Philosophisches Institut, RWTH Aachen

Hausarbeit im Rahmen des Seminars
Klassiker der Wissenschaftstheorie
im SS 2015

Dozent: Dr. Joachim Bromand

Lerntheorie und Induktion

Michael Krause (331069)
Studiengang: Informatik B.Sc.
michael.krause@rwth-aachen.de

Aachen, 07.08.2015

Inhaltsverzeichnis

1 Einleitung **1**

2 Grundlegende Ideen bei Putnam **2**

2.1 Carnaps Grad der Bestätigung und Putnams Lernmaschinen . . 3

2.2 Das Ziel induktiver Logik bei Putnam 4

3 Anwendungen der Lerntheorie **5**

3.1 Das „Rabenproblem“ 6

3.2 Hypothesen mit Ausnahmen 7

3.3 Das grue-Rätsel 7

3.4 Weitere Maßstäbe 8

4 Aus der Informatik: Mustersprachen **9**

4.1 Golds Lernmodell und Angluins Mustersprachen 9

4.2 Mustersprachen lernen 10

5 Ockhams Rasiermesser **11**

6 Schwächen der Lerntheorie **12**

7 Schluss **13**

Literaturverzeichnis **14**

1 Einleitung

Wie bringt man Computern das Lernen bei? Mit dieser Frage beschäftigt sich eine Disziplin der Informatik namens Machine Learning. „Lernen“ kann sich dabei auf vieles beziehen: von menschlicher Sprache über Muster in Bildern bis hin zu Brettspielen. Die algorithmische *Lerntheorie* wiederum ist ein Teilgebiet des Machine Learning und erforscht Grundlagenfragen wie etwa: Wie modelliert man einen Lernprozess? Welche Arten von Konzepten sind überhaupt lernbar? Wie müssen lernende Computerprogramme aussehen und wie lernt ein Programm in möglichst kurzer Zeit?

Die algorithmische Lerntheorie untersucht das Lernen innerhalb von abstrakten formalen Systemen. Insbesondere formale Sprachen spielen dabei eine Rolle (siehe auch Abschnitt 4). Die Methoden und Resultate dieses Gebiets haben allerdings auch Anwendungen in der Wissenschaftstheorie. Eine Reihe von Philosophen beschäftigt sich im Rahmen der *formellen Lerntheorie* mit diesen Verbindungen (vgl. Schulte 2014).

Gewissermaßen lässt sich auch die Arbeit eines Wissenschaftlers, der aus einer Reihe von Beobachtungen Hypothesen für zukünftige Ereignisse ableiten möchte, als Lernen auffassen. Dabei zieht der Wissenschaftler aus seinen Beobachtungen einen *induktiven Schluss*. Im Gegensatz zu deduktiven Schlüssen, bei denen Wahrheit der Annahmen auch Wahrheit der Schlussfolgerung garantiert (Beispiel: „Alle Menschen sind sterblich, also ist Sokrates sterblich“), gehen induktive Schlüsse von Annahmen aus, welche die Schlussfolgerung nur wahrscheinlich machen (Beispiel: „Es hat letzte Woche nur geregnet, also wird es auch heute regnen“).

David Hume warf die Frage auf, wie wir induktive Schlüsse begründen (Hume 1986): Aus Gewohnheit scheint es uns zwar so, als ob zwei Vorgänge, die wir immer wieder nacheinander beobachtet haben, einander bedingten. Tatsächlich können wir uns dessen aber nicht sicher sein: Das Gegenteil ist durchaus vorstellbar, so etwa, dass nach einer ganzen Woche Regenwetter wider erwarten die Sonne scheint.

Eine weitere Schwierigkeit bei der Rechtfertigung induktiver Schlüsse entdeckte Nelson Goodman (Goodman 1983): Sein *neues Rätsel der Induktion* fragt danach, wie wir zwischen Hypothesen entscheiden können, die zwar mit bisherigen Beobachtungen übereinstimmen aber völlig verschiedene Vorhersagen für die Zukunft treffen. Sein Beispiel: Auch wenn wir bisher nur grüne

Smaragde beobachtet haben, gibt uns dies keinen Anhaltspunkt, die Hypothese „Alle Smaragde sind grün“ gegenüber der Hypothese „Alle Smaragde sind grue“ zu bevorzugen (wobei ein Objekt grue ist genau dann, wenn es grün vor einem Zeitpunkt t in der Zukunft oder blau nach t ist).

Hume und Goodman lassen uns also an der Rechtmäßigkeit induktiver Schlüsse zweifeln. Dennoch sind solche Schlüsse Grundlage beinahe jeder wissenschaftlicher Arbeit: ein Ornithologe wird aus der Tatsache, dass alle beobachteten Raben schwarz waren, kaum schließen, dass der nächste Rabe grün würde. Und ein Physiker geht fest davon aus, dass die Gesetzmäßigkeiten, die er aufstellt, immer gelten, egal welcher Wochentag auch sei. Die praktische Relevanz induktiver Schlüsse macht daher eine nähere Untersuchung wünschenswert.

