und eine passende Antwort in einer Wissensbasis finden.
* Kontextabhängige Assistenten (Stufe 3): Diese Chatbots können nicht nur bidirektional interagieren, sondern sind auch kontextabhängig und basieren auf dem Gesprächsverlauf.

Es ist wahrscheinlich, dass sich in Zukunft weitere Stufen von Chatbots entwickeln werden. Ein Chatbot besteht aus drei Komponenten:

1. Natürliches Sprachverständnis (engl. Natural Language Understanding, NLU): Diese Komponente analysiert den eingegebenen Text und identifiziert die Absicht und die Entitäten des Benutzers (Benutzerinformationen).
2. Dialogverwaltungskomponente: Das Ziel dieser Komponente ist es, die vom NLU identifizierten Absichten und Entitäten im Kontext der Konversation zu interpretieren und die Reaktion des Bots zu bestimmen.
3. Nachrichtengeneratorkomponente: Basierend auf der Ausgabe der anderen Komponenten besteht die Aufgabe dieser Komponente darin, die Antwort des Chatbots zu generieren, indem sie entweder eine vordefinierte Vorlage ausfüllt oder einen freien Text erzeugt.

Chatbots können eine Menge Zeit und Geld sparen. Die Zahl der Anwendungsfälle nimmt daher ständig zu. Sie sind normalerweise rund um die Uhr zu vergleichsweise geringen Kosten verfügbar und können bei Bedarf leicht skaliert werden. Im Kundendienst können sie nicht nur die Anfragen der Kunden beantworten,

Natural Language Processing (NLP)

sondern auch Produktempfehlungen geben oder Reisearrangements wie Hotel- oder Flugreservierungen vornehmen. Wenn eine Anfrage zu kompliziert für einen Bot ist, gibt es in der Regel Schnittstellen, um Anfragen an ein menschliches Support-Team weiterzuleiten.

Im Marketing können Chatbots zur Lead-Generierung eingesetzt werden. Immer mehr Unternehmenswebsites verwenden bereits Bots, um ansonsten anonymen Besuchern verkaufsorientierte Fragen zu stellen. Die Antworten können dazu verwendet werden, neue Leads zu generieren. Außerdem kann der Chatbot den Benutzer beraten oder ihm neue Produkte oder Informationen über Sonderangebote vorstellen. Der Einsatz von Chatbots als persönliche Verkaufsagenten kann auch dazu beitragen, den Arbeitsaufwand von Menschen zu reduzieren, so dass diese sich auf komplexere Aufgaben konzentrieren können.

### Grundlegende NLP-Techniken

Frühe Systeme für NLP verwendeten Regeln, die auf sprachlichen Strukturen basierten. Diese Regeln wurden oft von Hand aufgeschrieben und galten nur für die Domäne, für die sie entworfen wurden. Heutzutage verwenden die meisten NLP-Systeme statistische Methoden des maschinellen Lernens.

###### Regelbasierte Techniken

Regelbasierte Techniken für NLP verwenden eine Reihe von vordefinierten Regeln, um ein bestimmtes Problem anzugehen. Diese Regeln versuchen, die Art und Weise zu reproduzieren, wie Menschen Sätze bilden. Ein einfaches Beispiel für ein regelbasiertes System ist die Extraktion von einzelnen Wörtern aus einem Text auf der Grundlage der sehr einfachen Regel „Teile den Text an jedem Leerzeichen“. Ein Blick auf Begriffe wie „New York“ zeigt bereits, wie schnell ein scheinbar einfaches Problem kompliziert werden kann. Daher basieren komplexere Systeme auf sprachlichen Strukturen unter Verwendung formaler Grammatiken.

Der regelbasierte Ansatz impliziert, dass Menschen in den Prozess einbezogen werden müssen, um ein System aufzubauen. Einer der größten Vorteile von regelbasierten Systemen ist daher die Erklärbarkeit: Da die Regeln von Menschen entworfen wurden, ist es einfach zu verstehen, wie eine Aufgabe bearbeitet wurde und Fehler zu finden.

Regelbasierte Systeme können flexibel entwickelt werden. Normalerweise ist es nicht notwendig, Änderungen am Kern einer Anwendung vorzunehmen, wenn Regeln geändert werden, sei es durch das Hinzufügen neuer Regeln oder die Korrektur bestehender Regeln. Ein weiterer Vorteil von regelbasierten Systemen ist, dass die Menge an Trainingsdaten, die für die Entwicklung des Systems erforderlich ist, vergleichsweise gering ist.

Der größte Nachteil des regelbasierten Ansatzes besteht darin, dass Experten benötigt werden, um geeignete Regeln zu erstellen. Darüber hinaus sind regelbasierte Systeme in der Regel domänenspezifisch aufgebaut, was es schwierig macht, ein System in einer Domäne einzusetzen, für die es nicht konzipiert wurde.

###### Statistikbasierte Techniken

Da die Rechenleistung in den letzten Jahrzehnten stetig zugenommen hat, haben Systeme, die auf statistischen Methoden basieren – die oft unter dem Begriff Machine Learning (Maschinelles Lernen) zusammengefasst werden – die meisten der frühen regelbasierten Systeme ersetzt. Diese Methoden folgen einem datenorientierten Ansatz. Die mit statistischen Methoden erstellten Modelle werden mit einer riesigen Menge an Trainingsdaten trainiert, um die Regeln für eine bestimmte Aufgabe abzuleiten. Danach können die Modelle verwendet werden, um einen Satz unbekannter Daten zu klassifizieren und Vorhersagen zu treffen. Im Gegensatz zu regelbasierten Systemen benötigen statistikbasierte Systeme kein Expertenwissen über die Domäne. Sie können leicht auf der Grundlage bestehender Methoden entwickelt und durch Bereitstellung geeigneter Daten verbessert werden. Auch die Übertragung des Modells auf eine andere Domäne ist erheblich einfacher als bei regelbasierten Systemen.

Ein Nachteil von Systemen, die auf statistischen Techniken des maschinellen Lernens beruhen, ist jedoch, dass viele annotierte Trainingsdaten erforderlich sind, um gute Ergebnisse zu erzielen, wohingegen regelbasierte Systeme gute Leistungen erbringen können, wenn nur wenige Daten für eine bestimmte Aufgabe zur Verfügung stehen.

###### Aufgaben

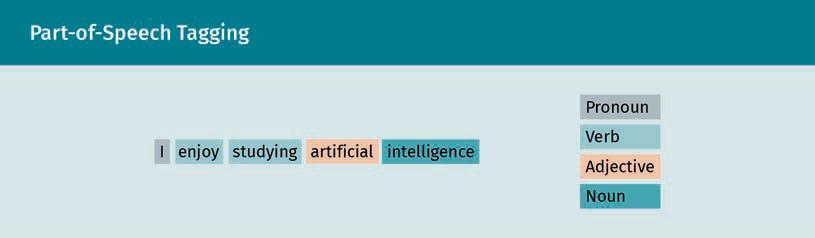
Im Allgemeinen lassen sich NLP-Aufgaben in vier Kategorien unterteilen: Syntax, Semantik, Diskurs und gesprochene Sprache. Im Folgenden geben wir einen Überblick über diese Aufgaben.

Syntax

Syntaktische Aufgaben im NLP befassen sich mit den Merkmalen der Sprache wie Kategorien, Wortgrenzen und grammatikalischen Funktionen. Typische Aufgaben, die sich mit der Syntax befassen, sind die Tokenisierung und das Part-of-Speech-(POS)-Tagging.

Das Ziel der Tokenisierung ist es, einen Text in einzelne Einheiten wie Wörter, Sätze oder Unterworte zu zerlegen. Der Satz „Es macht mir Spaß, Künstliche Intelligenz zu studieren.“ könnte zum Beispiel in „Es“ „macht“ „mir“ „Spaß“ „,“ „Künstliche“ „Intelligenz“ „zu“ „studieren“ „.“ tokenisiert werden.

POS-Tagging – auch grammatikalisches Tagging genannt – geht noch einen Schritt weiter und fügt dem Text grammatikalische Wortfunktionen und Kategorien hinzu. Das folgende Beispiel veranschaulicht, wie ein Satz mithilfe von POS-Tagging analysiert werden kann.



Natural Language Processing (NLP)

Syntaktische Mehrdeutigkeit, d. h. Wörter, die sich nicht eindeutig einer Kategorie zuordnen lassen, sind eine große Herausforderung im NLP.

Ein häufig verwendetes Beispiel für syntaktische Zweideutigkeit ist der englische Satz

„Time flies like an arrow“ („Die Zeit fliegt wie ein Pfeil“),

der auf viele verschiedene Arten interpretiert werden kann. Zwei der möglichen Interpretationen sind

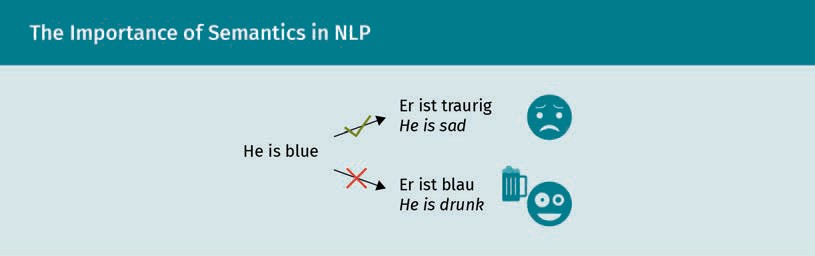
1. Die Zeit vergeht so schnell wie ein Pfeil fliegt
2. Es gibt einen bestimmten Pfeil, den jede „Zeitfliege“ („time fly“, Insekt) mag

In der ersten Interpretation ist „like“ („wie“) eine vergleichende Präposition, während es in der zweiten Interpretation ein Verb („mögen“) ist.

Semantik

Der Schwerpunkt der semantischen Aufgaben liegt auf der Bedeutung von Wörtern und Sätzen. Das Verständnis der Bedeutung eines Textes ist für die meisten Anwendungsbereiche von NLP unerlässlich. Bei der Stimmungsanalyse werden die subjektiven Aspekte der Sprache analysiert. Bei der Analyse von Beiträgen in sozialen Medien ist es zum Beispiel wichtig zu verstehen, was der Text bedeutet, um zu erkennen, ob er positiv oder negativ ist. Die Erkennung von benannten Entitäten (engl. Named Entity Recognition, NER) ist ein weiteres Forschungsgebiet, in dem die Semantik für korrekte Klassifizierungsergebnisse wichtig ist. Die Identifizierung von Entitäten, wie Namen, Orten oder Daten, aus einem gegebenen Text ist nicht möglich, ohne die Semantik eines Textes zu verstehen. Bei der Themenerkennung wird ein bestimmter Text mit einem Thema versehen. Deshalb ist es wichtig zu verstehen, worum es in dem Text geht. Zeitungsartikel könnten zum Beispiel mit Themen wie „Politik“, „Kultur“ oder „Wetter“ gekennzeichnet werden.

Wenn NLP zur Beantwortung von Fragen verwendet wird, muss ein Computer eine passende Antwort auf eine bestimmte Frage erstellen. Angenommen, ein Algorithmus zur Beantwortung von Fragen wurde mit diesem Studienskript trainiert, dann könnte der Algorithmus diesen Abschnitt anzeigen, wenn er gefragt wird: „Was sind die typischen Aufgaben im NLP?“ Zu diesem Zweck muss die Semantik dieses Abschnitts richtig interpretiert werden. Auch bei der maschinellen Übersetzung ist es wichtig, die richtige Bedeutung eines Textes zu verstehen. Andernfalls wird die Übersetzung schwer verständliche oder sogar falsche Ergebnisse liefern.



Die obige Abbildung zeigt, wie wichtig es ist, die Semantik eines Textes richtig zu verstehen.

Diskurs

Ein Diskurs befasst sich mit Texten, die länger als ein einziger Satz sind. Dies ist wichtig für Aufgaben wie die Themenerkennung und die Textzusammenfassung, bei denen ein Algorithmus eine Zusammenfassung eines gegebenen Textes erstellt, indem er die wichtigsten Sätze extrahiert. Die Analyse des Verlaufs eines Textes umfasst mehrere Teilaufgaben, wie die Identifizierung der Themenstruktur, die Analyse der Koreferenz und der Gesprächsstruktur.

