Heribert Wankerl, Maike L. Stern

Lernen durch Versuch und Irrtum: Bestärkendes Lernen

Wie lernen Menschen sinnvoll zu handeln?

In ihrem Leben machen Menschen gute und schlechte Erfahrungen. Im besten Fall lernen wir daraus, wie wir in bestimmten Situationen handeln müssen, um möglichst gute Erfahrungen zu sammeln und schlechte zu vermeiden. Auf diese Weise erwerben wir Verhaltensweisen, mit denen wir unsere Umwelt also zu unserem Vorteil beeinflussen können. Algorithmen des bestärkenden Lernens (engl. reinforcement learning, RL) implementieren eben diese Paradigmen aus Wahrnehmung, Manipulation und Bewertung der Umwelt.

Dafür wird, wie in Bild 1 angedeutet, ein sog. Agent in einer virtuellen Realität platziert, die er durch Aktionen beeinflussen kann. Durch dieses Eingreifen ändert sich der Status der Umwelt und die damit assoziierte Bewertung. Da wir möglichst gute Erfahrungen bevorzugen, sollten die Handlungen natürlich nicht willkürlich sein, sondern vielmehr die zu erwartende Bewertung maximieren. Dazu muss der Agent die Beziehung zwischen dem ursprünglichen Status der Umwelt, der getroffenen Aktion, dem anschließenden Status und dessen Bewertung verstehen, um seine Handlungsweisen anzupassen.

Der Agent lernt – ähnlich wie der Mensch – mit seinem Verhalten gute Erfahrungen langfristig wahrscheinlicher zu machen und schlechte zu vermeiden. Dafür kann es auch notwendig sein, kurzzeitig schlechte Erfahrungen (Lernen für eine Prüfung) hinzunehmen, wenn es nur so möglich ist, eine gute Bewertung (Prüfungsnote) zu erhalten. Solche strategischen Aspekte sind kennzeichnend für menschliches Handeln.

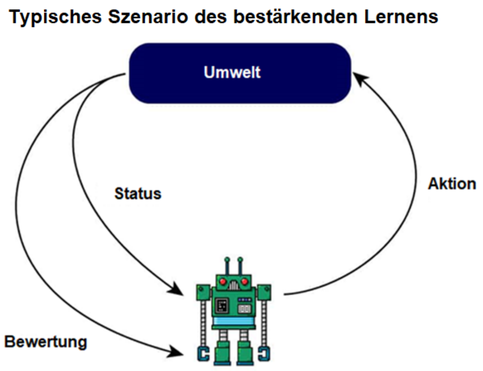

Bild 1: Typisches Szenario des bestärkenden Lernens: Ein Agent oder Mensch beobachtet den Status seiner Umwelt und kann diesen durch Aktionen beeinflussen. Jeder solcher Status ist mit einer Bewertung assoziert, die es dem Agenten ermöglicht, seine Handlungsweisen zu reflektieren.

Hier geht's zum Video

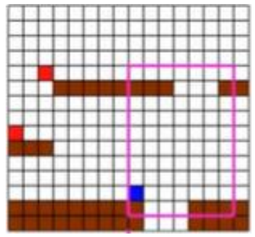

Bild 2: Ein Bildschirmausschnitt (Frame) aus dem Videospiel Super Mario Bros (links, [1, QR-code]) und seine maschinenlesbare Repräsentation (rosa 16x16-Feld, rechts), die der Agent wahrnimmt und basierend auf welcher er seine Handlungsentscheidungen (Aktion) trifft. Wenn er sich etwa dazu entscheidet, einfach in das Loch zu laufen, endet das Spiel mit einer schlechten Erfahrung. Die Verhaltensweise des Agenten wird entsprechend so angepasst, dass er künftig vermeidet, in Löcher zu fallen: Zum Beispiel durch Springen!

Bestärkendes Lernen und Kindheitserinnerungen

Erinnerst du dich noch an den Game Boy? – Vielleicht sogar an Super Mario?
Nicht erschrecken, aber dann warst du selbst damals ein Agent - eine Form von Intelligenz also - die Mario, wie in Bild 2 angedeutet, gesteuert hat! Basierend auf dem, was sich auf dem Bildschirm (Status) des Spiels (Umwelt) abgespielt hat, hast du gewisse Knöpfe (Aktionen) gedrückt.
Während du gespielt und so Erfahrungen gesammelt hast, haben sich deine Verhaltensweisen hoffentlich so angepasst, dass du möglichst lange überlebt und das Level geschafft hast: Du deine Aktionen also positiv bewertet hast!