Versucht man, das Lernen zu verstehen und zu modellieren, ergeben sich ganz ähnliche Fragestellungen wie die klassischen Probleme der Wissenschaftstheorie. Denn im Grunde sind induktive Schlüsse solche, bei denen wir aus vergangenen Beobachtungen lernen. Die formelle Lerntheorie formuliert Ansätze, um Antworten auf die aufgeworfenen Fragen zu finden. Grundsätzlich werden dabei Induktionsmethoden (bzw: Lernmethoden) nach ihrer Aussicht auf Erfolg beurteilt: ist es mit einer Methode möglich, langfristig zu einer wahren Hypothese über Beobachtungen zu gelangen (bzw: ein Konzept richtig zu lernen), so ist sie akzeptabel. Darüberhinaus bieten sich weitere Kriterien an, z.B. können Methoden bevorzugt werden, die möglichst selten ihre Hypothesen wechseln.

Ziel dieser Hausarbeit ist es, die für die Wissenschaftstheorie relevanten Grundkonzepte der Lerntheorie auszuarbeiten und an Beispielen zu veranschaulichen. Dazu werden zunächst zwei Aufsätze von Hilary Putnam, der zwar den Begriff „Lerntheorie“ nicht verwendete aber viele ihrer Ideen einführte, und anschließend Beispiele aus Informatik und Philosophie betrachtet. Schließlich werden noch Begründungen für Ockhams Rasiermesser im Rahmen der Lerntheorie sowie Schwächen des Ansatzes besprochen.

2 Grundlegende Ideen bei Putnam

Den Grundstein für die formelle Lerntheorie und ihre Beschäftigung mit Induktion legte Hilary Putnam in zwei Aufsätzen zur induktiven Logik von Rudolf Carnap (Putnam 1963 und Putnam 1979). In diesen zeigt er, dass die von Carnap vorgeschlagenen Bewertungsfunktionen für Hypothesen in gewissem Sinne unzureichend für eine erfolgreiche induktive Logik sind. Dieses Argument

soll hier kurz dargelegt werden, da es viele zentrale Ideen der Lerntheorie enthält.

2.1 Carnaps Grad der Bestätigung und Putnams Lernmaschinen

Putnams Kritik an Carnap dreht sich um dessen Konzept des *Grads der Bestätigung*, den eine Hypothese durch gegebene Daten erhält (Putnam 1963, S. 767). Für einen gegebenen Satz S (die Schlussfolgerung) und eine Menge an Sätzen E (die Daten) ist der Grad der Bestätigung von S unter E das Ausmaß, in dem S durch E gestützt wird. Dieser Grad ist eine reelle Zahl zwischen 0 und 1, größere Zahlen bedeuten ein höheres Maß an Bestätigung. Als Analogie kann man sich E als eine Reihe von Beobachtungen und S als eine Hypothese vorstellen, welche die Beobachtungen erklären soll.

In seinen Arbeiten vergleicht Putnam Induktion bildlich mit der Arbeit von *Lernmaschinen* (Putnam 1979, S. 297). Solche Maschinen leiten aus Beobachtungen Regelmäßigkeiten ab und treffen damit Vorhersagen für zukünftige Ereignisse. Nach Putnam ist das Ziel induktiver Logik die Untersuchung und Verbesserung von Strategien für solche Lernmaschinen. Die zentrale Anforderung, die Putnam an Lernmaschinen stellt, ist *Zuverlässigkeit*: Die Maschine muss nach irgendeiner Zahl von Beobachtungen zur korrekten Hypothese gelangen und diese danach nicht mehr fallenlassen. Stellt man diese Anforderung nicht, so wird man mithilfe der Lernmaschine die Wahrheit gewisser Hypothesen niemals herausfinden können (Putnam 1963, S. 763,766).

Angenommen nun, Carnaps Grad der Bestätigung ließe sich vollständig für alle Sätze unserer Sprache definieren (für Putnams Argument reicht eine Sprache aus, welche Anordnungen von Objekten ausdrücken kann, siehe ebd., S. 763). Für beliebige Hypothesen S und Daten E der Sprache ließe sich also eine *Bewertungsfunktion* $c(S, E) \in [0, 1]$ aufstellen, die genau diesen Grad der Bestätigung berechnet. Damit wäre implizit die Strategie einer optimalen Lernmaschine festgelegt, die bei gegebenen Beobachtungen E immer diejenige Hypothese annimmt, die durch E den höchsten Grad an Bestätigung erhält (Putnam 1979, S. 298). Bei jeder Datenlage wüssten wir also, welche Hypothese am wahrscheinlichsten wird - eine vollständige Beschreibung des induktiven Schließens.

Putnam zeigt, dass es eine solche optimale Lernmaschine nicht geben kann (Putnam 1979, S. 299, 1963, S. 769). Damit ist auch Carnaps ursprünglicher Ansatz, im Rahmen einer induktiven Logik Bewertungsfunktionen für jede Art von Hypothesen und Beobachtungen aufzustellen, nicht durchführbar. Putnams Argument ist, dass durch die Strategie einer optimalen Lernmaschine eine Beobachtungsfolge festgelegt wird, welche diese Strategie scheitern lässt.