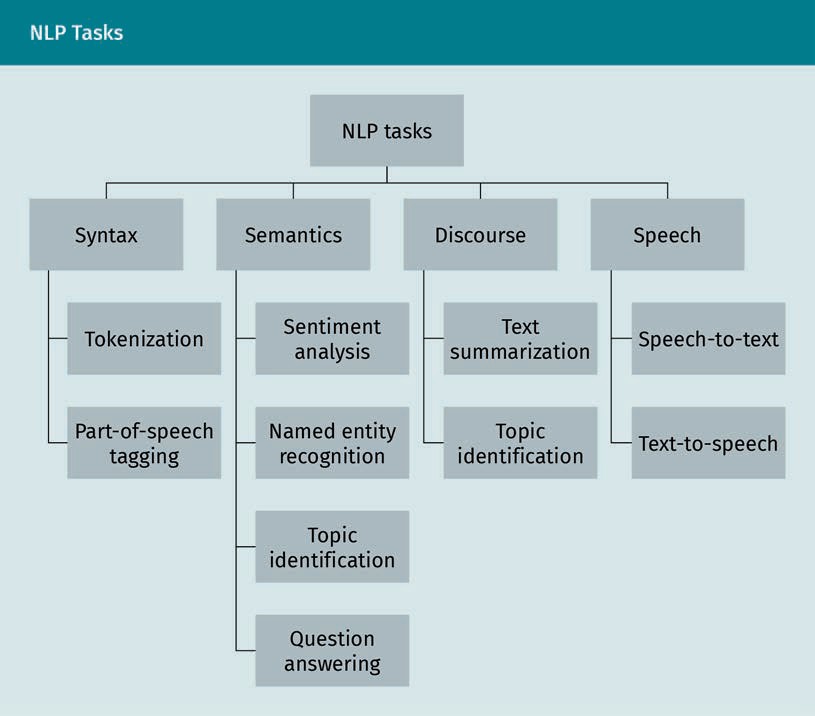
Gesprochene Sprache

Bei Sprachaufgaben im NLP dreht sich alles um gesprochene Sprache. Bei Sprachaufgaben können zwei Teilaufgaben unterschieden werden:

* 1. Speech-to-Text (STT): Sie wird auch als automatische Spracherkennung (ASR) bezeichnet und wandelt gesprochene Sprache in Text um.
  2. Text-to-Speech (TTS), oder Sprachsynthese, die sich mit der Umwandlung eines geschriebenen Textes in gesprochene Sprache beschäftigt.

Beides ist wichtig für Konversationsschnittstellen, wie z. B. Sprachassistenzsysteme wie Siri oder Alexa. Die folgende Abbildung fasst die typischen Aufgaben im NLP zusammen.

Natural Language Processing (NLP)



### Daten vektorisieren

Im Machine Learning akzeptieren die Algorithmen nur numerische Eingaben. Wenn wir also Informationen aus einem unstrukturierten Text extrahieren wollen, müssen wir einen Weg finden, wie der Computer sie verarbeiten kann. Zu diesem Zweck muss der Text in ein numerisches Format umgewandelt werden, das der Computer verarbeiten kann.

Im Folgenden wollen wir zwei Ansätze vorstellen, wie Wörter in einen semantischen Vektorraum eingebettet werden können: den Bag-of-Words-Ansatz, der einfach ist, und das leistungsfähigere Konzept der neuronalen Wort- und Satzvektoren.

###### Bag-of-Words

Einer der einfachsten Ansätze, Textinformationen in Zahlen umzuwandeln, ist das Bag-of-Words-(BoW)-Modell. Beim BoW wird ein Text durch einen Vektor dargestellt, der die Anzahl der Wortvorkommen in einem bestimmten Textdokument beschreibt. Der Begriff „Bag“, also „Tasche“, bezieht sich auf die Tatsache,

dass alle Informationen über die Struktur oder die Reihenfolge der Wörter in einem Text verworfen werden, sobald die Wörter in die eindeutige Menge von Wörtern, die einen Text beschreiben, eingeordnet sind. Um den BoW-Ansatz zu verstehen, werden wir den folgenden Beispieltext verwenden:

* + - Darren liebt Hunde
    - Katzen mag Darren nicht
    - Katzen sind nicht wie Hunde

Im ersten Schritt müssen wir alle eindeutigen Wörter des Textes identifizieren. Zu diesem Zweck verwenden wir die Tokenisierung.

In dem obigen Text werden die folgenden Worte verwendet:

Darren, liebt, Hunde, Katzen, mag, nicht, sind, wie

Im nächsten Schritt müssen wir die Wörter in jedem Satz bewerten. Da wir wissen, dass unser Vokabular aus 8 Wörtern besteht, wird der resultierende Vektor eine Länge von 8 haben. Die BoW-Vektoren für die obigen Sätze sehen wie folgt aus:

* + - [1, 1, 1, 0, 0, 0, 0, 0]
    - [1, 0, 0, 1, 1, 1, 0, 0]
    - [0, 0, 1, 1, 0, 1, 1, 1]

Es gibt verschiedene Methoden, um die Wörter im BoW-Modell zu bewerten. In den obigen Sätzen kommt jedes Wort nur einmal vor, daher sind die resultierenden Vektoren eine binäre Darstellung des Textes. Wenn der gesamte Text aus dem obigen Beispiel in einem Vektor zusammengefasst wurde, stehen die folgenden Optionen zur Verfügung:

* + - Boolesche Darstellung: der Vektor gibt einfach an, ob ein Wort vorkommt oder nicht [1, 1, 1, 1, 1, 1, 1, 1].
    - Anzahl der Wörter: der resultierende Vektor gibt an, wie oft ein Wort vorkommt [2,1,2,2,1,2,1,1].

Wie Sie feststellen werden, enthält diese Darstellung keine Informationen mehr über die ursprüngliche Reihenfolge der Wörter.

Beschränkungen von Bag-of-Words

Alles in allem ist das BoW-Modell einfach, was einige große Nachteile mit sich bringt:

Natural Language Processing (NLP)

* Auswahl der Vokabeln: Das Vokabular des Modells muss mit großer Sorgfalt ausgewählt werden. Das Gleichgewicht zwischen der Größe des Modells und der Besetzung muss immer im Auge behalten werden. Je größer das Vokabular ist, desto dünner besetzt sind die Vektoren.
* Risiko dünner Besetzung: Aus computertechnischen Gründen ist es schwieriger, eine dünnbesetzte Repräsentation von Daten zu modellieren, da die Komplexität von Zeit und Raum mit zunehmend dünner Besetzung steigt. Außerdem ist es schwieriger, die Daten zu nutzen, wenn nur wenige Informationen in einem großen Darstellungsraum enthalten sind.
* Verlust von Bedeutung: Bei der Verwendung von BoW werden weder die Wortstellung noch der Kontext oder der Sinn berücksichtigt. In unserem Beispiel gehen die unterschiedlichen Bedeutungen von „like“ (einmal als Präposition und einmal als Verb) völlig verloren. In solchen Situationen schneidet das BoW-Modell nicht gut ab.

###### Wortvektoren

Um Wörter in einen semantischen Vektorraum einbetten zu können, können sie als Wortvektoren dargestellt werden. Lineare Operationen können angewendet werden, um Wortanalogien und Ähnlichkeiten zu finden. Diese Wortähnlichkeiten können zum Beispiel auf der Kosinusähnlichkeit basieren. Vor allem aber können Wörter, sobald sie in Wortvektoren umgewandelt wurden, als Eingabe für Machine-Learning-Modelle wie künstliche neuronale Netze und lineare Klassifikatoren verwendet werden. Im Folgenden werden drei Vektorisierungsmethoden vorgestellt: Word2Vec, TD-IDF und GloVe.

Word2Vec

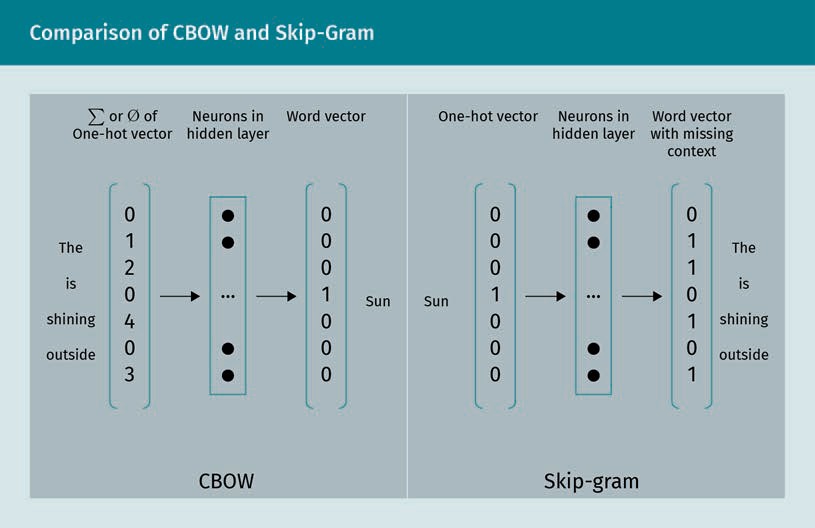
Das Word2Vec-Modell basiert auf einem einfachen neuronalen Netz. Das neuronale Netz generiert Worteinbettungen auf der Grundlage von nur einer verborgenen Schicht (Mikolov *et al*., 2013). Als Google Research das Modell im Jahr 2013 veröffentlichte, war damit ein Meilenstein in der Forschung erreicht. Die Eingabeschicht des neuronalen Netzes erwartet einen „One-Hot-Vektor“. Der One-Hot-Vektor ist ein BoW-Vektor für ein einzelnes Wort. Das bedeutet, dass alle Indizes dieses Vektors auf 0 gesetzt werden, außer dem Index des Wortes, das analysiert wird. Dieser Index wird auf 1 gesetzt.

Das Training des neuronalen Netzes für Word2Vec erfordert einen großen Textkorpus. Dies könnte zum Beispiel ein Wikipedia-Speicherauszug sein. Beim Training wird ein Wortfenster fester Länge mit der Länge N über den Korpus geschoben. Typische Werte für N wären z. B. N = 5 oder N = 10.

In Word2Vec gibt es zwei Vorhersagemodelle:

1. Continuous Bag-of-Words (CBOW): Dieses Modell kann verwendet werden, wenn das Ziel darin besteht, ein fehlendes Wort in einem festen Fenster im Kontext der anderen N – 1 Wörter vorherzusagen. Als Eingabevektor können wir entweder den Durchschnitt oder die Summe des One-Hot-Vektors verwenden.
2. Skip-Gram: Wenn wir ein Word innerhalb eines festen Fensters haben, können wir mit diesem Modell die verbleibenden N – 1 Kontextwörter vorhersehen.

Der Unterschied zwischen den beiden Modellen ist in der folgenden Abbildung dargestellt.



Ein wichtiger Aspekt von CBOW ist, dass das Ergebnis der Vorhersage nicht von der Reihenfolge der Kontextwörter beeinflusst wird. Beim Skip-Gram werden nah zusammenliegende Kontextwörter stärker gewichtet als weiter entfernte Kontextwörter. Während CBOW im Allgemeinen schneller arbeitet, ist die Skip-Gram-Architektur besser für selten auftretende Wörter geeignet.

Beim Training von Word2Vec besteht das Ziel darin, die Wahrscheinlichkeiten für die Wörter zu maximieren, die in dem festen Fenster der analysierten Stichprobe aus dem für das Training verwendeten Datenkorpus erschienen sind. Die Funktion, die wir durch diesen Prozess erhalten, ist die Zielfunktion.

Bei einer NLP-Aufgabe besteht das Ziel in der Regel nicht darin, ein Modell zur Vorhersage des nächsten Wortes auf der Grundlage eines gegebenen Textausschnitts zu finden, sondern die Syntax und Semantik eines gegebenen Wortes oder Textes zu analysieren. Wenn wir die Ausgabeschicht des Modells, das wir zuvor erstellt haben, entfernen und uns stattdessen die verborgene Schicht ansehen, können wir den Ausgabevektor aus dieser Schicht extrahieren. Neuronale Netze entwickeln in der Regel eine starke Abstraktions- und Generalisierungsfähigkeit auf ihren letzten Schichten. Es ist daher möglich, den Ausgabevektor der versteckten Schicht als abstrakte Darstellung der Merkmale aus den Eingabedaten zu verwenden. Wir können ihn also als Einbettungsvektor für das Wort verwenden, das wir analysieren möchten.

Heute gibt es viele vortrainierte Word2Vec-Modelle für verschiedene Sprachen, die leicht für spezifische NLP-Aufgaben angepasst werden können.

Term Frequency-Inverse Document Frequency

Bei BoW spiegelt die Häufigkeit der Vokabeln nur die Wörter wider, die im Dokument enthalten sind. Daher wird bei der Analyse eines Textes allen Wörtern das gleiche Gewicht beigemessen, unabhängig von ihrer Bedeutung. Term Frequency-Inverse Document Frequency (TF-IDF) ist ein statistisches Maß aus dem Information Retrieval, das sich mit diesem Problem befasst und eines der am häufigsten verwendeten Gewichtungsschemata im Information Retrieval (Beel *et al*., 2016). In TF-IDF wird die Begriffshäufigkeit (TF) mit der inversen Dokumenthäufigkeit

Natural Language Processing (NLP)

(IDF) kombiniert. Die Relevanz eines Wortes steigt mit seiner Häufigkeit in einem bestimmten Text, wird aber durch die Worthäufigkeit im gesamten Datensatz bestimmt. Für die Berechnung von TF-IDF benötigen wir die folgenden Parameter:

* Die Vorkommenshäufigkeit (TF) gibt an, wie oft ein Begriff t in einem Dokument d vorkommt. Die Wortreihenfolge ist in diesem Fall nicht relevant. Die Anzahl der Vorkommen wird nach der Gesamtzahl der Begriffe im Dokument gewichtet:

TF t, d = Anzahl der Vorkommen von t in d

Anzahl der Wörter in d

* Die Dokumenthäufigkeit (DF) gibt den Prozentsatz der Dokumente an, die einen bestimmten Begriff t enthalten, im Verhältnis zur Gesamtzahl der Dokumente D. Dies kann als Indikator für die Bedeutung des Textes angesehen werden.

