Information: Währung des Lebens

Sommerakademie Olang, September 2008

Professor Dr. Daniel Polani, Stefan Winter,
Richard Bormann, Paul Brodersen, Gregor Däubler, Felix Dollinger, Fynn Feldpausch, Alexander Fromm, Matthias Grey, Malte Harder, Lennart Hilbert, Frederik John, Matthias Linden, Georg Ostrovski, Saskia Pohl, Hanna Rademaker, Andreas Reiserer, Jan Schlüter, Malte Schomers, Malte Schwarzkopf, Jan Simon, Jan Hendrik Stockemer, Martin Tanz, Julien Wolf, Jingbo Wu

Zusammengetragen von Jan Simon

> *„Life may be defined operationally as an information processing system - a structural hierarchy of functioning units - that has acquired through evolution the ability to store and process the information necessary for its own accurate reproduction."*
>
> Lila Gatlin, 1972

Information ist einer der zentralen Begriffe unseres Lebens. Ihre Bedeutung ist fundamental und allgegenwärtig. Information ist wertvoll. Information ist teuer. Information ist gefährlich. Information ist Macht. Informationsverarbeitung, Informationsgesellschaft, Informationsministerium, Informationszeitalter!

Information ist alt. Lange bevor es Menschen gab, verstand die Natur es bereits, Information in Form des Erbmaterials zu speichern, zu reproduzieren und im Prozess der Evolution auch zu verändern. Dabei sind intelligent operierende Lebewesen entstanden, die wiederum hervorragend angepasst waren an die Aufnahme von Information aus ihrer Umgebung und deren Umrechnung in sinnvolle Handlungen. Schließlich gelang es Organismen auch, miteinander zu kommunizieren, also Information untereinander auszutauschen.

Es erscheint nahe liegend, dass, wer die evolutionäre Entstehung intelligenten Lebens verstehen und letztendlich auch nachbilden will, das Wesen der Information und ihre Gesetzmäßigkeiten verstehen muss.

Die moderne Informationstheorie entstand um 1949. Einer ihrer Vordenker war Claude Shannon. Er schlug ein quantitatives Maß für Information im stochastischen Kontext vor, die so genannte

Shannon-Information. Ist X eine diskrete Zufallsvariable, und x ein möglicher Wert, den X zufällig annehmen kann, so ist dessen Shannon-Information definiert als

$$\boxed{h(x) := \log_2 \frac{1}{P(x)}\,\mathrm{bit}}. \tag{1}$$

Hinter dieser Definition steckt anschaulich *die Anzahl maximal erforderlicher Fragen, um das Ereignis x bei optimaler Fragestrategie zu identifizieren*, wenn die Variable X diskret und gleichverteilt ist:

Beispielsweise möge X die 32 Spielkarten eines Skatspiels beschreiben. Die Information, die in der Auswahl einer einzelnen Karte steckt, ist nach Shannon dann durch 5 bit quantifiziert. Eine optimale Strategie, um diese eine Karte zu bestimmen besteht nämlich darin, die Möglichkeiten durch jede Frage zu halbieren. Wir können dann $\frac{1}{P(x)}$ als die Anzahl der möglichen Karten verstehen, und es sind $\log_2 32 = 5$ Fragen zur Identifikation erforderlich.

Eine naivere Strategie besteht darin, einzelne Karten abzufragen. Die Wahrscheinlichkeit, nicht auf Anhieb richtig getippt zu haben, ist jedoch sehr groß, nämlich $\frac{31}{32}$, sodass in diesem ungünstigen Fall die Information lediglich $\log_2 \frac{32}{21} \approx 0{,}0458\,\mathrm{bit}$ beträgt.

Es soll jedoch nicht verschwiegen werden, dass in der Praxis nicht immer eine optimale Fragestrategie möglich ist. Beispielsweise kann ein Tresor nur Fragen nach einzelnen Ziffernkombinationen beantworten. Die Shannon-Information berücksichtigt das nicht, eignet sich aber dennoch als Referenzwert.

Hinzu kommt, dass sie rein *objektiv* ist und keinen Aussage über die *subjektive* Bedeutung oder Relevanz der Information für Organismen trifft. Zwei Zeichenfolgen gleicher Länge haben den gleichen Shannon-Informationsgehalt, unabhängig von der tatsächlichen Wahl der Buchstaben. Die folgenden Zeichenketten haben alle dieselbe Shannon-Information von $\log_2 \frac{1}{\left(\frac{1}{26}\right)^{11}} \approx 51{,}7\,\mathrm{bit}$ (, bei 26 gleichwahrscheinlichen Zeichen).