Deutlich wird das an Putnams Beispiel. Eine optimale Lernmaschine beobachte eine Reihe von schwarzen und roten Kugeln, die nach einem ihr unbekannten System aus einer Urne gezogen werden. Ziel der Maschine ist, die Farbe der nächsten Kugel vorherzusagen. Beobachtet sie nun eine sehr lange Folge von roten Kugel, wird die Maschine ab irgendeinem Punkt die Vorhersage treffen, dass beim nächsten Mal (nennen wir diesen Zeitpunkt n_1) eine rote Kugel gezogen wird. Denn täte sie dies nicht, würde sie bereits auf einem so simplem Beispiel wie einer rein roten Folge von Kugeln das System nicht erkennen können - und wäre demnach keine optimale Lernmaschine.

Als solche muss sie jedoch bei jeder Beobachtungsfolge ab irgendeinem Punkt korrekte Voraussagen treffen. Zum Beispiel auch dann, wenn zum Zeitpunkt n_1 eine schwarze Kugel auftaucht. Problematisch ist nicht, dass die Vorhersage einer roten Kugel zum Zeitpunk n_1 falsch war. Das Problem liegt vielmehr darin, dass auf die selbe Weise die Maschine immer wieder falsche Hypothesen generieren wird:

Angenommen, nach n_1 tauchen wiederum nur rote Kugeln auf. Dann wird es einen Zeitpunkt n_2 geben, für den die Maschine eine rote Kugel voraussagen wird. Sei nun jedoch die Kugel zum Zeitpunkt n_2 schwarz. Dann lag die Maschine wieder mit ihrer Vorhersage falsch. Das Prinzip lässt sich klarerweise für alle n_i wiederholen. Die „optimale“ Lernmaschine wäre in diesem Fall also nicht zuverlässig, da es eine Beobachtungsfolge gibt (nur rote Kugeln bis auf schwarze Kugeln zu den Zeitpunkten $n_1, n_2, n_3, \ldots$), auf der sie unendlich viele Fehler macht und niemals zu einem Punkt gelangt, an dem sie nur noch wahre Hypothesen ausgibt.

2.2 Das Ziel induktiver Logik bei Putnam

Eine optimale Lernmaschine kann es demnach nicht geben. Für Putnam bedeutet dies aber keineswegs, dass induktive Logik ein hoffnungsloses Unterfangen sei. Im Gegensatz zu Carnap, der seine Bewertungsfunktionen über alle in

einer Sprache ausdrückbaren Hypothesen zu definieren versucht, setzt Putnam der induktiven Logik ein anderes Ziel: Diese soll Methoden bereitstellen, um aus einer *gegebenen* Menge an Hypothesen immer so auszuwählen, dass wahre Hypothesen zuverlässig gefunden und falsche Hypothesen langfristig verworfen werden.

Steht die Menge der in Frage kommenden Hypothesen vorab fest, so lässt sich das obige Argument nicht mehr anwenden (Putnam 1963, S. 773): Sind die zur Verfügung stehenden Hypothesen zum Beispiel „Alle Kugeln sind schwarz“ und „Alle Kugeln sind rot“, so fände eine Lernmaschine auf Basis dieser beiden Hypothesen die korrekte heraus, falls tatsächlich eine Folge gleichfarbiger Kugeln betrachtet wird. Sobald jedoch der erste Farbwechsel beobachtet würde, bräche die Maschine mit der Erkenntnis ab, dass keine der Hypothesen wahr sei.

Insgesamt setzt Putnam folgende Regeln fest, nach denen Induktion funktionieren soll: Aus einer gegebenen Menge an Hypothesen werden nach einer bestimmten Strategie immer nur solche ausgewählt, die mit den bisherigen Beobachtungen übereinstimmen. Die Untersuchung solcher Strategien - und damit der Konstruktion immer besserer Lernmaschinen - ist Ziel der induktiven Logik. Ein entscheidendes Kriterium für gute Strategien ist, dass sie zuverlässig entweder die wahre Hypothese herausfinden oder - wenn keine der gegebenen Hypothesen wahr ist - die gesamte Hypothesenmenge ablehnen.

3 Anwendungen der Lerntheorie

Putnam legt in seinen Arbeiten die grundlegenden Maßstäbe fest, nach denen Induktion in der Lerntheorie betrachtet wird. Insbesondere kann es nicht mehr Ziel sein, eine universelle Methode zur Prüfung induktiver Hypothesen entwickeln zu wollen - damit wäre man wieder bei Carnaps Bewertungsfunktionen angelangt. Stattdessen müssen Strategien zur Auswahl von Hypothesen aus einer zur Verfügung stehenden Hypothesenmenge untersucht werden. Laut Putnam sollen Induktionsstrategien zuverlässig sein, das heißt: egal, welche Beobachtungen tatsächlich eintreten werden, führen die Strategien zur korrekten Hypothese und verwerfen alle falschen. Dieses Prinzip soll nun anhand von einfachen Beispielen verdeutlicht werden. Die Ergebnisse in diesem Abschnitt stammen - leicht abgewandelt - von Schulte (2014).