DF t, d, D = Anzahl der Dokumente d, die t enthalten

Gesamtzahl der Dokumente D

* Die inverse Dokumenthäufigkeit (IDF) testet die Relevanz eines bestimmten Begriffs. Wie der Name schon sagt, handelt es sich dabei um die Umkehrung der Dokumenthäufigkeit, die logarithmisch skaliert ist:

IDF t = log 1

DF t, d, D

Der endgültige TF-IDF-Wert für einen Begriff kann wie folgt berechnet werden:

TFIDF t, d = TF t, d · IDF t

Hohe TF-IDF-Werte weisen auf Wörter hin, die häufig in einem Dokument vorkommen, während die Anzahl der Dokumente, die den jeweiligen Begriff enthalten, im Vergleich zur Gesamtmenge der Dokumente gering ist. Daher kann TF-IDF dabei helfen, die wichtigsten Begriffe in einem Dokument zu finden, die in einem Text vorkommen.

GloVe

Globale Vektoren für die Wortdarstellung (GloVe) sind eine weitere Vektorisierungsmethode, die häufig im NLP verwendet wird. Während Word2Vec ein prädiktives Modell ist, ist GloVe ein unüberwachter Ansatz, der auf der Anzahl der Wörter basiert. Sie wurde entwickelt, weil Pennington *et al*. (2014) zu dem Schluss kamen, dass der Skip-Gram-Ansatz in Word2Vec die statistischen Informationen in Bezug auf das gemeinsame Vorkommen (Kookkurenz) von Wörtern nicht vollständig berücksichtigt. Deshalb haben sie den Skip-Gram-Ansatz mit den Vorteilen der **Matrixfaktorisierung** kombiniert. Das GloVe-Modell verwendet eine Kookkurenzmatrix, die Informationen über den Wortkontext enthält. Es hat sich gezeigt, dass das entwickelte Modell verwandte Modelle übertrifft, insbesondere bei der Erkennung von benannten Entitäten und Ähnlichkeitsaufgaben (Pennington *et al*., 2014).

Matrixfaktorisierung Diese Methode wird verwendet, um eine Matrix in ihre Komponenten zu zerlegen und so komplexe Verfahren zu vereinfachen.

###### Satzvektoren

Bis jetzt haben wir gelernt, wie Wörter als Vektoren darstellt werden. Verschiedene NLP-Aufgaben, wie die Beantwortung von Fragen oder die Stimmungsanalyse, erfordern jedoch nicht nur die Analyse eines einzelnen Wortes, sondern eines ganzen Satzes oder Absatzes. Daher brauchen wir auch eine Möglichkeit, eine Wortfolge zu kodieren, um sie mit einem Lernalgorithmus verarbeiten zu können.

Ein Ansatz besteht darin, einen Durchschnitt der Vektoren eines Satzes aus Word2Vec zu bilden und die resultierenden Vektoren als Eingabe für ein Modell zu verwenden. Diese Methode hätte jedoch den Nachteil, dass die Wortreihenfolge nicht mehr in den Wortkodierungen enthalten ist. Zum Beispiel enthalten die Sätze „Ich gebe den Kaffee in die Tasse“ und „Ich gebe die Tasse in den Kaffee“ die gleichen Wörter. Nur die Wortreihenfolge sorgt für den Bedeutungsunterschied.

Um das Problem des Umgangs mit Textausschnitten unterschiedlicher Länge zu bewältigen, gibt es verschiedene Ansätze. In den folgenden Abschnitten werden wir eine Auswahl dieser Algorithmen vorstellen. Bitte beachten Sie, dass der Begriff „Satz“ im Folgenden auch für einen ganzen Textabsatz verwendet wird, nicht nur für einen Satz im strengen grammatikalischen Sinn.

Skip-Thought

Im Ansatz der Skip-Thought-Vektoren (Kiros *et al*., 2015) wird das Konzept der Skip-Gram-Architektur, das wir zuvor im Abschnitt über den Word2Vec-Ansatz vorgestellt haben, auf die Ebene der Sätze übertragen.

Wie Word2Vec benötigt auch Skip-Thought einen großen Textkorpus zum Trainieren des Modells. Im Gegensatz zu Word2Vec wird bei Skip-Thought kein gleitendes Wortfenster verwendet, sondern ein Tripel aus drei aufeinanderfolgenden Sätzen analysiert. Das resultierende Modell ist ein typisches Beispiel für eine Encoder-Decoder-Architektur. Der mittlere Satz des Tripels wird als Eingabe für den Encoder verwendet. Der Encoder erzeugt eine Ausgabe, die mit dem Decoder verbunden ist. Es gibt zwei Möglichkeiten, das Modell zu optimieren: Der Decoder kann entweder den folgenden oder den vorherigen Satz des Satzes vorhersagen, den der Encoder empfangen hat.

Es gibt einige NLP-Aufgaben, die kein Vorhersagemodell erfordern. Für diese Aufgaben wird der Decoderteil nach dem Training nicht mehr benötigt und kann verworfen werden. Um die Vektordarstellung des Satzes zu erhalten, können wir den Ausgabevektor des Encoders verwenden.

Wenn wir das Modell nur zur Vorhersage des folgenden oder des vorhergehenden Satzes verwenden, ist das Ergebnis ein Uni-Skip-Vektor. Wenn wir zwei Uni-Skip-Vektoren so miteinander verknüpfen, dass der eine den vorherigen und der andere den nächsten Satz vorhersagt, wird das Ergebnis als Bi-Skip-Vektor bezeichnet. Wenn n-dimensionale Uni-Skip-Vektoren mit n-dimensionalen Bi-Skip-Vektoren kombiniert werden, ist das Ergebnis ein 2n-dimensionaler Combine-Skip-Vektor. In einem Vergleich verschiedener Kombinationsmodelle hat sich gezeigt, dass das Combine-Skip-Modell etwas besser abschneidet.

Es gibt ein vortrainiertes englisches Sprachmodell, das der Öffentlichkeit zur Verfügung steht und auf dem BookCorpus-Datensatz basiert.

Natural Language Processing (NLP)

Universal Sentence Encoder (USE)

Der Universal Sentence Encoder (USE) ist eine Familie von Modellen für die Satzeinbettung, die von Google Research entwickelt wurde (Cer *et al*., 2018). Es gibt zwei Architekturvarianten des USE. Eine Variante verwendet ein Deep Averaging Network (DAN) (Iyyer *et al*., 2015) und ist schneller, aber weniger genau, während die andere Variante ein Transformermodell verwendet.

Auch für USE gibt es bereits trainierte Modelle, die der Öffentlichkeit zur Verfügung stehen: ein englisches Modell und ein mehrsprachiges Modell (Chidambaram *et al*., 2019). Diese Modelle basieren beide auf der DAN-Architektur.

Bidirectional Encoder Representations from Transformers (BERT)

Wie der Name schon sagt, basiert BERT (Devlin *et al*., 2018) auf der Transformer-Architektur. Wie USE wurde dieses Modell von Google Research eingeführt. Das Sprachmodell ist als Open-Source verfügbar und wurde auf einem großen Textkorpus auf zwei kombinierte und unüberwachte Arten trainiert: maskiertes Sprachmodell und Vorhersage des nächsten Satzes.

Im Modus des maskierten Sprachmodells wird ein Satz aus dem Trainingsdatensatz genommen. Im nächsten Schritt werden etwa 15 Prozent der Wörter in diesem Satz maskiert, also verborgen. So wurden in dem Satz

„Ich [Maske1] morgens gerne eine Tasse Kaffee mit [Maske2]“

die Wörter „trinken“ und „“Milch“ maskiert. Das Modell wird dann darauf trainiert, die fehlenden Wörter im Satz vorherzusagen. Der Schwerpunkt des Modells liegt darauf, den Kontext der Wörter zu verstehen. Die Verarbeitung der Textdaten erfolgt nicht mehr unidirektional von links nach rechts oder von rechts nach links.

Wenn wir die Vorhersage des nächsten Satzes als Trainingsmethode verwenden, erhält das Modell ein Paar von zwei Sätzen. Das Ziel des Modells ist es, vorherzusagen, ob auf den ersten Satz ein zweiter Satz folgt. Daher konzentriert sich das resultierende Modell hauptsächlich darauf, wie ein Satzpaar zusammenhängt.

Beide Modelle wurden zusammen trainiert, um die kombinierte Verlustfunktion der beiden Strategien zu minimieren.

**Zusammenfassung**

Die Verwendung von NLP in der Informatik reicht bis in die 1950er Jahre zurück. Es gibt eine Vielzahl von Anwendungsbereichen für NLP, darunter Themen wie die Beantwortung von Fragen, die Stimmungsanalyse, die Erkennung von benannten Entitäten und die Themenerkennung.

Um Sprache mit Computern verarbeiten zu können, werden Vektorisierungstechniken wie Bag-of-Words, Wortvektoren und Satzvektoren eingesetzt. Diese Modelle sind jedoch mit einigen Einschränkungen verbunden. Wenn zum Beispiel Bag-of-Words verwendet wird, verlieren wir alle Informationen über die Wortreihenfolge. Daher kann dieses Modell nur verwendet werden, wenn die Wortfolge

nicht entscheidend ist. Außerdem haben einige Modelle, darunter BERT, Einschränkungen hinsichtlich der Eingabelänge eines Textes (z. B. 256-Wort-Token). Ein größerer Textabschnitt kann nur mithilfe von Tricks eingebettet werden, z. B. indem er in kleinere Teile zerlegt wird.

Nichtsdestotrotz hat es in den letzten Jahren enorme Fortschritte im Bereich NLP gegeben, da die Rechenleistung drastisch gestiegen ist und immer größere Datenkorpora zum Trainieren von Sprachmodellen zur Verfügung stehen.



# Lektion 5

## Computer Vision

#### LERNZIELE

Nach der Bearbeitung dieser Lektion können Sie …

… Computer Vision definieren.

… erklären, wie Bilder als Pixel dargestellt werden.

… zwischen Erkennung, Beschreibung und Abgleich von Merkmalen unterscheiden.

… Verzeichnungen mit Kalibrierungsmethoden korrigieren.

DL-E-DLBDSEAIS01-U05

1. Computer Vision

### Einführung

In dieser Lektion werden die grundlegenden Prinzipien der Computer Vision besprochen. Wir beginnen mit einer Definition des Themas, dem historischen Hintergrund und einem Überblick über die wichtigsten Aufgaben der Computer Vision. Danach lernen Sie, wie ein Bild als ein Array von Pixeln dargestellt werden kann und wie Bilder mithilfe von Filtern verändert werden können. Wir zeigen Ihnen, wie Sie Merkmale wie Kanten, Ecken und Flecken in Bildern erkennen können. Dieses Wissen nutzen wir, um zu veranschaulichen, wie wir die Kalibrierung nutzen und mit Verzeichnungen umgehen können.

Darüber hinaus befasst sich diese Lektion mit dem Thema der semantischen Segmentierung, mit der sich Pixel in Kategorien einteilen lassen.

### Einführung in Computer Vision

Computer Vision ist ein Thema, das mehrere Disziplinen vereint. Es handelt sich um eine Mischung aus Informatik (insbesondere Künstliche Intelligenz) und Ingenieurwissenschaften (Wiley & Lucas, 2018). Computer Vision versucht, die menschliche visuelle Wahrnehmung durch die Verarbeitung und Analyse visueller Daten zu modellieren. Bei diesen Daten kann es sich zum Beispiel um statische Bilder oder Videos von Kameras handeln. Das Ziel ist es, ein umfassendes Verständnis für die visuellen Aspekte der realen Welt zu erlangen (Wiley & Lucas, 2018). Computer Vision umfasst Aufgaben, wie die Klassifizierung von Objekten oder die Bewegungserkennung.