Diese Verhaltensweisen kann durch bestärkendes Lernen auch ein Algorithmus erlernen. Im sog. deep reinforcement learning ist der Agent durch ein neuronales Netzwerk implementiert, das jeweils den Status der Umwelt verarbeitet und so entscheidet, welche Knöpfe gedrückt werden sollen.
Solche neuronalen Netze können ähnlich wie der Mensch durch gemachte Erfahrungen darauf trainiert werden, Peach vor Bowser zu retten. Dazu muss der Algorithmus die Zusammenhänge zwischen Status und Aktionen verstehen, um die Bewertung seiner Erfahrungen zu maximieren.

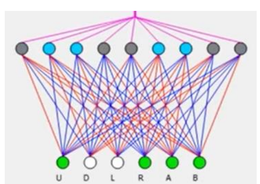

Fig. 3: Mögliche Aktionen in der Umwelt des Videospiels Super Mario Bros: Up, down, left, right, A und B. Welche Aktion in einem gegebenen Status getroffen wird, entscheidet hier ein neuronales Netz; sozusagen das Gehirn des Agenten. In diesem Fall drückt der Agent up, right, A und B gleichzeitig: Sollte also reichen, um über das Loch in Bild 1 zu kommen! Für die Retro-zocker unter uns: Der Agent hat scheinbar noch nicht gelernt, das Up und B hier keinen Einfluss auf den Status haben.

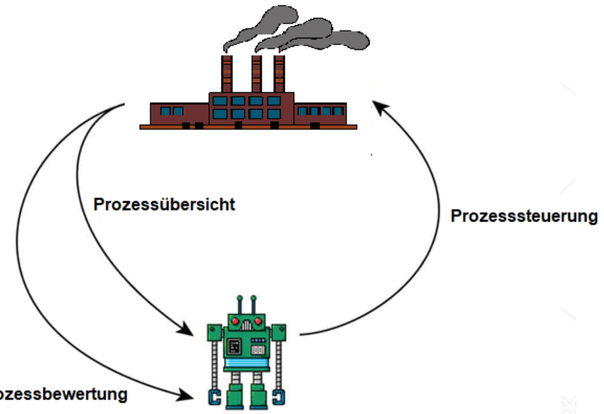

Bild 3: Ein Agent steuert eine Fabrik – oder zumindest Teilprozesse davon. Einem (vorerst) unerfahrenen Agenten direkt die Steuerung über eine reale Fabrik zu überlassen, erscheint dann aber doch etwas riskant. Deshalb werden die entsprechenden Teilprozesse in einer Simulation virtuell nachgebildet. Diese Simulationen dienen dann als Umwelt – ähnlich einem Videospiel oder Flugsimulatoren in der Ausbildung von Linienpilotinnen und -piloten. Hat der Agent dort ausreichend Erfahrungen gemacht ohne die Fabrik in die Luft zu jagen, kann er nach und nach in die reale Welt überführt werden.

Und wie kommt das jetzt in die Industrie?

Super Mario ist natürlich ein unbestrittener Klassiker. Keine Frage!
Trotzdem ist man in der Industrie eher weniger daran interessiert, Mario durch verschiedene Level zu navigieren. Aber man kann sich die erwähnten Methoden durchaus zu Nutze machen, um in der Industrie auftretende Prozesse etwa in Produktionsanlagen zu steuern.

Wie in Bild 3 angedeutet, muss der betreffende Prozess dazu simuliert werden. Durch virtuelle Steuereingriffe ändert der Agent dann die Einstellungen des Prozesses und beobachtet, wie sich die Abläufe in der Fabrik ändern. Sinkt durch seine Aktionen beispielsweise die Produktivität der Anlagen, wird er für sein Handeln bestraft. Schafft er es dagegen, die Produktivität zu erhöhen oder die Qualität der gefertigten Produkte zu verbessern, macht er dadurch positive Erfahrungen.
Wie auch der Mensch lernt der Agent sowohl durch gute als auch durch schlechte Erfahrungen, wie er sich in bestimmten Situationen verhalten muss.

Hat der Agent eine vielversprechende Verhaltensweise erlernt, kann diese verwendet werden, um Abläufe in der realen Fabrik zu optimieren oder besser zu verstehen.
Wie bestärkendes Lernen zudem in der physikalischen Forschung eingesetzt werden kann, erfährst du auf dem nächsten Poster!