3.1 Das „Rabenproblem“

Das erste Problem, das hier betrachtet werden soll, ist die Frage: „Sind alle Raben schwarz?“ Die relevanten Beobachtungen, um dies bejahen oder verneinen zu können, sind natürlich Beobachtungen von schwarzen oder nicht-schwarzen Raben. Um die Frage zu beantworten könnten wir eine feste Anzahl von Beobachtungen festlegen, ab der nur noch von schwarzen Raben ausgegangen wird. Doch so eine feste Grenze wäre offensichtlich willkürlich. Da es immer denkbar ist, dass wir doch noch auf einen roten Raben treffen, können wir überhaupt nicht feststellen, ob denn nun alle Raben schwarz sind.

Der lerntheoretische Ansatz ist ein anderer. Die Frage lautet weniger „Sind alle Raben schwarz?“, als: „Welche Strategie, nach der wir unsere Hypothesen über die Farben von Raben auswählen, wird zuverlässig zur Wahrheit führen?“ Das bedeutet: Wenn wirklich alle Raben schwarz sind soll gefolgert werden: alle Raben sind schwarz (I). Gibt es jedoch einen nicht-schwarzen Raben soll gefolgert werden: nicht alle Raben sind schwarz (II).

Denkbar wäre nun folgendes Vorgehen - man könnte es die Hume-Strategie nennen: Wird ein nicht-schwarzer Rabe beobachtet, so folgern wir (II). Wird jedoch ein schwarzer Rabe beobachtet, so folgern wir daraus nichts. Denn nach Hume kann ein einzelner schwarzer Rabe kein entscheidender Indikator für die Frage „Sind alle Raben schwarz?“ sein. Für den ersten Fall führt diese Strategie zuverlässig zum Ziel: Sind nicht alle Raben schwarz, so wird auch die Hume-Strategie dies erkennen. Für den zweiten Fall jedoch nicht: Wenn tatsächlich alle Raben schwarz sind, wird die Hume Strategie dies niemals folgern. Damit ist diese Induktionsstrategie nicht zuverlässig.

Eine zuverlässige Induktionsstrategie ist hingegen diejenige, die bereits nach dem ersten schwarzen Raben (I) folgert. Denn ist (I) korrekt, so führt die Strategie zuverlässig zur Wahrheit. Gibt es hingegen einen nicht-schwarzen Raben, so wird dieser irgendwann beobachtet und die Strategie folgert (II). In beiden Fällen ist das Ergebnis die jeweils wahre Hypothese. Nach den hier angelegten Maßstäben ist also eine Induktionsstrategie, die mit geringer empirischer Grundlage bereits (I) folgert, gegenüber einer eher skeptischen Einstellung im Vorteil!

3.2 Hypothesen mit Ausnahmen

Wie sieht es mit komplizierteren Fragestellungen aus, etwa: „Sind alle *bis auf endlich viele* Raben schwarz?“ Eine analoge Strategie zur zuverlässigen Strategie aus dem vorherigen Beispiel würde nach dem ersten (und allen danach) beobachteten schwarzen Raben folgern: alle bis auf endlich viele Raben sind schwarz (I′). Diese Strategie wäre jedoch nicht zuverlässig. Denn angenommen, alle bis auf endlich viele Raben seien weiß, nicht schwarz. Dann bliebe diese Strategie bei ihrer ersten, falschen Folgerung, auch wenn es sich in Wirklichkeit bei allen beobachteten Raben um Teile der endlich vielen Ausnahmen handelte.

Schulte (2014) schlägt für dieses Szenario stattdessen folgende, zuverlässige Induktionsstrategie vor: Sind 50% der bisher beobachteten Raben schwarz, so folgere (I′), sonst: alle bis auf endlich viele Raben sind nicht-schwarz (II′). Angenommen nun, es sind tatsächlich alle bis auf endlich viele Raben schwarz. Dann werden - egal in welcher Reihenfolge die Beobachtungen stattfinden - irgendwann über 50% der beobachteten Raben schwarz sein, denn es gibt nur endlich viele Ausnahmen. Somit wird die Strategie zur richtigen Hypothese (I′) gelangen. Sind hingegen alle Raben weiß und nur endlich viele schwarz, gibt die Strategie aus dem analogen Grund irgendwann die korrekte Antwort (II′). Diese Induktionsstrategie ist also zuverlässig.

3.3 Das grue-Rätsel

Nun soll mit Hilfe des selben Ansatzes das neue Rätsel der Induktion nach Goodman betrachtet werden (siehe Einleitung). Angenommen, wir beobachten eine Reihe von Smaragden und alle sehen grün aus - dies stützt gleichsam die Hypothesen „Alle Smaragde sind grün“ wie auch „Alle Smaragde sind grue“. Aus beiden Möglichkeiten ergeben sich Induktionsstrategien. Die grün-Strategie folgert bei Beobachtung eines grünen Smaragds: „Alle Smaragde sind grün“ und erst bei Beobachtung eines blauen Smaragds zum Zeitpunkt t: „Alle Smaragde sind grue“. Die grue-Strategie hingegen folgert auch bei Beobachtung grüner Smaragde: „Alle Smaragde sind grue“.