###### Historische Entwicklung

Die Forschung im Bereich Computer Vision begann in den 1960er Jahren an einigen der führenden Universitäten für Robotik und KI, wie der Stanford University, dem Massachusetts Institute of Technology und der Carnegie Mellon University. Das Ziel dieser frühen Forschung war es, das visuelle System des Menschen zu imitieren (Szeliski, 2022). Forscherinnen und Forscher versuchten, Roboter intelligenter zu machen, indem sie den Prozess der Bildanalyse mit einer an den Computer angeschlossenen Kamera automatisierten. Der große Unterschied zwischen digitaler Bildverarbeitung und Computer Vision bestand damals darin, dass die Forschenden versuchten, die 3D-Struktur aus der realen Welt zu rekonstruieren, um ein besseres Verständnis der Szene zu erhalten (Szeliski, 2022).

Frühe Grundlagen von Algorithmen wie Linienbeschriftung, Kantenextraktion, Objektdarstellung und Bewegungsabschätzung gehen auf die 1970er Jahre zurück (Szeliski, 2022). In den 1980er Jahren verlagerte sich der Schwerpunkt auf die quantitativen Aspekte der Computer Vision und die mathematische Analyse. Es entwickelten sich Konzepte, wie z. B. die Ableitung von Formen aus Merkmalen wie Textur, Schattierung, Konturmodellen und Fokus. In den 1990er Jahren wurden Methoden aus der **Fotogrammetrie** verwendet, um Algorithmen für sparsame 3D-Rekonstruktionen von Szenen auf der Grundlage mehrerer Bilder zu entwickeln. Die Ergebnisse führten zu einem besseren Verständnis der Kamerakalibrierung. Statistische Methoden, insbesondere Eigengesichter (engl. Eigenfaces), wurden für die Gesichtserkennung aus Bildern verwendet. Aufgrund der zunehmenden Interaktion zwischen Computer Vision und

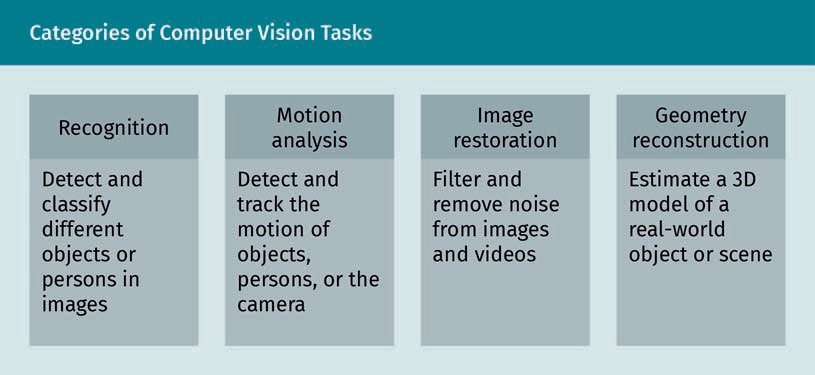
Computer Vision

Computergrafik in den letzten Jahren haben sich Methoden wie Morphing, bildbasierte Modellierung und Rendering, Image Stitching, Light-Field-Rendering und Interpolation von Ansichten stark verändert (Szeliski, 2022).

Die aktuellen Entwicklungen gehen in Richtung Optimierungs-Frameworks und Machine-Learning-Ansätze für diese merkmalsbasierten Techniken. Weitere Impulse für den Bereich der Computer Vision kommen von den jüngsten Entwicklungen im Bereich des Deep Learning. Diese neuen Methoden übertreffen die klassischen Methoden bei Benchmark-Computerbilddatensätzen in vielen Aufgaben wie Segmentierung und Klassifizierung oder optischem Fluss (O‘Mahony *et al*., 2020).

###### Typische Aufgaben

Es gibt vier Hauptkategorien in der Computer Vision: Erkennungsaufgaben, Bewegungsanalyse, Bildwiederherstellung und Geometrierekonstruktion. Die folgende Abbildung veranschaulicht diese Aufgaben.



Erkennung

In der Computer Vision gibt es verschiedene Arten von Erkennungsaufgaben. Typische Aufgaben sind die Erkennung von Objekten, Personen, Posen oder Bildern. Die Objekterkennung befasst sich mit der Abschätzung verschiedener Klassen von Objekten, die in einem Bild enthalten sind (Zou *et al*., 2019). Zum Beispiel könnte ein sehr einfacher Klassifikator verwendet werden, um zu erkennen, ob sich auf einem Bild ein Gefahrstoffetikett befindet oder nicht. Wenn Sie den Klassifikator spezifischer machen, könnten Sie zusätzlich Informationen über die Art des Etiketts erkennen, wie z. B. „entflammbar“ oder „giftig“. Die Objekterkennung ist auch im Bereich des autonomen Fahrens von Bedeutung, wenn es darum geht, andere Fahrzeuge oder Fußgänger zu erkennen.

Bei der Objektidentifizierung werden Objekte oder Personen in einem Bild anhand eindeutiger Merkmale identifiziert (Barik & Mondal, 2010). Für die Identifizierung von Personen kann ein Computer-Vision-System zum Beispiel Merkmale wie Fingerabdrücke, Gesichter oder Handschriften

Fotogrammetrie

Eine Gruppe von berührungslosen Methoden zur Ableitung der Position und Form von physischen Objekten direkt aus fotografischen Bildern wird Fotogrammetrie genannt.

verwenden. Die Gesichtserkennung zum Beispiel verwendet biometrische Merkmale aus einem Bild und vergleicht sie mit den biometrischen Merkmalen anderer Merkmale aus einer bestimmten Datenbank. Die Personenidentifizierung wird üblicherweise verwendet, um die Identität einer Person für die Zugriffskontrolle zu überprüfen.

Posenabschätzungsaufgaben spielen beim autonomen Fahren eine wichtige Rolle. Das Ziel ist es, die Ausrichtung und/oder Position eines bestimmten Objekts relativ zur Kamera zu bestimmen (Chen *et al*., 2020). Das kann z. B. der Abstand zu einem anderen Fahrzeug oder ein Hindernis auf der Straße sein.

Bei der optischen Zeichenerkennung (OCR) wird handgeschriebener oder gedruckter Text aus einem Bild erkannt und in eine Zeichenkette umgewandelt, die von einer Maschine verarbeitet werden kann (Islam *et al*., 2017). Beim Online-Banking kann OCR beispielsweise verwendet werden, um die relevanten Informationen für Überweisungen, wie den Betrag oder die Kontodaten, aus einer Rechnung zu extrahieren.

Bewegungsanalyse

In der klassischen Odometrie werden Bewegungssensoren verwendet, um die Veränderung der Position eines Objekts im Laufe der Zeit zu schätzen. Bei der visuellen Odometrie hingegen wird eine Bildsequenz von Hand analysiert, um Informationen über die Position und Ausrichtung der Kamera zu sammeln (Aqel *et al*., 2016). Autonome Reinigungsroboter können diese Informationen zum Beispiel nutzen, um den Standort in einem bestimmten Raum abzuschätzen.

Bei Verfolgungsaufgaben wird ein Objekt in aufeinanderfolgenden Frames lokalisiert und verfolgt. Ein Frame kann als ein einzelnes Bild in einer längeren Bildsequenz definiert werden, z. B. in Videos oder Animationen (Yilmaz *et al*., 2006). So können zum Beispiel Menschen, Fahrzeuge oder Tiere verfolgt werden.

Rauschen In der Computer Vision bezieht sich Rauschen auf einen Qualitätsverlust eines Bildes, der durch eine Signalstörung

verursacht wird.

Bildwiederherstellung

Bei der Bildwiederherstellung geht es darum, ein verschwommenes oder verrauschtes Bild als ein Bild mit besserer und klarerer Qualität wiederherzustellen. Das können zum Beispiel alte Fotos sein, aber auch Filme, die im Laufe der Zeit beschädigt wurden. Um die Bildqualität wiederherzustellen, können Filter wie Median- oder Tiefpassfilter das **Rauschen** beseitigen (Dhruv *et al*., 2017). Heutzutage können Methoden der Bildwiederherstellung auch dazu verwendet werden, fehlende oder beschädigte Teile eines Kunstwerks wiederherzustellen.

Geometrierekonstruktion

Bei der Geometrierekonstruktion werden virtuelle 3D-Modelle von Szenen aus Videos oder Bildern oder sogar realen Objekten geschätzt (Han *et al*., 2021). Dies geschieht in der Regel auf der Grundlage mehrerer Bilder, die aus verschiedenen Blickwinkeln aufgenommen wurden.

###### Herausforderungen im Bereich Computer Vision

In der Computer Vision gibt es fünf große Herausforderungen, die bewältigt werden müssen (Szeliski, 2022):

Computer Vision

* Die Beleuchtung eines Objekts ist sehr wichtig. Wenn sich die Lichtverhältnisse ändern, kann dies zu unterschiedlichen Ergebnissen beim Erkennungsprozess führen. Zum Beispiel kann Rot leicht als Orange erkannt werden, wenn die Umgebung hell ist.
* Bei Erkennungsaufgaben kann auch die Unterscheidung ähnlicher Objekte schwierig sein. Wenn ein System darauf trainiert ist, einen Ball zu erkennen, könnte es auch versuchen, ein Ei als Ball zu identifizieren.
* Die Größe und das Seitenverhältnis von Objekten in Bildern oder Videos stellen eine weitere Herausforderung in der Computer Vision dar. Auf einem Bild erscheinen weiter entfernte Objekte kleiner als nähere Objekte, selbst wenn sie eigentlich gleich groß sind.
* Algorithmen müssen in der Lage sein, mit der Rotation eines Objekts zurecht zu kommen. Wenn wir zum Beispiel einen Bleistift auf einem Tisch betrachten, kann er entweder wie eine Linie aussehen, wenn wir ihn von oben betrachten, oder wie ein Kreis, wenn wir in eine andere Perspektive wechseln.
* Die Position von Objekten kann variieren. In der Computer Vision wird dieser Effekt als Translation bezeichnet. Um auf unser Beispiel mit dem Bleistift zurückzukommen: Für den Algorithmus sollte es keinen Unterschied machen, ob sich der Bleistift in der Mitte des Papiers oder daneben befindet.

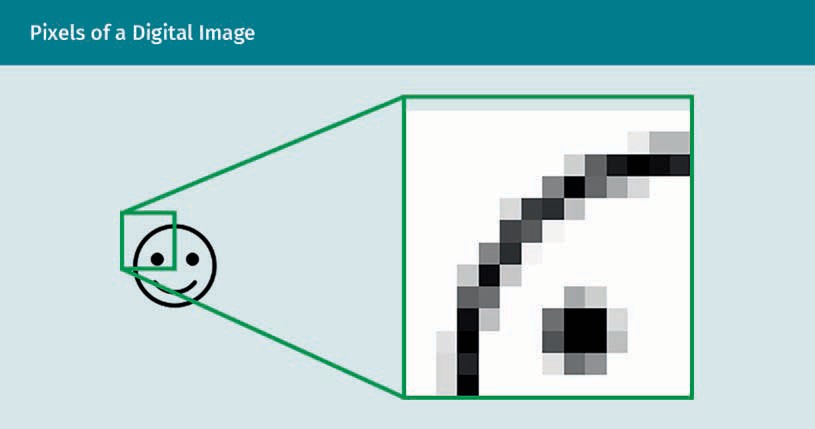
Aufgrund dieser Herausforderungen wird viel an Algorithmen geforscht, die größen-, rotations- und/oder translationsinvariant sind (Szeliski, 2022).

### Bilddarstellung und Geometrie

Bei Computer Vision geht es um die Verarbeitung digitaler Bilder. Um Bilder mit einem Computer verarbeiten zu können, wird in diesem Abschnitt zunächst erklärt, wie Bilder in Form von numerischen Daten dargestellt werden. Zu diesem Zweck führen wir das Konzept der Pixel ein. Im Folgenden werden wir uns mit Filtern befassen und damit, wie Bilder mit Filtern verändert werden können.

###### Pixel

Bilder werden als zweidimensionales Pixel-Array konstruiert (Lyra *et al*., 2011). Ein Pixel ist die kleinste Einheit eines Bildes. Das Wort setzt sich aus den beiden Begriffen „pictures“ (pix) und „element“ (el) zusammen (Lyon, 2006). Ein Pixel wird normalerweise als ein einzelnes Quadrat mit einer Farbe dargestellt. Dies wird sichtbar, wenn Sie sehr weit in ein digitales Bild hineinzoomen. Ein Beispiel für die Pixel eines Bildes sehen Sie in der folgenden Abbildung.



Mit der „Auflösung“ eines Bildes wird die Anzahl der Pixel angegeben. Je höher die Auflösung, desto mehr Details sind auf dem Bild zu erkennen. Umgekehrt kann das Bild bei einer niedrigen Auflösung unscharf oder verschwommen wirken.

Farbdarstellungen

Es gibt verschiedene Möglichkeiten, die Farbe eines Pixels als numerischen Wert darzustellen. Am einfachsten ist es, monochrome Bilder zu verwenden. In diesem Fall wird die Farbe eines Pixels durch ein einziges Bit repräsentiert, das 0 oder 1 ist. In einem True-Color-Bild wird ein Pixel mit 24 Bit dargestellt.