References

[1] AI Learns to Play Super Mario Bros!
https://www.youtube.com/watch?v=CI3FRsSAa_U

[2] Künstliche Intelligenz in der Halbleiterfertigung von OSRAM
https://www.osram.de/cb/referenzen/podcast/episode-7-kuenstliche-intelligenz-in-der-halbleiterfertigung-von-osram/index.jsp

Die Strahlkraft künstlicher Intelligenz in der Ingenieurswissenschaft

Heribert Wankerl, Maike L. Stern

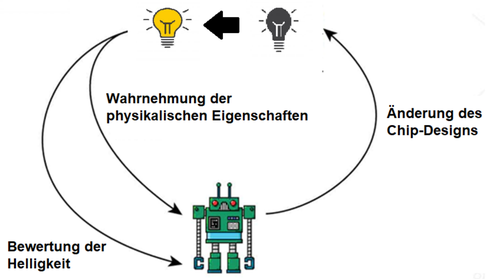

Bild 1: Ein Agent kann das Chip-Design einer LED verändern. Dadurch wird die LED hoffentlich heller. Der Agent nimmt also neben einigen anderen veränderten physikalischen Eigenschaften auch diese Helligkeitszunahme war und wird in seinem Verhalten bestärkt. Auf diese Weise erkennt der Algorithmus nach und nach wesentliche Aspekte helligkeitssteigernder Aktionen kennen und erstellt basierend auf diesem Wissen immer bessere LEDs.

Künstliche Intelligenz als Ingenieurin?

In der Halbleiterindustrie werden Agenten darauf trainiert, Prozesse in Fabriken zu kontrollieren, logistische Probleme zu lösen oder optische Bauteile wie LEDs zu optimieren.
Im letzten Fall unterstützen sie also Ingenieurinnen und Ingenieure immer hellere, langlebigere und umweltfreundlicher LEDs zu entwickeln (Bild 1).

Als optische Bauteile können LEDs eine unvorstellbar hohe Anzahl unterschiedlicher Chip-Designs aufweisen, die sich in Variationen von Schichtdicken oder Materialzusammensetzungen (Bild 2) äußern. Jedes dieser Designs führt zu gewissen physikalischen Eigenschaften (Bild 3) des Bauteils, die durch Simulationen teilweise vorhergesagt werden können. Wollen wir nun etwa eine Helligkeitssteigerung oder andere anwendungsbezogene Charakteristika (Bild 3) erreichen, würde ein naives Ausprobieren (Simulieren) aller denkbaren Designs zum Auffinden der besten Lösung etwa fünf Milliarden Mal so lange dauern, wie unser Universum alt ist. Hier kann Reinforcement Learning, ein Teilbereich des maschinellen Lernens, Abhilfe schaffen: Es erlaubt nämlich, diese gewaltige Anzahl von Designs zielgerichtet nach Optima zu durchsuchen.

Bild 2: Vereinfachtes Chip-Design bestehend aus acht Schichten, die aus verschiedenen, farblich unterschiedenen (orange, grau, blau) Materialien bestehen. Zusätzlich ist jede Schicht durch eine Dicke charakterisiert.

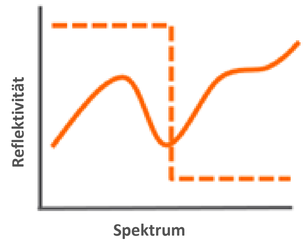

Bild 3: Ein Chip-Design wie in Bild 2 dargestellt besitzt bestimmte optische Eigenschaften wie etwa die Reflektivität über Spektrum. Diese gibt an, welcher spektrale Anteil eines einfallenden Lichtstrahls vom Chip-Design reflektiert wird. Für das Chip-Design in Bild 2 entspricht dies der durchgezogenen Linie. Für unsere Anwendung wünschen wir uns die gestrichelte Linie. Ziel des Agenten wäre es hier also, durch gezielte, sukzessive Änderungen am Chip-Design, möglichst eine Übereinstimmung der beiden Linien zu erreichen.

Eine strahlende(re) Zukunft dank KI?

Um die Charakteristika einer LED zu optimieren, nimmt der Algorithmus sukzessive Änderungen am bestehenden Aufbau eines Chip-Designs vor (Siehe Bild 4). Bildlich kann man sich hier eine Wissenschaftlerin vorstellen, die das Design einer LED durch Eingriffe manipuliert, indem sie beispielsweise Dicke und Material der nächsten Schicht vorgibt. Verbessern sich die physikalischen Charakteristika der LED durch die Veränderung, erhält der Algorithmus (die Wissenschaftlerin) Zuspruch und wird in seinem (ihrem) Verhalten bestärkt. Werden hingegen Verschlechterungen beobachtet, führt dies zur Vermeidung entsprechender Designs und damit assoziierter Designänderungen. Mit der Bewertung des Handelns verhält es sich konzeptionell also wie mit Schulnoten. Der Algorithmus erhält demnach bessere Noten, wenn durch sein Handeln gewünschte Charakteristika erreicht werden. Dem menschlichen Streben nach Belohnung folgend entwickelt der Algorithmus nach und nach Designs, die zu möglichst guten Bewertungen führen. Dabei nutzt er einerseits Erfahrungen aus der Vergangenheit, versucht andererseits jedoch auch neuartige LED-Designs zu entdecken – und entwickelt dabei zumeist bessere Designs, als das mit herkömmlichen Methoden möglich gewesen wäre.