Es ist klar einzusehen, dass die grün-Strategie zuverlässig ist. Sie wird irgendwann das korrekte Prädikat - grün oder grue - folgern. Die grue-Strategie hingegen ist nicht zuverlässig: sind tatsächlich alle Smaragde grün, so wird die grue-Strategie dies niemals herausfinden. Aus lerntheoretischer Sicht ist damit grün gegenüber grue im Vorteil. Die grün-Strategie arbeitet zuverlässig,

weiß jedoch nie, zu welchem Zeitpunkt die korrekte Hypothese ausgewählt wurde. Anders ausgedrückt: stehen als Hypothesen „Alle Smaragde sind grün“ und „Alle Smaragde sind grue mit t=‚Jahr 3001‘“ zur Verfügung müssen wir noch ein Jahrtausend warten, bis wir uns über die richtige Antwort sicher sein können.

3.4 Weitere Maßstäbe

Bisher wurden Induktionsstrategien nur im Hinblick auf ihre Zuverlässigkeit untersucht. Es kommen aber durchaus noch weitere Maßstäbe in Betracht. So ließe sich etwa beurteilen, wie oft eine Induktionsstrategie zwischen verschiedenen Hypothesen wechselt, bevor sie zur wahren Hypothese gelangt. Intuitiv erscheint eine zuverlässige Strategie umso besser, je seltener sie ihre Hypothesen ändern muss.

Kehren wir noch einmal zum grue-Rätsel zurück. Es stehe nun neben der grün- und der grue-Strategie von zuvor noch eine weitere zuverlässige Strategie zur Auswahl: Diese hält sich bis zur Beobachtung des 100en grünen Smaragds an die grue-Hypothese, folgert danach jedoch die grün-Hypothese (bis unter Umständen ein blauer Smaragd gesichtet und wieder zur grue-Hypothese gewechselt wird). Egal, ob nun alle Smaragde grün oder grue sind, diese Methode wird zuverlässig zur wahren Hypothese gelangen. Allerdings benötigt sie für ersteren Fall einen Hypothesenwechsel mehr als die grün-Strategie, die von Anfang an die richtige Hypothese folgert. Und falls zwar alle Smaragde grue sind, aber zunächst 101 grüne Smaragde beobachtet wurden, braucht diese Strategie ebenfalls einen Hypothesenwechsel mehr als die grün-Strategie. Wenn man als Maßstab die Zahl der Hypothesenwechsel einer Strategie anlegt, würde man daher die grün-Strategie bevorzugen.

Zusammengefasst verdeutlichen diese Beispiele einen Grundsatz der Lerntheorie: Weniger wichtig als kurzfristig korrekte Überzeugungen ist für Lerntheoretiker eine Methode, die langfristig richtige Hypothesen liefert. Weiterhin können zusätzliche Maßstäbe wie etwa die Anzahl der Hypothesenwechsel oder die Einfachheit der Hypothesen (siehe auch Abschnitt 5) angelegt werden.

4 Aus der Informatik: Mustersprachen

Das von E.M. Gold entwickelte Modell des *Lernens im Limit* (Gold 1967) hat viele Gemeinsamkeiten mit Putnams Ideen für eine induktive Logik und spielte darüber hinaus in der Entwicklung der Lerntheorie in der Informatik eine bedeutende Rolle. Hier soll das Lernmodell von Gold (genauer gesagt: eines der von ihm vorgeschlagenen) kurz skizziert werden. Ziel ist es, mit Hilfe des Modells ein anschauliches Beispiel für praktische Induktionsprobleme und deren Lösung zu bieten.

4.1 Golds Lernmodell und Angluins Mustersprachen

Golds Modell beschäftigt sich mit *formalen Sprachen*. Eine formale Sprache ist eine Menge von Zeichenketten (= *Wörtern*), die aus Buchstaben bestehen. Die Menge aller möglichen Buchstaben wird *Alphabet* genannt. Ein Beispiel für ein Alphabet wäre $\{0, 1\}$. Mögliche Worte bestehend aus Buchstaben dieses Alphabets wären $0, 1, 01, 10, 101, 0010$ und so weiter. Ein Beispiel für eine formale Sprache über diesem Alphabet wäre die Sprache aller Worte mit einer geraden Anzahl an 0en (dazu gehören $1, 11, 001, 01011, 0100110, \ldots$). Die deutsche Sprache ließe sich als Menge aller grammatikalisch korrekter Sätze über dem deutschen Alphabet auffassen. In der Regel haben die zu untersuchenden formalen Sprachen unendlich viele Elemente.