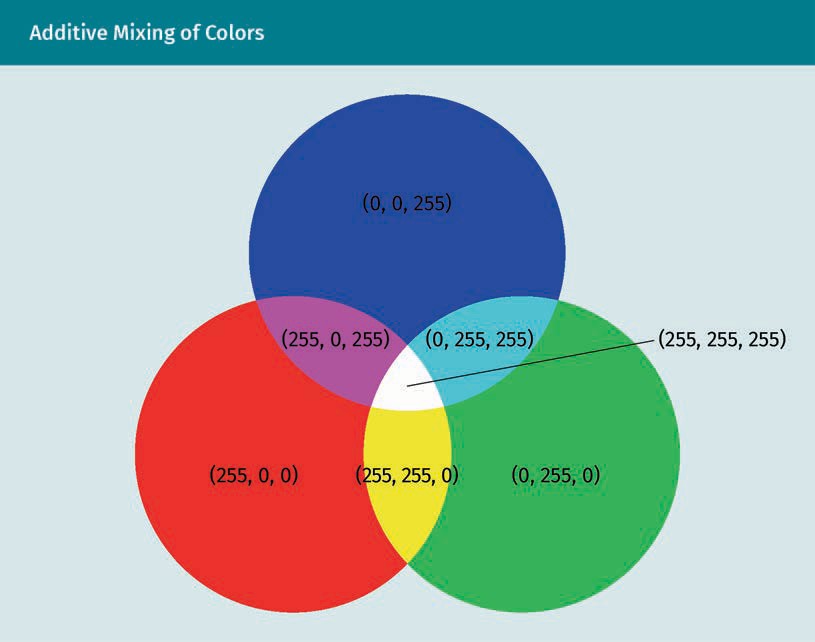
Die folgende Tabelle zeigt die wichtigsten Farbdarstellungen mit der jeweiligen Anzahl der verfügbaren Farben (Farbtiefe).

|  |  |  |
| --- | --- | --- |
| **Farbdarstellungen in Bildern** | | |
| Name | Farbdarstellung | Farbtiefe |
| Monochrom | 1 Bit | 2 Farben |
|  | 8 Bit | 28 = 256 Graustufen-Intensitätsstufen oder Farben |
| Real Color | 15 Bit | 215 = 32.768 Farben |
| High Color | 16 Bit | 216 = 65.536 Farben |
| True Color | 24 Bit | 224 = 16.777.216 Farben |

Computer Vision

|  |  |  |
| --- | --- | --- |
| Name | Farbdarstellung | Farbtiefe |
| Deep Color | 30 – 48 Bit | 230 - 248 Farben |

Eine Möglichkeit, Farben darzustellen, ist die RGB-Farbdarstellung. Wir veranschaulichen dies anhand der 24-Bit-Farbdarstellung. Bei RGB werden die 24 Bit eines Pixels in drei Teile von jeweils 8 Bit Länge aufgeteilt. Jeder dieser Teile steht für die Intensität einer Farbe zwischen 0 und 255. Der erste ist die Farbe Rot (R), die zweite Grün (G), und die letzte Blau (B). Aus diesen drei Komponenten können alle anderen Farben additiv gemischt werden. Zum Beispiel ergibt der Farbcode RGB(0, 255, 0) 100 Prozent Grün. Wenn alle Werte auf 0 gesetzt werden, ist die resultierende Farbe Schwarz. Wenn alle Werte auf 255 gesetzt sind, ergibt das Weiß. Die Abbildung unten zeigt, wie die Farben additiv gemischt werden.



Eine weitere Möglichkeit, Farben darzustellen, ist das CMYK-Modell. Im Gegensatz zur RGB-Darstellung handelt es sich dabei um ein subtraktives Farbmodell, das aus Cyan, Magenta, Gelb und der sogenannten Key-Farbe (Schwarz) besteht. Die Farbwerte in CMYK reichen von 0 bis 1. Um Farben von RGB in CMYK zu konvertieren, müssen die RGB-Werte daher zunächst durch 255 geteilt werden. Daher können die Werte für Cyan, Magenta, Gelb und Key wie folgt berechnet werden:

K=1 − max R , G , B

255

1 − R − K

255

C= 1 − K

1 − G − K

255

M= 1 − K

1 − B − K

255

Y =

1 − K

255

255

Während RGB besser für die digitale Darstellung von Bildern geeignet ist, wird CMYK in der Regel für gedrucktes Material verwendet.

Bilder als Funktionen

Wir werden uns nun ansehen, wie ein Bild aus einzelnen Pixeln aufgebaut werden kann. Dazu benötigen wir eine Funktion, die eine zweidimensionale Koordinate (x,y) auf einen bestimmten Farbwert abbilden kann. Auf der x-Achse beginnen wir links mit einem Wert von 0 und fahren nach rechts fort, bis die maximale Breite eines Bildes erreicht ist. Auf der y-Achse beginnen wir oben bei 0 und erreichen unten die Höhe eines Bildes.

Betrachten wir die Funktion f x, y für ein 8-Bit-Graustufenbild. Die Funktionswerte von f 42, 100 = 0 würden bedeuten, dass wir ein schwarzes Pixel 42 Pixel rechts und 100 Pixel unterhalb des Startpunkts haben. Bei einem 24-Bit-Bild wäre das Ergebnis der Funktion ein dreifacher Wert, der die RGB-Intensität des angegebenen Pixels angibt.

###### Filter

Filter spielen in der Computer Vision eine wichtige Rolle, wenn es darum geht, Effekte auf ein Bild anzuwenden, Techniken wie Glättung oder Inpainting zu implementieren oder nützliche Informationen aus einem Bild zu extrahieren, wie z. B. die Erkennung von Ecken oder Kanten. Sie kann als eine Funktion definiert werden, die ein Bild als Eingabe erhält, Änderungen an diesem Bild vornimmt und das gefilterte Bild als Ausgabe zurückgibt (Szeliski, 2022).

2D-Faltung

Eine häufig verwendete Technik zum Filtern von Bildern ist die 2D-Faltung. Wenn die 2D-Faltung auf ein Bild angewendet wird, wird eine kleine Matrix (auch Faltungsmatrix oder Kernel genannt) Pixel für Pixel über die Matrix eines Bildes bewegt und mit den Werten der Matrix multipliziert. Die Faltungsmatrix besteht normalerweise aus 3x3 oder 5x5 Elementen (Smith, 1997).

Die Faltung eines Bildes I mit einem Kernel k mit einer Größe von n und einer Zentralkoordinate a

kann wie folgt berechnet werden:

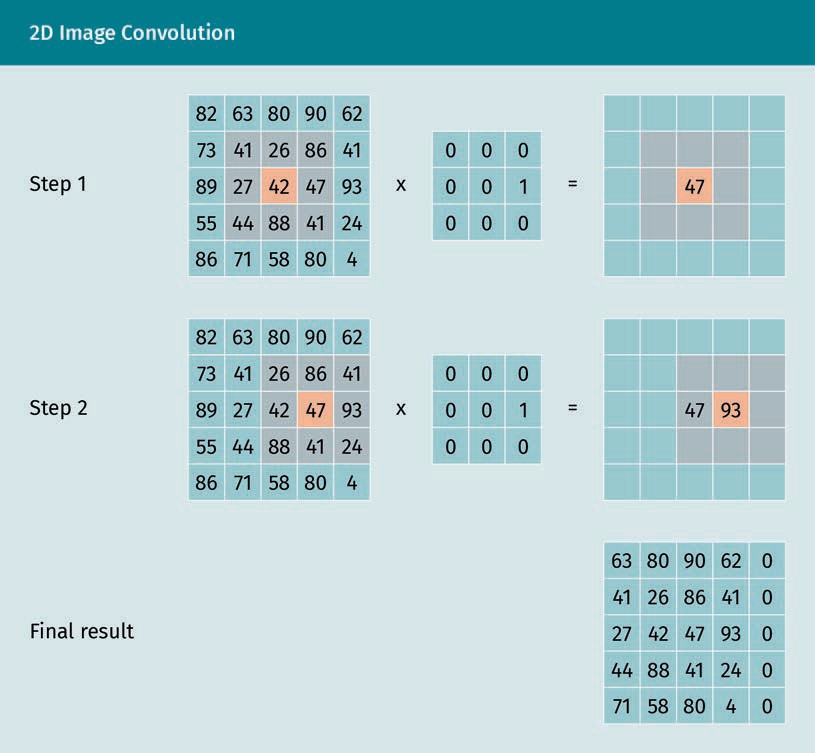
n n

I ⋅ x, y = ∑ ∑I x − 1 + a, y − j + a k i, j

i = 1 i = j

Computer Vision

wobei I · x, y der Wert des resultierenden Bildes I · an der Position x, y ist, während I das Originalbild ist. Um den Prozess zu verstehen, werden wir das folgende Beispiel einer 3x3-Faltung verwenden. Die für die Faltung verwendete Kernelmatrix ist in der mittleren Spalte der Abbildung zu sehen.



Die Kernelmatrix wird über jede Position des Eingabebildes bewegt. In unserem Eingabebild ist die aktuelle Position orange markiert. In unserem Beispiel beginnen wir im Zentrum des Bildes und multiplizieren das Bild an dieser Position mit den Werten der Kernelmatrix. Der resultierende Wert für das Zentrum unseres gefilterten Bildes wird wie folgt berechnet:

0 · 41 + 0 · 26 + 0 · 86 + 0 · 27 + 0 · 42 + 1 · 47 + 0 · 44 + 0 · 88 + 0 · 41 = 47

Im nächsten Schritt verschieben wir die Kernelmatrix an die nächste Position und berechnen den neuen Wert des gefilterten Bildes:

0 · 26 + 0 · 86 + 0 · 41 + 0 · 42 + 0 · 47 + 1 · 93 + 0 · 88 + 0 · 41 + 0 · 24 = 93

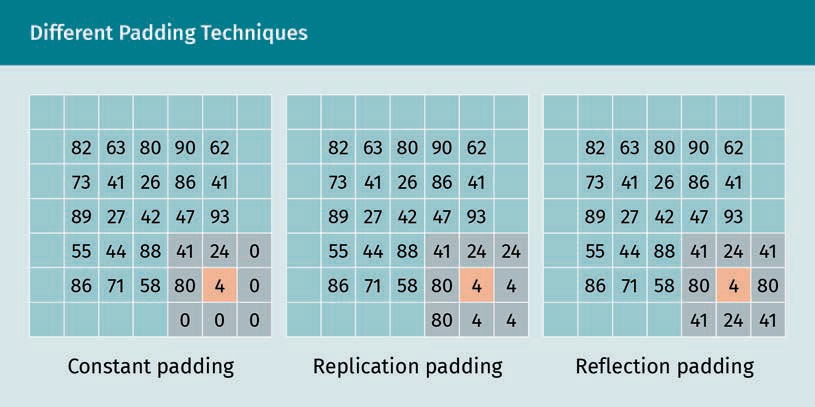
Die untere Zeile in unserer Abbildung zeigt das Ergebnis, nachdem alle Positionen des Bildes mit der Kernelmatrix multipliziert wurden.

Auffülltechniken

Wenn Faltungstechniken auf Bilder angewendet werden, stehen wir vor dem Problem, dass in den ersten und letzten Zeilen und Spalten eines Bildes nicht genügend Werte vorhanden sind, um die Matrixmultiplikation mit der Faltungsmatrix durchzuführen. Um dieses Problem zu lösen, können wir zusätzliche Werte am Rand unserer Eingabebilder hinzufügen. Dieser Vorgang wird als Padding (dt. Auffüllen) bezeichnet (Szeliski, 2022).

Es gibt drei gängige Auffülltechniken: Constant Padding, Replication Padding und Reflection Padding.

Beim Constant Padding wird eine konstante Zahl (z. B. Null) verwendet, um die leeren Zellen zu füllen. Beim Replication Padding werden die Werte aus den nächstgelegenen Nachbarzellen repliziert. Beim Reflection Padding wird der Wert von der gegenüberliegenden Seite eines Pixels verwendet, um die Zelle zu füllen. Zum Beispiel wird die Zelle oben links mit dem Wert unten rechts gefüllt (Szeliski, 2022).



Die Abbildung oben zeigt, wie die drei Auffülltechniken auf ein Bild angewendet werden.