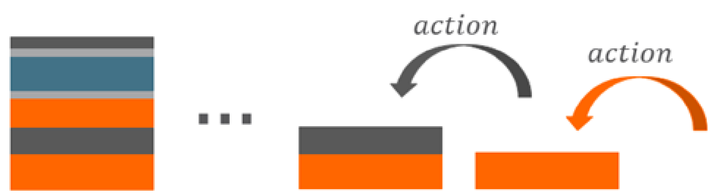

Bild 4: Was bedeutet hier eine Änderung des Chip-Designs? Was sind hier die Aktionen des Agenten? - In Anlehnung an Bild 2 enstpricht eine Änderung am Chip-Design der Aufbringung einer zusätzlichen Schicht. Der Agent entscheidet also, welches Material von welcher Dicke auf ein bestehendes Design aufgetragen warden soll. Er beginnt dabei auf einer leeren Werkbank und kann sich jederzeit auch dazu entscheiden, keine weitere Schicht aufzutragen.

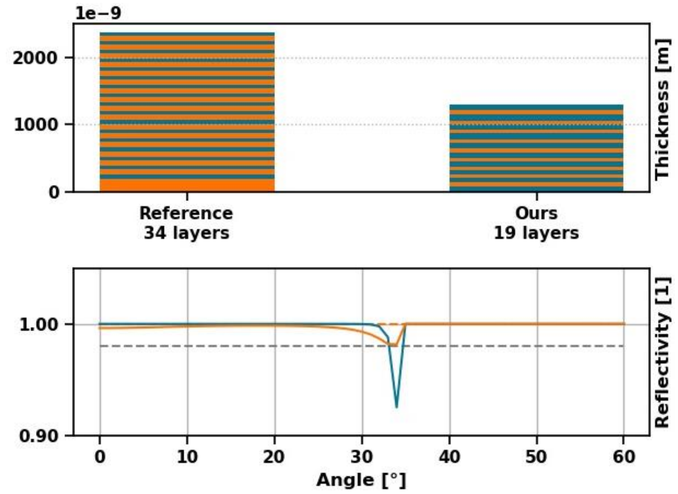

Bild 5: Vergleich eines Chip-Designs unserer Ingenieurinnen und Ingenieure (oben links) mit dem Chip, den unserer RL-Algorithmus gefunden hat (oben rechts). Letzterer weist nicht nur eine deutlich geringerer Gesamtdicke auf, sondern zeichnet sich auch durch eine kleinere Anzahl an gestapelten Schichten aus: Beides bietet prozesstechnisch einige Vorteile. Gleichzeitig erfüllt der RL-Algorithmus die Kundenspezifikationen im Bereich zwischen 30° und 35° besser. Die Farben der Linien werden im einzelnen im Text unten erklärt.

Ist das bereits die Realität?

Wir vergleichen ein momentan verwendetes Referenzdesign (Bild 5, oben links) bestehend aus Schichten verschiedener Dicken und zwei Materialien (blau, orange) mit einem weiteren, das unsere RL-Wissenschaftlerin (Bild 5, oben rechts) gefunden hat. Es zeigt sich, dass wir nicht nur die Gesamtdicke von weit über 2000 nm auf unter 1300 nm reduzieren konnten, sondern auch mit 19 statt 34 Schichten auskommen. Dies hat nicht nur positive Auswirkungen auf die Fertigungsstabilität der Chips. So zeigt die untere Abbildung, dass unsere Lösung (RL, orange Linie) nicht nur die Kundenanforderung (graue gestrichelte Linie) für alle betrachteten Winkel erreicht, sondern dass wir uns verglichen mit dem Referenzdesign (Reference, blaue Linie) gerade im Bereich zwischen 30° und 35° deutlich verbessern konnten. Als Ziel wurde vorab eine Reflektivität von 1.0 für alle Winkel (Target, orange gestrichelte Linie) definiert. Der jeweils gegebene Reward für die Wissenschaftlerin richtet sich danach, wie weit die simulierten Reflektivitätskurven ihrer Designvorschläge von der Zielkurve abweichen.

Trotzdem bleiben nach wie vor die Ingenieurinnen und Ingenieure oft treibende Innovationskraft, da eine KI bisher oft nur einzelne Teilprobleme effizient lösen kann und in Sachen genereller Kreativität und Problemlösefähigkeit dem Menschen bisher unterlegen bleibt.
Für besonderes Neugierige sei auf unser Paper [1, QR-code] oder Video hier im Showroom verwiesen.

Zum Paper

Zum Video