Gold modelliert den Vorgang des Lernens folgendermaßen: Ein *Lerner* (ähnlich den Lernmaschinen bei Putnam) wird nacheinander mit Wörtern aus einer formalen Sprache konfrontiert und soll herausfinden, um welche Sprache es sich handelt. Die Elemente kommen ungeordnet und mehrfach (aber jedes mindestens einmal) vor. Immer wenn der Lerner ein weiteres Element erhält, gibt er eine Hypothese darüber aus, um welche Sprache es sich handelt.

Wenn durch die Strategie, mit der ein Lerner seine Hypothesen generiert, sichergestellt wird, dass ab irgendeinem Zeitpunkt die korrekte Hypothese genannt und danach nicht mehr verändert wird, so spricht Gold davon, dass der Lerner die Sprache *im Limit lernt*.

Diese Definition ähnelt stark der aus Putnams induktiver Logik. Statt anhand formaler Sprachen kann man den Vorgang auch so erklären: Ein Wissenschaftler wird mit immer neuen Vorkommnissen eines Phänomens konfrontiert und wendet eine Methode an, um Hypothesen über das Phänomen aufzustellen. Gelangt er irgendwann zur wahren Hypothese und gibt sie anschließend

auch nicht mehr auf, so könnte man sagen, er hätte das Phänomen erklärt. Die von ihm verwendete Methode hat ihn zur korrekten Induktion aus den Beobachtungen geführt.

Forscher, die sich in der Informatik mit Lerntheorie beschäftigen, untersuchen nun, für welche Arten von formalen Sprachen es Lernstrategien gemäß Golds Definition gibt. Ein interessantes und anschauliches Beispiel sind die sogenannten *Mustersprachen* nach Angluin (1980).

Ein *Muster* ist eine Zeichenkette bestehend aus Konstanten und Variablen (z.B. $11x0yx$, wobei die Zahlen Konstanten und die Buchstaben Variablen sind). Die Wörter einer Mustersprache entstehen aus einem Muster, indem die Variablen des Musters durch Zeichenketten aus Konstanten ersetzt werden. Aus obigem Muster entsteht zum Beispiel durch eine Ersetzung $x = 10$ und $y = 1$ das Wort 11100110. Die zu $11x0yx$ gehörige Mustersprache enthält also das Wort 11100110, sowie auch $110010, 1101000010, 1100110$ und viele weitere.

4.2 Mustersprachen lernen

Angenommen, ein Lerner wird mit den Wörtern einer ihm unbekannte Mustersprache konfrontiert. Welches ist das Muster, aus dem die Wörter generiert wurden? Für die Menge aller möglichen Mustersprachen ist also eine Strategie gesucht, die zu einer Menge von Eingabewörtern das passende Muster findet.

Für dieses spezielle Lernproblem gibt es tatsächlich Lösungsstrategien. Eine ist der Algorithmus von Lange und Wiehagen (1991). Der Algorithmus findet Muster für Eingabewörter beliebiger Länge. Der Einfachheit halber nehmen wir hier aber an, dass alle Wörter die selbe Länge n besitzen. Weiterhin bezeichne w_i das i-te Zeichen der Zeichenkette w. Dann lautet der Algorithmus:

Wann immer ein neues Wort w der Mustersprache gesehen wird, berechne aus w und der zuletzt ausgegebenen Hypothese h eine neue Hypothese h^+ zeichenweise für $i = 1, i = 2, \ldots, i = n$ durch folgende Vorschrift.

$$h_i^+ = \begin{cases} h_i & \text{wenn } h_i = w_i & (1) \\ x & \text{wenn } h_i \neq w_i \text{ und es eine Position } k < i \text{ gibt, wo} & (2) \\ & w_k = w_i \text{ und } h_k = h_i \text{ und } h_k^+ = x & \\ y & \text{sonst, wobei } y \text{ für eine neue Variable steht} & (3) \end{cases}$$

In Worten: wir durchlaufen das Wort w vom Anfang bis zum Ende. Wenn die Zeichen in h und w gleich sind, entsprechen sie denen in h^+. Gibt es

unterschiedliche Zeichen in h und w, werden sie in h^+ durch eine Variable ersetzt. Die selbe Variable wird für gleiche Kombinationen von Zeichen in h und w verwendet. Die entstehende Hypothese h^+ wird als Ausgangshypothese h für das nächste Wort w verwendet.

Ein Beispiel. Durch bisher gesehene Wörter seien wir zur Hypothese $h = 0xx011$ gelangt, daher: wir nehmen auf Grund der bisher gesehenen Wörter an, dass h das zu findende Muster ist. Nun erhalten wir ein neues Wort $w = 011100$ und passen damit unsere Hypothese an:

i	1	2	3	4	5	6
h_i	0	x	x	0	1	1
w_i	0	1	1	1	0	0
Fall	1	3	2	3	3	2
h_i^+	0	x_1	x_1	x_2	x_3	x_3

Nach Erhalt des Wortes w lautet die neue Hypothese also $h^+ = 0x_1x_1x_2x_3x_3$. Lange und Wiehagen zeigen, dass diese Strategie genügt, um jede beliebige Mustersprache im Limit zu lernen. Wie bei den Beispielen im vorigen Abschnitt sind dabei zwei Aspekte zu beachten: zum Einen wurden die möglichen Hypothesen zu Anfang klar eingegrenzt - nur gültige Muster sind Hypothesen. Zum Anderen ist dem Lerner hier zwar bekannt, dass er ab irgendeinem Zeitpunkt das korrekte Ergebnis berechnet - aber nicht, wann dieser Zeitpunkt ist.