###### Verzeichnung

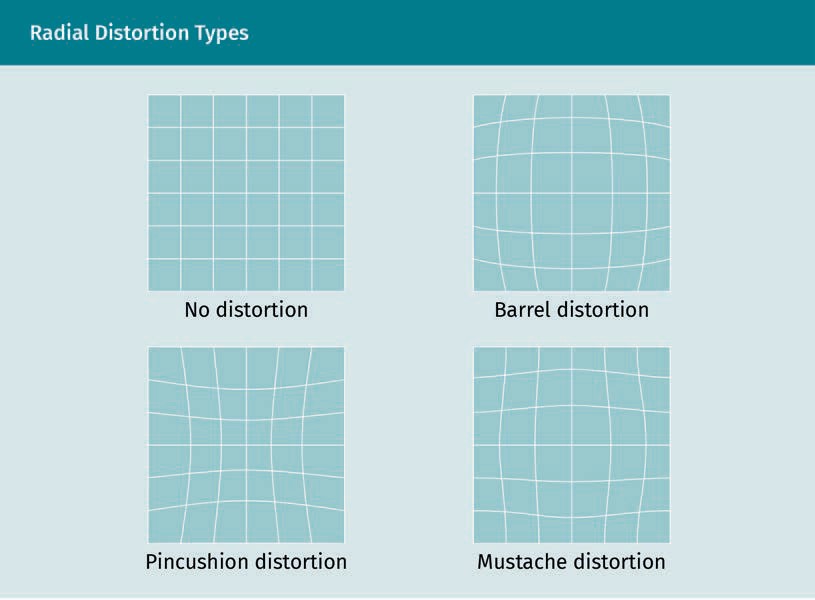
Bei der Bildverarbeitung in der Computer Vision wird normalerweise davon ausgegangen, dass ein Bild, das wir von einer Kamera erhalten, eine lineare Projektion einer Szene ist. Das heißt, wenn wir in der realen Welt eine gerade Linie vor uns haben, können wir erwarten, dass diese auch in der digitalen Darstellung des Bildes eine gerade Linie ist (Szeliski, 2022). In der Praxis verursachen Kameraobjektive jedoch oft Verzeichnungen. Es gibt zwei Arten von Verzeichnungen – radiale und tangentiale –, die im Folgenden erklärt werden.

Computer Vision

Radiale Verzeichnung

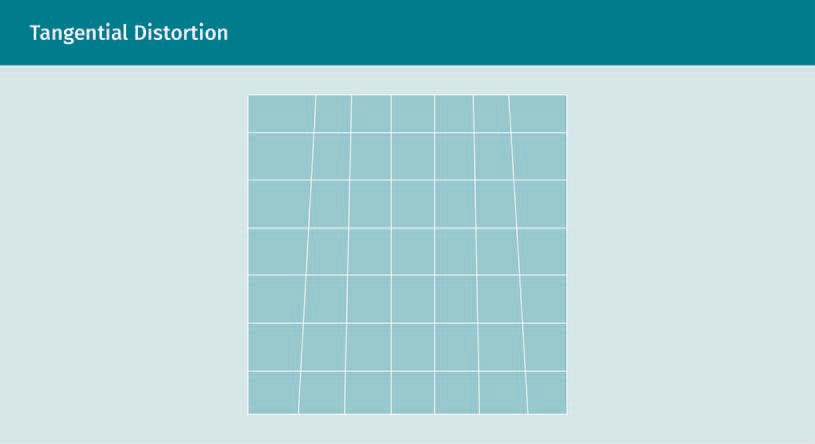
Eine radiale Verzeichnung tritt auf, wenn Linien, die normalerweise gerade sind, sich zum Rand des Kameraobjektivs hin biegen (Wang *et al*., 2009). Die Stärke der Verzeichnung hängt von der Größe des Objektivs ab. Bei kleineren Objektiven kommt es zu einer stärkeren Verzeichnung. Außerdem ist die radiale Verzeichnung bei der Verwendung von Weitwinkelobjektiven stärker ausgeprägt. Im Allgemeinen gibt es vier Arten der radialen Verzeichnung (Szeliski, 2022):

1. Tonnenförmige Verzeichnung/positive radiale Verzeichnung: Linien in der Mitte eines Bildes werden nach außen gebogen.
2. Kissenförmige Verzeichnung/negative radiale Verzeichnung: Linien in der Mitte eines Bildes werden nach innen gebogen.
3. Komplexe Verzeichnung/radiale Schnurrbartverzeichnung: Linien mit einer Kombination aus positiver und negativer Verzeichnung.
4. Radiale Fischaugenverzeichnung: Tritt bei Ultraweitwinkelobjektiven auf, z. B. bei einem Guckloch.



Tangentiale Verzeichnung

Neben der radialen Verzeichnung ist die tangentiale Verzeichnung ein weiterer Effekt, der bei der digitalen Bildbearbeitung häufig zu beobachten ist. Tangentiale Verzerrungen werden verursacht, wenn die Bildsensoreinheit und das Kameraobjektiv nicht korrekt aufeinander abgestimmt sind. Wenn das Kameraobjektiv und die Bildebene nicht parallel sind, sehen die Verzeichnungen wie in der folgenden Grafik dargestellt aus.



Um Verzeichnungen in der digitalen Bildverarbeitung zu behandeln, können mathematische Modelle wie das Brown-Conrady-Modell (Brown, 1966) verwendet werden, um die Auswirkungen der Verzeichnung zu beschreiben und zu korrigieren. Um diese Modelle anwenden zu können, ist es wichtig, dass die extrinsischen und intrinsischen Parameter der Kamera bekannt sind. Diese Parameter können durch Kalibrierung bestimmt werden.

###### Kalibrierung

Die Kamerakalibrierung schätzt die extrinsischen und intrinsischen Parameter einer Kamera (Szeliski, 2022). Nur wenn diese Parameter bekannt sind, kann die Kamera richtig kalibriert werden. Die Kalibrierung ermöglicht es, Verzeichnungen aus den Bildern zu extrahieren.

Extrinsische Merkmale einer Kamera sind z. B. die Ausrichtung in Koordinaten der realen Welt und die Position der Kamera. Zu den intrinsischen Eigenschaften gehören Parameter wie das optische Zentrum, die Brennweite und die Verzeichnungsparameter des Objektivs.

Wenn die Kamera richtig kalibriert ist, können Bilder zuverlässig von Verzeichnungen befreit werden. Dies ermöglicht es uns zum Beispiel, Entfernungen und Größen auf diesen Bildern in Einheiten wie Meter zu messen und somit ein 3D-Modell des zugrunde liegenden Szenarios aus der realen Welt zu rekonstruieren.

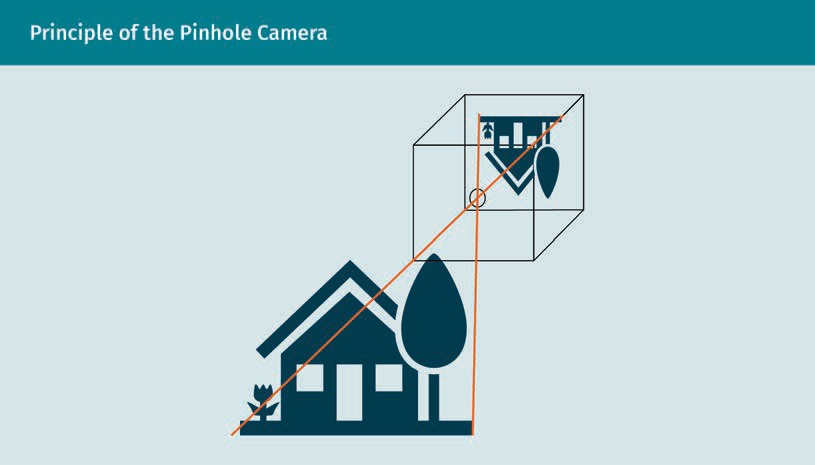
Lochkamera Die Lochkamera war die allererste Kamera. Es handelt sich um eine Schachtel

Techniken

Um die Kalibrierungsparameter einer Kamera bestimmen zu können, ist es wichtig, die Koordinaten der ursprünglichen 3D-Darstellung der realen Welt sowie die entsprechenden Koordinaten des 2D-Bildes zu kennen (Szeliski, 2022). Ein gutes Beispiel, um den Prozess der Übertragung eines 3D-Bildes in ein 3D-Bild in einem vereinfachten Modell zu veranschaulichen, ist die **Lochkamera**. Bei diesem Prozess verwenden wir drei Koordinatensysteme:

Computer Vision

1. das 3D-Koordinatensystem der Kamera
2. das 3D-Koordinatensystem der realen Welt
3. das 2D-Koordinatensystem des projizierten Bildes



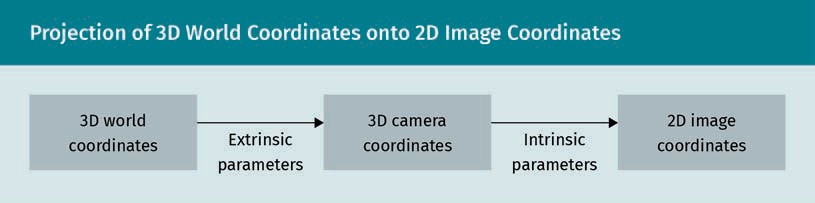
Der Projektionsprozess erfolgt in zwei Schritten:

1. Transformieren der Koordinaten aus der 3D-Welt in die 3D-Kamerakoordinaten. Für diesen Schritt werden extrinsische Parameter wie die Rotation und Translation der Informationen verwendet.
2. Transformieren der 3D-Kamerakoordinaten in die 2D-Bildkoordinaten. In diesem Schritt werden intrinsische Parameter wie Brennweite, Verzeichnungsparameter und optisches Zentrum angewendet.

Um die 3D-Koordinaten aus der realen Welt auf ein zweidimensionales Bild abzubilden, wird eine 3x4-Projektionsmatrix (oft auch als Kameramatrix bezeichnet) verwendet. Wenn wir die 3D-Koordinaten mit dieser Matrix multiplizieren, erhalten wir die 2D-Koordinaten des projizierten Punktes auf dem Bildausschnitt.

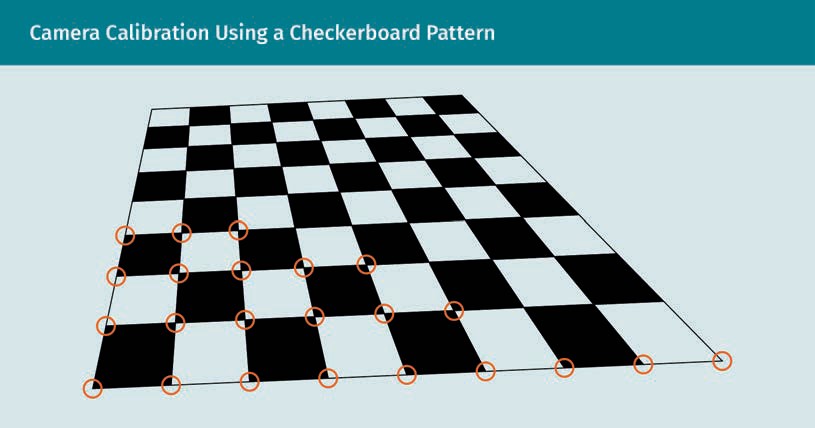
Die folgende Abbildung veranschaulicht die Schritte des Projektionsprozesses, wenn 3D-Koordinaten der realen Welt in 2D-Bildkoordinaten transformiert werden.

mit einer kleinen Lochblende, die ein Bild auf der gegenüberliegenden Seite der Schachtel erzeugt.



Um die oben dargestellten Projektionsschritte anzuwenden, müssen wir die intrinsischen und extrinsischen Kameraparameter kennen. Diese können mithilfe der Kamerakalibrierung geschätzt werden. Um die praktische Umsetzung des Kalibrierungsprozesses zu verstehen, werden wir uns flexible Techniken zur Kamerakalibrierung ansehen (Zhang, 2000).

Diese Technik verwendet zwei oder mehr Bilder als Eingabe sowie die Größe des Objekts. Ein gutes Objekt für die Kamerakalibrierung ist z. B. ein Schachbrettmuster. Nach dem Kalibrierungsprozess erhalten wir die extrinsischen Parameter Rotation und Translation sowie die intrinsischen Kameraparameter optisches Zentrum, Brennweite und Verzeichnung.



Der Kalibrierungsprozess funktioniert folgendermaßen:

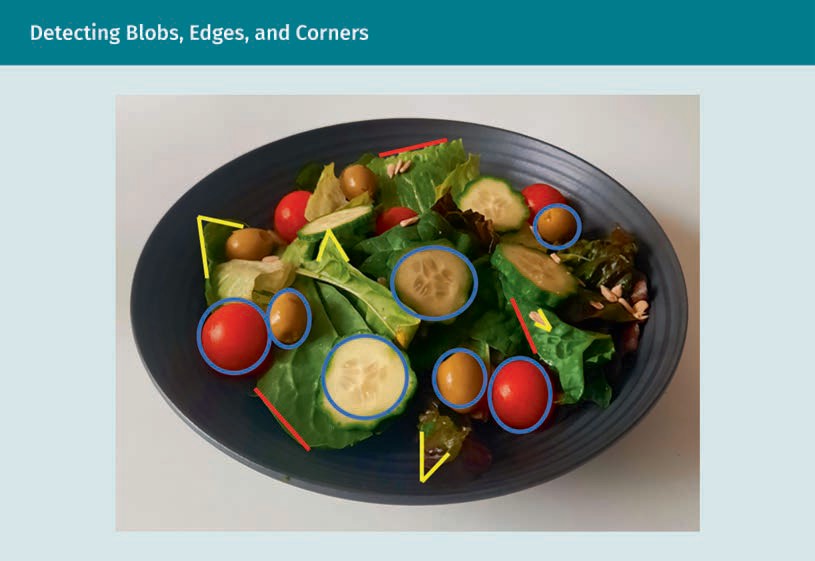
* 1. Wir wählen mindestens zwei Beispielbilder aus, bei denen es sich um gut strukturierte Muster handeln sollte, z. B. ein Schachbrettmuster.
  2. Wir identifizieren markante Punkte in jedem Bild. Wenn wir ein Schachbrettmuster verwenden, können dies zum Beispiel die Ecken der einzelnen Quadrate sein. Aufgrund der klaren Struktur des Schachbrettmusters mit den schwarzen und weißen Quadraten sind die Ecken leicht zu erkennen. Sie haben in beiden Richtungen ein hohes Gefälle an den Ecken.
  3. Lokalisierung der Ecken der Quadrate. Für das Schachbrettmuster kann dies auf eine sehr robuste Weise geschehen. Um die 3D-Koordinaten der Ecken in der realen 3D-Welt zu identifizieren, müssen wir die Größe des Schachbretts kennen und benötigen zwei oder mehr Beispielbilder. Außerdem kennen wir die 2D-Koordinaten der Ecken im Bild aus dem Foto, das die Kamera aufgenommen hat. Mit diesen Informationen können wir die Kameramatrix und die Verzeichnungskoeffizienten berechnen. Die Verzeichnungskoeffizienten können durch Anwendung des Brown-Conrady-Modells (Brown, 1966) verwendet werden.

Computer Vision

### Merkmalserkennung

Im Zusammenhang mit Computer Vision können Merkmale als Punkte von Interesse in einem Bild definiert werden, die die erforderlichen Informationen zur Lösung eines bestimmten Problems enthalten (Hassaballah *et al*., 2016). Um diese Merkmale in einem Bild zu finden, gibt es eine große Auswahl an Algorithmen zur Merkmalserkennung. Sobald die Merkmale erkannt wurden, können die semantischen Informationen über sie extrahiert werden. Die Koordinaten eines Merkmals, d. h. die Position, an der es sich in einem Bild befindet, ist der Schlüsselpunkt des Merkmals. Die semantischen Informationen, die über ein Merkmal extrahiert werden, werden in einem Vektor gespeichert, der auch als Merkmalsdeskriptor oder Merkmalsvektor bezeichnet wird. Die Erkennung und Extraktion von Merkmalen ist oft ein wichtiger Teil der Vorverarbeitung in Anwendungen des maschinellen Lernens. Die extrahierten Merkmalsvektoren können anschließend als Eingabe für die Bildklassifizierung verwendet werden. Bei der Bewegungsverfolgung oder der Erkennung von Personen oder ähnlichen Objekten in mehreren Bildern kann der Abgleich von Merkmalen verwendet werden.

Die häufigsten Arten von Merkmalen sind Blobs, Kanten und Ecken. Blobs werden aus einer Gruppe von Pixeln gebildet, die einige Eigenschaften gemeinsam haben. Regionen, die sich in ihren Eigenschaften unterscheiden, gehören zu verschiedenen Blobs. Dies kann z. B. eine andere Farbe oder Helligkeit im Vergleich zu den umliegenden Bereichen einer Region sein. Kanten werden durch eine signifikante Änderung der Helligkeit der Pixel angezeigt. Sie lassen sich an einer Diskontinuität der Bildintensität erkennen, d. h. an einer plötzlichen Veränderung der Helligkeit eines Bildes (Jain *et al*., 1995). Ecken sind die Verbindung zwischen zwei Kanten. Das Bild unten veranschaulicht den Unterschied zwischen Blobs (blau), Kanten (rot) und Ecken (gelb).



Wenn wir alle Tomaten auf dem Bild erkennen wollen, können wir einen Algorithmus verwenden, der alle Blobs erkennt. Es bleibt jedoch die Aufgabe, Tomaten von anderen runden Objekten wie Oliven oder Gurken zu unterscheiden. Diese Aufgabe lässt sich bewältigen, wenn wir einen Algorithmus zur Merkmalsbeschreibung verwenden, um die für eine Tomate charakteristischen Informationen zu extrahieren und daraus einen Merkmalsdeskriptor zu konstruieren. Der Merkmalsdeskriptor könnte z. B. Informationen über die n umgebenden Pixelwerte oder die Farbe der Pixel enthalten.

Sobald wir den Merkmalsdeskriptor für unseren Gurkenkandidaten haben, können wir ihn mit anderen Merkmalsdeskriptoren von Gurkenbildern vergleichen, indem wir einen Algorithmus zum Abgleich der Merkmale verwenden. Mit diesem Algorithmus zum Abgleich von Merkmalen können wir alle Gurkenscheiben auf dem Bild erkennen. Wie wir in unserem Beispiel gesehen haben, wird das Feature Engineering in der Regel in drei Schritten durchgeführt:

1. Merkmalserkennung
2. Merkmalsbeschreibung/-extraktion
3. Merkmalsabgleich

Merkmalserkennung

Um Merkmale wie Kanten oder Ecken zu erkennen, gibt es verschiedene Methoden. Um Kanten in Bildern zu erkennen, kann eine 2D-Faltung verwendet werden. Kanten zeichnen sich durch einen signifikanten Unterschied der Pixelwerte zu den umgebenden Pixeln aus. Wenn wir eine Kante betrachten, gibt es einen deutlichen Unterschied in der Helligkeit und/oder Farbe im Vergleich zu den umliegenden Pixeln.

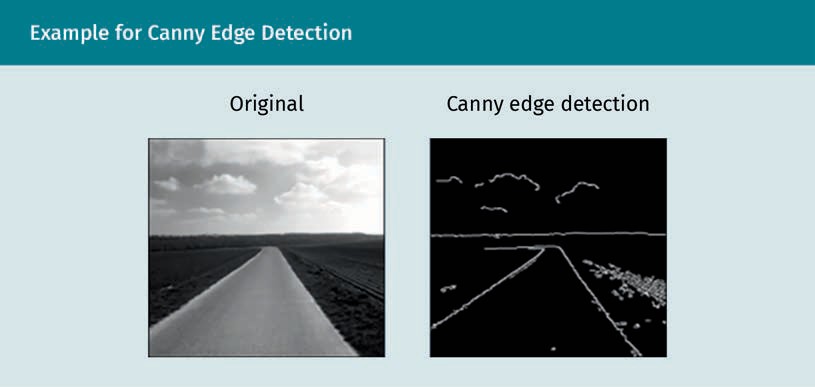


Die Abbildung oben zeigt ein Beispiel für die Kantenerkennung. Die Kante zwischen der Straße und dem umgebenden Gras ist in diesem Beispiel deutlich zu erkennen. Im oberen linken Teil des vergrößerten Bildes können wir einige Variationen von dunkelgrünen Farben sehen, der untere rechte Teil ist mit Variationen von Hellgrau gefüllt. Die Kante trennt beide Teile des Bildes.

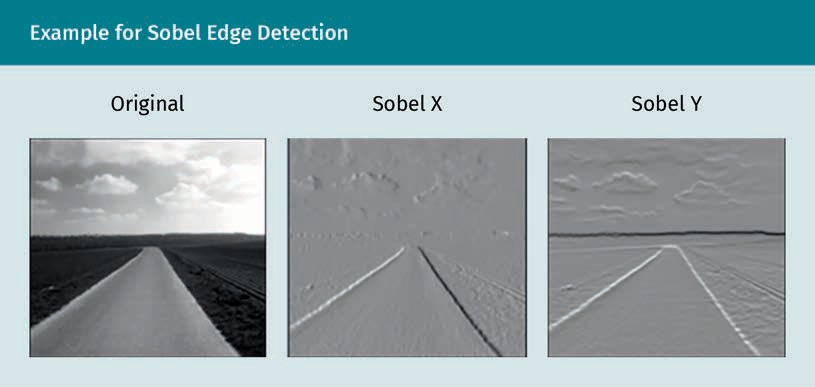
Zwei Techniken, die häufig für die Kantenerkennung verwendet werden, sind der Canny-Kantenerkenner und der Sobel-Filter. Die Canny-Kantenerkennung (Canny, 1986) analysiert die Veränderung zwischen den Pixelwerten. Zu diesem Zweck werden die Ableitungen der x- und y-Koordinaten

Computer Vision

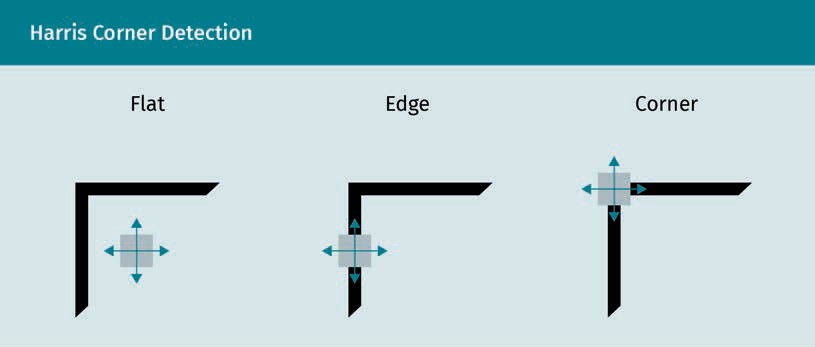
verwendet. Der Algorithmus arbeitet mit zweidimensionalen Werten, d. h. er funktioniert nur bei einfarbigen Bildern wie z. B. Graustufenbildern. Die Abbildung unten zeigt das Ergebnis der Canny-Kantenerkennung in unserem Beispielbild.



Bei der Verwendung von Sobel-Filtern zur Kantenerkennung werden zwei spezielle Kernelmatrizen verwendet, eine für jede der Achsen. Diese Sobel-Operatoren verwenden Faltung, um das Originalbild in ein Verlaufsbild zu übertragen. Hohe Frequenzen im Verlaufsbild weisen auf Bereiche mit den größten Änderungen der Pixelintensität hin, bei denen es sich wahrscheinlich um Kanten handelt. Daher wird der Algorithmus in einem zweiten Schritt oft mit einer Schwellenwertfunktion kombiniert, um die Bilder zu erkennen. Die Abbildung unten zeigt die Sobel-Kantenerkennung für die x- und y-Richtung.



Einer der bekanntesten Algorithmen zur Erkennung von Ecken in Bildern ist die Harris-Eckenerkennung (Harris & Stephens, 1988). Dieser Algorithmus analysiert die Veränderung der Pixelwerte in einem gleitenden Fenster, das in verschiedene Richtungen bewegt wird. Das Schiebefenster kann z. B. 7x7 Pixel groß sein. Die Abbildung zeigt, wie flache Bereiche, Kanten und Ecken mit der Schiebefenstertechnik erkannt werden können.



Das linke Bild zeigt das Fenster in einem flachen Bereich ohne Ecken und Kanten. Im darunter liegenden Fenster gibt es keine signifikante Änderung der Werte der Pixel, wenn das Fenster in eine beliebige Richtung bewegt wird. Im mittleren Bild wird das Fenster an einer Kante verschoben, berührt aber nicht die andere Kante. Das bedeutet, dass sich ein Pixelwert nur ändert, wenn wir das Bild in horizontaler Richtung bewegen. Wenn wir das Bild in vertikaler Richtung bewegen, ändert sich der Pixelwert nicht. In der Abbildung rechts wird das Schiebefenster über eine Ecke bewegt. In diesem Bild gibt es eine signifikante Änderung des Pixelwerts, egal in welche Richtung wir das Bild bewegen.

Wenn wir also Ecken erkennen wollen, müssen wir das Fenster finden, in dem die Veränderung der zugrunde liegenden Pixel in alle Richtungen maximiert ist. Um diese Idee mathematisch zu formalisieren, verwendet die Harris-Eckenerkennung die Sobel-Operatoren, die zuvor erläutert wurden.

Merkmalsbeschreibung

Für die weitere Verarbeitung der bei der Merkmalserkennung erkannten Merkmale ist es wichtig, diese Merkmale so beschreiben zu können, dass ein Computer sie verwenden und voneinander unterscheiden kann. Zu diesem Zweck verwenden wir Merkmalsvektoren/Merkmalsdeskriptoren, die semantische Informationen über die Merkmale enthalten. Eine Möglichkeit, Merkmale zu beschreiben, ist der BRIEF-Algorithmus (Binary Robust Elementary Features) (Calonder *et al*., 2010). Um ein Merkmal zu beschreiben, wird ein binärer Vektor verwendet.