5 Ockhams Rasiermesser

Ein Prinzip, das Wissenschaftler in ihren induktiven Schlüssen verfolgen können, ist *Ockhams Rasiermesser*: Wenn mehrere Hypothesen ein Phänomen adäquat erklären, wähle die einfachste (für weitere Formulierungen und deren historischen Ursprung siehe De Wolf 1997). Wie die Induktion verlangt auch dieses Prinzip nach einer Begründung, die über bloße Intuition hinausgeht.

Im Kontext der Lerntheorie lässt sich das Prinzip als ein Baustein für Induktionsstrategien auffassen: Passen mehrere Hypothesen zu den bisherigen Beobachtungen, so wird die einfachste ausgewählt. Wieso aber die einfachste Hypothese wählen? Wieso nicht die komplizierteste, oder diejenige, die besonders spannend klingt?

Mit Modellen der Lerntheorie haben mehrere Autoren neue Begründungen für das Prinzip von Ockhams Rasiermesser gefunden. Im Abschnitt 3.4 wurde

bereits für ein Beispiel gezeigt, dass eine intuitiv einfachere Strategie (die sofort „Alle Smaragde sind Grün“ folgert) gegenüber einer komplexeren Strategie (die erst hundert mal „Alle Smaragde sind grue“ folgert) seltenere Hypothesenwechseln aufweist.

Ein Ansatz zur Begründung von Ockhams Rasiermesser ist daher, zu zeigen, dass mittels dieses Prinzips seltener zwischen Hypothesen gewechselt werden muss. Das Bevorzugen von Einfachheit macht Induktion also effizienter. Dies wurde für eine Reihe von praktischen Induktionsproblemen (z.B. der Auswahl neuer Erhaltungssätze in der Teilchenphysik) gezeigt (siehe Schulte 2014).

Ein anderer Ansatz bedient sich einem Modell aus der Informatik, welches PAC-Learning genannt wird. Im Gegensatz zum bisher Betrachteten handelt es sich hierbei um ein statistisches Modell. Gesucht sind nicht mehr Induktionsstrategien, die mit Sicherheit ein korrektes Ergebnis liefern werden, sondern stattdessen Methoden, mit denen auf Basis von Beobachtungen *ungefähr korrekte* Hypothesen aufgestellt werden. Ungefähr korrekt bedeutet dabei: mit immer weiteren Eingabebeobachtungen wird der PAC-Lerner immer bessere Hypothesen aufstellen, bis hin zu einer beliebig kleinen Fehlertoleranz.

De Wolf (1997) zeigt, dass eine Maschine, die aus einer Menge von Hypothesen immer die einfachste passende auswählt, auch ein PAC-Lerner für die Beobachtungen ist. Das bedeutet: wählt man immer die einfachsten Erklärungen, kommt man der Wahrheit beliebig nahe (mit steigender Anzahl an Beobachtungen). Einfachheit misst sich in seinem Beweis mit der Kolmogorov-Komplexität einer Hypothese. Dies ist die Länge des kürzesten Computerprogramms (genauer: Programms einer Turing-Maschine), welches diese Hypothese ausgibt (ebd., S. 50).

Für diesen Ansatz müssten sich die gemäß Kolmogorov-Komplexität einfachsten Hypothesen aus einer Hypothesenmenge herausfinden lassen. Da die Kolmogorov-Komplexität einer Hypothese nicht berechenbar ist, ist dies praktisch unmöglich. De Wolf zeigt also in erster Linie einen theoretischen Zusammenhang zwischen niedriger Kolmogorov-Komplexität und effizientem Lernen.

6 Schwächen der Lerntheorie

Mit den vorgestellten Grundsätzen lassen sich Induktionsstrategien für die Auswahl aus einer Hypothesenmenge analysieren und bewerten. Auf eine wichtige

Frage kann die Lerntheorie aber keine Antwort geben: Woher sollen die zur Auswahl stehenden Hypothesen stammen?

Bereits Putnam stellte dieses Problem fest und hebt die Wichtigkeit wissenschaftlicher Theorien hervor (Putnam 1963, S. 779). Aus diesen leiten sich Hypothesen ab, die für die Lösung eines Induktionsproblems benutzt werden können. Die Schwierigkeit, gute Theorien und Hypothesen zu finden, ist für ihn der Grund, weshalb in der Wissenschaft Genies gebraucht werden (ebd., S. 781).

Scott Aaronson bringt an, dass Lernmodelle unzureichend sind, um die revolutionären Erkenntnisse etwa Albert Einsteins erfassen zu können. Ein weiteres Ziel müsse daher sein, auch Modelle für die Vorhersage völlig neuer, bisher nicht beobachteter Phänomene zu entwickeln (Aaronson 2012, S. 289).

7 Schluss

In dieser Hausarbeit sollten die Grundideen der Lerntheorie und ihre Anwendung auf Induktionsprobleme dargestellt werden. Dazu wurden zunächst Hilary Putnams ursprüngliche Arbeiten betrachtet und ihre Ideen Anhand von Beispielen illustriert. Ein kurzer Abstecher in die Informatik, wo im Rahmen der algorithmischen Lerntheorie viele Erkenntnisse zu den theoretischen Grenzen des Lernens und der effizienten Lösung von Lernproblemen gewonnen werden, demonstrierte die praktische Anwendbarkeit des Lernmodells nach E. M. Gold. Schließlich wurden Arbeiten vorgestellt, in denen mit lerntheoretischen Modellen Begründungen für Ockhams Rasiermesser gefunden wurden.

Induktion und menschliches Lernen haben viele Gemeinsamkeiten. Wir haben für sie keine abschließenden Erklärungen, aber dennoch lernen Menschen immer wieder etwas Neues und stellen Wissenschaftler gewagte Vorhersagen auf. Eine präzise Erklärung wie zuverlässige Induktion funktioniert wäre deshalb äußerst interessant. Als Student der Informatik ist es sehr erfreulich zu sehen, dass Informatik und Wissenschaftstheorie in diesen Fragen voneinander profitieren können.

Literatur

Aaronson, Scott. „Why Philosophers Should Care About Computational Complexity“. In: *Computability: Turing, Gödel, Church, and Beyond*. Hrsg. von B. Jack Copeland, Carl J. Posy und Oron Shagrir. Cambridge MA–London: MIT Press, 2012, S. 261–328.

Angluin, Dana. „Finding patterns common to a set of strings“. In: *Journal of Computer and System Sciences* 21.1 (1980), S. 46–62.

De Wolf, R. M. *Philosophical applications of computational learning theory: Chomskyan innateness and Occam's razor*. Masterarbeit, Erasmus Universiteit Rotterdam, 1997.

Gold, E. Mark. „Language identification in the limit“. In: *Information and Control* 10.5 (1967), S. 447–474.

Goodman, Nelson. „The New Riddle of Induction“. In: *Fact, Fiction, and Forecast*. Cambridge MA: Harvard University Press, 1983, S. 72–81.

Hume, David. *Eine Untersuchung über den menschlichen Verstand*. Stuttgart: Philipp Reclam Jun., 1986, S. 41–77, 82–105.

Lange, Steffen und Rolf Wiehagen. „Polynomial-time inference of arbitrary pattern languages“. In: *New Generation Computing* 8.4 (1991), S. 361–370.

Putnam, Hilary. „“Degree of Confirmation” and Inductive Logic“. In: *The Philosophy of Rudolf Carnap*. Hrsg. von P.A. Schilpp. La Salle: Open Court, 1963, S. 761–783.

— „Probability and Confirmation“. In: *Mathematics, Matter and Method*. 2. Aufl. Cambridge: Cambridge University Press, 1979, S. 293–304.

Schulte, Oliver. „Formal Learning Theory“. In: *The Stanford Encyclopedia of Philosophy (Spring 2014 Edition)*. Hrsg. von Edward N. Zalta. `http://plato.stanford.edu/archives/spr2014/entries/learning-formal/`, 2014.

E r k l ä r u n g

Hiermit erkläre ich, dass ich die vorgelegte Arbeit selbständig verfasst und - einschließlich eventuell beigefügter Abbildungen und Skizzen - keine anderen als die im Literaturverzeichnis angegebenen Quellen, Darstellungen und Hilfsmittel benutzt habe. Dies gilt in gleicher Weise für gedruckte Quellen wie für Quellen aus dem Internet.

Ich habe alle Passagen und Sätze der Arbeit, die dem Wortlaut oder dem Sinne nach anderen Werken entnommen sind, in jedem einzelnen Fall unter genauer Angabe der Stelle ihrer Herkunft (Quelle, Seitenangabe bzw. entsprechende Spezifizierung) deutlich als Entlehnung gekennzeichnet.

Außerdem erkläre ich, dass die vorgelegte Arbeit zuvor weder von mir noch - soweit mir bekannt ist - von einer anderen Person an dieser oder einer anderen Universität eingereicht wurde.

Die elektronische Fassung der Hausarbeit ist mit der gedruckten Fassung identisch.

Mir ist bekannt, dass Zuwiderhandlungen gegen diese Erklärung eine Benotung der Arbeit mit der Note "nicht ausreichend" zur Folge haben. Ich weiß, dass Verletzungen des Urheberrechts sowie Betrugsversuche strafrechtlich verfolgt werden können und dass, wer vorsätzlich gegen eine die Täuschung betreffende Regelung verstößt, ordnungswidrig handelt. Die Ordnungswidrigkeit kann mit einer Geldbuße bis zu 50.000 Euro geahndet werden. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann außerdem eine Exmatrikulation erfolgen.

_____ _____

(Datum) (Unterschrift)