Der Vektor wird mithilfe eines Bildfeldes konstruiert, das durch den Vergleich der Intensität eines Pixelpaares entsteht. In einem ersten Schritt wird ein Fleck p an der Position x zunächst geglättet. Anschließend wird die Pixelintensität p x berechnet. In einem Test τ wird das Ergebnis des Vergleichs in einen binären Wert gemäß der folgenden Gleichung kodiert:

τ p; x, y : =

1 wenn p x < p y

0 sonst

Der größte Vorteil des BRIEF-Algorithmus ist, dass er schnell zu berechnen und einfach zu implementieren ist. Allerdings ist die Merkmalsextraktion für Merkmale, die um mehr als 35 Grad gedreht sind, nicht mehr genau (Hassaballah *et al*., 2016). Algorithmen wie Oriented FAST und Rotated BRIEF (ORB) versuchen, diese Einschränkung zu überwinden (Rublee *et al*., 2011).

Computer Vision

Ein weiterer Algorithmus zur Merkmalsbeschreibung ist der SIFT-Algorithmus (Scale-Invariant Feature Transform) (Lowe, 1999). Der SIFT-Algorithmus wurde durch den SURF-Algorithmus (Speeded-Up Robust Features) erweitert (Bay *et al*., 2008), der eine leistungsstärkere Variante des SIFT-Algorithmus darstellt. Da beide Algorithmen jedoch patentiert sind, können sie nicht so frei verwendet werden wie z. B. ORB. Außerdem ist ihre Genauigkeit im Vergleich zu ORB geringer und der Rechenaufwand höher (Rublee *et al*., 2011).

Merkmalsabgleich

Das Ziel des Merkmalsabgleichs ist es, ähnliche Merkmale in verschiedenen Bildern zu identifizieren. Dies kann zum Beispiel der Fall sein, wenn Sie dieselbe Person in verschiedenen Szenarien erkennen. Der Merkmalsabgleich ist eine wichtige Komponente bei Aufgaben wie Kamerakalibrierung, Bewegungsverfolgung, Objekterkennung und -verfolgung.

Eine sehr einfache Technik für den Merkmalsabgleich ist der Brute-Force-Abgleich, bei dem die Merkmalsdeskriptoren von Ausgangs- und Zielbild verglichen werden und der Abstand zwischen diesen Bildern berechnet wird (Jakubovic & Velagic, 2018). Für numerische Werte der Merkmalsvektoren können wir den euklidischen Abstand verwenden (Wang *et al*., 2005). Für binäre Vektoren, die bei der Verwendung des BRIEF-Algorithmus erzeugt werden, ist die Hamming-Distanz ein geeigneter Ansatz zur Berechnung des Abstands (Torralba *et al*., 2008).

Insbesondere bei großen Datensätzen und hochdimensionalen Merkmalsvektoren bietet die FLANN (Fast Library for Approximate Nearest Neighbors) eine ausgefeiltere Methode für den Merkmalsabgleich. Sie enthält eine Reihe von Algorithmen, die eine Nearest-Neighbors-Suche (Suche nach den nächsten Nachbarn) verwenden und hat einen geringeren Rechenaufwand als ein Brute-Force-Abgleich. Der am besten geeignete Algorithmus wird je nach Datensatz automatisch ausgewählt. Dies ist jedoch weniger genau als der Brute-Force-Abgleich (Muja & Lowe, 2009).

###### Wichtige Merkmale für die Merkmalserkennung und -extraktion

Laut Hassaballah *et al*. (2016) gibt es mehrere Eigenschaften, die ein guter Algorithmus zur Merkmalserkennung und -extraktion aus Bildern haben sollte: Robustheit, Wiederholbarkeit, Genauigkeit, Allgemeinheit, Effizienz und Quantität. Die Merkmale werden in der folgenden Tabelle erläutert.

|  |  |
| --- | --- |
| Wichtige Merkmale eines Algorithmus zur Merkmalserkennung und -extraktion | |
| Robustheit | Zuverlässige Merkmalserkennung auch unter schwierigen Bedingungen, z. B. bei unterschiedlichen Lichtverhältnissen, Rauschen (gestörte Bildsignale) oder Änderungen der Position, der Größe oder der Rotation des Merkmals |
| Wiederholbarkeit | Replizierbarkeit der Merkmalserkennung unabhängig von Blickwinkel und Perspektive |

|  |  |
| --- | --- |
| Wichtige Merkmale eines Algorithmus zur Merkmalserkennung und -extraktion | |
| Genauigkeit | Genaue Lokalisierung eines Merkmals in einem Bild auf der Grundlage seiner Pixelposition |
| Allgemeinheit | Anwendung der Merkmalserkennung und -extraktion in einem anderen Anwendungsfall ohne zusätzliche Anpassungen |
| Effizienz | Geringer Rechenaufwand |
| Quantität | Fähigkeit des Algorithmus, (fast) alle in einem Bild vorhandenen Merkmale zu erkennen, um eine aussagekräftige Darstellung dieses Bildes erzeugen zu können |

###### Herausforderungen bei der Merkmalserkennung

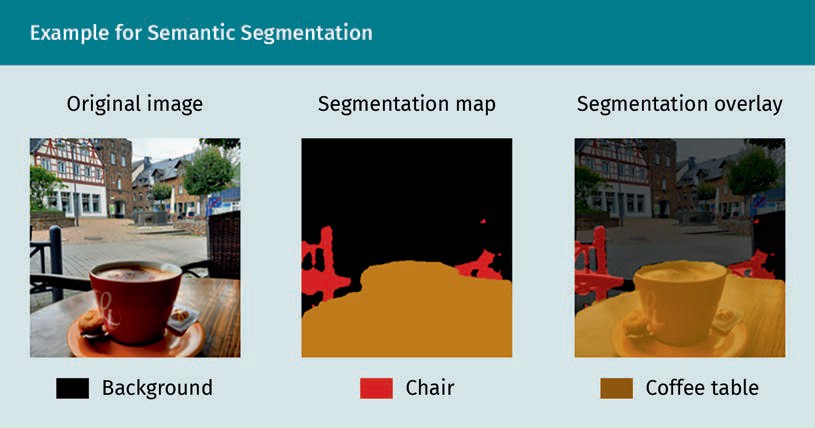
Bei der Merkmalserkennung und -extraktion in einem Bild gibt es mehrere Herausforderungen. Während Menschen Objekte unabhängig von ihrer Lage oder Beleuchtung leicht erkennen können, stellen diese Unterschiede für einen Computer eine große Herausforderung dar. Es wird daher noch intensiv geforscht, um Algorithmen zu entwickeln, die weniger anfällig für Faktoren wie Rauschen, wechselnde Lichtverhältnisse, Änderungen der Kameraperspektive, Rotation oder Verschiebung von Objekten und Größenveränderungen sind.

### Semantische Segmentierung

Bei der semantischen Segmentierung, die auch als Bildsegmentierung bezeichnet wird, werden Teile eines Bildes, die zur gleichen Objektklasse gehören, dem gleichen Cluster zugeordnet (Li *et al*., 2018; Thoma, 2016). Die Vorhersage wird auf Pixelebene durchgeführt, d. h. jedes Pixel eines Bildes wird entsprechend seiner Kategorie klassifiziert.

Zur Durchführung der semantischen Segmentierung erhält der Algorithmus ein Bild mit einem oder mehreren Objekten als Eingabe und gibt ein Bild aus, in dem jedes Pixel entsprechend seiner Kategorie markiert ist. Die folgende Abbildung zeigt, wie die semantische Segmentierung auf ein Bild angewendet werden kann. In dem Bild wird jedes Pixel entweder als Hintergrund, Stuhl oder Kaffeetisch kategorisiert.

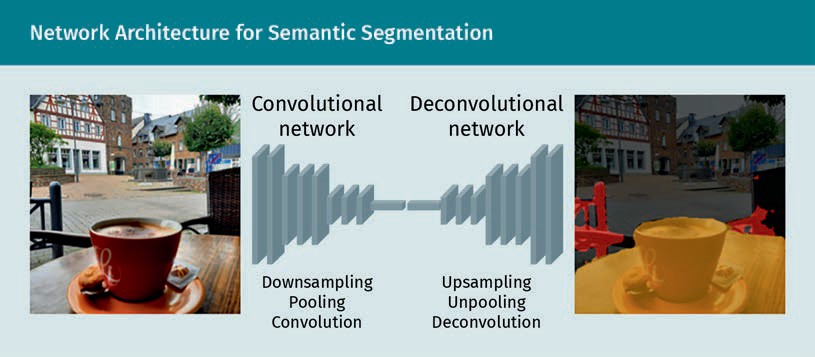
Computer Vision



###### Techniken der semantischen Segmentierung

Algorithmen für die semantische Segmentierung können als strukturiertes Markierungsproblem auf Pixelebene betrachtet werden und basieren häufig auf Convolutional Neural Networks (CNNs, dt. etwa „faltende neuronale Netze“) (Chen *et al*., 2016; Long *et al*., 2015).

Wenn eine CNN-basierte Strategie verwendet wird, kann ein auf den Faltungsschichten aufbauendes Netz gelernt werden. Dieses Netz besteht aus Deconvolution Layers und Convolution Layers. Nach dem Training des Netzes kann es auf die Vorschläge der einzelnen Objekte angewendet werden. Die endgültige Zuordnung der semantischen Segmentierung wird als Kombination der Ergebnisse aus den instanziellen Segmentierungen erstellt (Noh *et al*., 2015). Die Architektur des gesamten neuronalen Netzes ist in der folgenden Abbildung dargestellt.



Der Convolution-Teil des Netzes wird für die Merkmalsextraktion verwendet. Er wandelt das Bild aus der Eingabe in eine mehrdimensionale Darstellung seiner Merkmale um. Das Deconvolution-Netz verwendet die Merkmale, die aus dem

Conditional Random

Field Ein ungerichtetes probabilistisches Modell, das auch benachbarte Stichproben für die Klassifizierung berücksichtigt, wird als Conditional Random Field (CRF) bezeichnet.

Convolution-Netz extrahiert wurden, um die Formen der Objektsegmentierung zu erzeugen. Die Unpooling- und Deconvolution-Layer werden verwendet, um anhand der Pixel Klassenlabels zu identifizieren und die Segmentierungsmasken vorherzusagen. Dies erzeugt eine Wahrscheinlichkeitszuordnung als Ausgabe, die die gleiche Größe wie das Eingabebild hat. Für jedes Pixel zeigt diese Wahrscheinlichkeitszuordnung die Wahrscheinlichkeit an, dass es zu einer der angegebenen Klassen gehört (Noh *et al*., 2015). Um die Label-Zuordnung zu verfeinern, ist es außerdem möglich, vollständig verbundene **CRFs** auf die Ausgabe des Netzes anzuwenden (Krähenbühl & Koltun, 2012).

###### Anwendungsfälle

Die semantische Bildsegmentierung kann in vielen Anwendungsfällen hilfreich sein:

* Autonomes Fahren: Erkennen von anderen Fahrzeugen, Fahrspuren, Fußgängern oder Gehwegen (Kaymak & Uçar, 2019).
* GeoSensing: Analyse von Informationen über die Landnutzung wie landwirtschaftliche Flächen, Wälder oder Wasserflächen aus Satellitenbildern (Pollatos *et al*., 2020).
* Posenschätzung/Bewegungserfassung: Identifizierung und Verfolgung von Körperteilen wie Beinen, Armen, Kopf oder Augen (Liu *et al*., 2013).
* Medizin: Erkennung der von Tumoren betroffenen Gehirnbereiche (Işın *et al*., 2016).

**Zusammenfassung**

Die Computer Vision ist ein interdisziplinäres Gebiet, das Methoden aus der Informatik, dem Ingenieurwesen und der Künstlichen Intelligenz kombiniert. Sie geht auf die 1960er Jahre zurück, als Forscherinnen und Forscher erstmals versuchten, das visuelle System des Menschen nachzuahmen. Typische Aufgaben in der Computer Vision befassen sich mit Themen wie Erkennungsaufgaben, Bildwiederherstellung, Bewegungsanalyse und Geometrierekonstruktion.

In der Computer Vision werden Bilder mithilfe von Pixeln dargestellt. Modelle wie das Brown-Conrady-Modell können verwendet werden, um die Verzeichnung digitaler Bilder zu beseitigen. Darüber hinaus ist es auch wichtig, die Kalibrierungsparameter einer Kamera zu kennen, um radiale und tangentiale Verzeichnungen zu vermeiden.

Algorithmen zur Merkmalserkennung in der Computer Vision können verwendet werden, um relevante Bildinhalte zu ermitteln. Nach der Merkmalserkennung werden die Merkmale in Merkmalsvektoren umgewandelt, die dann für den Merkmalsabgleich verwendet werden können.

Mit Methoden der semantischen Segmentierung können die Pixel eines Bildes in verschiedene Kategorien eingeteilt werden, um den Inhalt eines Bildes zu klassifizieren.