INFORMATION INFORRMATOM NIOFMRTAONI MEEGEZURWMI

Als Randbemerkung sei auf die *Kolmogoroff-Komplexität* hingewiesen, welche eine Alternative zur Shannon-Information darstellt bei der Frage, welche Information eine Zeichenkette beinhaltet. Unter ihr versteht man *die Länge des kürzesten Programms, welches die Zeichenkette erzeugen kann.*

Eine Zufallsvariable X kann im Allgemeinen mehrere Werte x annehmen. *Die durchschnittliche Fragenanzahl der verschiedenen Werte* ist eine Eigenschaft der gesamten Variablen. Sie heißt

Entropie. Ist X eine diskrete Zufallsvariable, so ist ihre Entropie definiert als

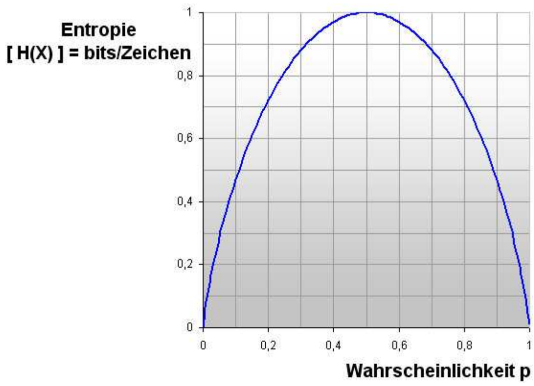
$$\boxed{H(X) := \sum_{x\in X} P(x)h(x) = -\sum_{x\in X} P(x)\log_2 P(x)}. \tag{2}$$

Die Entropie ist nie negativ. Sie misst den Grad der Unordnung eines Systems (genauer: die Unsicherheit über den Zustand eines Systems), welches von X beschrieben wird, weil die Identifikation des Zustands eines unstrukturierten Systems mehr Fragen erfordert.

Im Fall einer kontinuierlichen Zufallsvariable X muss die Summation durch eine Integration ersetzt werden:

$$\boxed{H(X) = -\int_{-\infty}^{+\infty} P(x)\log_2 P(x)dx}. \tag{3}$$

Für den einfachen Fall, dass X nur zwei Werte x_1 und x_2 mit den Wahrscheinlichkeiten p und 1-p annehmen kann, ist in nebenstehendem Diagramm die Entropie in Abhängigkeit von p aufgetragen. Dabei findet die Formel (2) mit $P(x_1)=p$ und $P(x_2)=1-p$ Verwendung. Ist die Wahrscheinlichkeit der beiden Zustände ungefähr gleich (um p=0,5), so enthält das Wissen, welcher der beiden Zustände tatsächlich vorliegt, viel Information und die Entropie ist maximal. Ist hingegen der Zustand des Systems bereits im Vorfeld gut bekannt (p nahe 0 oder 1), so ist der zu erwartende Informationsgewinn bei Erfahren des tatsächlichen Zustands gering: Wahrscheinlich kennt man ihn bereits, ohne überhaupt eine Frage stellen zu müssen.



Ist X – um auch ein kontinuierliches Beispiel zu nennen – normalverteilt mit der Varianz σ^2, so ist nach (3) und Rechnung die Entropie $H(X) = \frac{1}{2}\log_2\left(2\pi e\sigma^2\right)$. Übrigens ist die Normalverteilung diejenige Verteilung mit der größten Entropie bei gegebener Varianz. Sie enthält also in diesem Sinn a priori minimale Information. Viele Fragen sind erforderlich, um den tatsächlichen Zustand zu erfahren. Daher enthält das Wissen a posteriori entsprechend viel Information.

Wir stellen fest: Die Entropie misst auch den Unterschied von a priori- zur a posteriori-Information.

In der Thermodynamik wird die physikalische Entropie definiert als $S := k\ln W$, wobei k die Boltzmann-Konstante und W die Anzahl möglicher Mikrozustände ist. Diese Anzahl stellt im Übrigen auch eine prinzipielle Grenze für Informationsspeicher dar. Es sei ferner angemerkt, dass ein inhaltlicher Zusammenhang der physikalischen Entropie zur Shannonschen Definition in (2) besteht, der die identische Bezeichnung rechtfertigt.

Ein weiterer wichtiger Begriff ist die

Bedingte Entropie. Sind X und Y diskrete Zufallsvariablen, so ist die bedingte Entropie von Y gegeben X definiert als

$H(Y \mid X) := \sum_{x\in X} P(x)H(Y \mid x)$, mit $H(Y \mid x) := -\sum_{y\in Y} P(y \mid x)\log_2 P(y \mid x)$ in Analogie zu (2). Es ist also

$$\boxed{H(Y \mid X) = -\sum_{x\in X} P(x)\sum_{y\in Y} P(y \mid x)\log_2 P(y \mid x)}. \tag{4}$$

Die bedingte Entropie misst also *die nach Kenntnis von* X *verbleibende Unsicherheit über* Y. Damit eignet sie sich, um zu definieren, in welchem informatorischen Zusammenhang zwei Variablen stehen. Dies ist die

Gegenseitige Information. Zwei Zufallsvariablen enthalten übereinander die gegenseitige Information

$$\boxed{I(X;Y) := H(Y) - H(Y \mid X) = H(X) - H(X \mid Y) = H(X) + (Y) - H(X,Y)}. \tag{5}$$

Mit ihr wird ausgedrückt, *welche Information beim Erfragen des Zustands von* X *verschwendet wird, wenn dabei die Kenntnis von* Y *unberücksichtigt bleibt.* Wir betrachten die beiden Extremfälle

1. X=Y. Aus (4) folgt mit $\log_2 P(x \mid x) = \log_2 1 = 0$ H(X|X)=0, also I(X;X)=H(X). Die verschwendete Information ist maximal. Als nächstes seien
2. X, Y unabhängig. Dann folgt wegen H(Y|X)=H(Y) bereits I(X;Y)=H(Y)-H(Y)=0. Hier enthalten die beiden Variablen also keine gegenseitige Information.

Ein wichtiger Spezialfall ist die

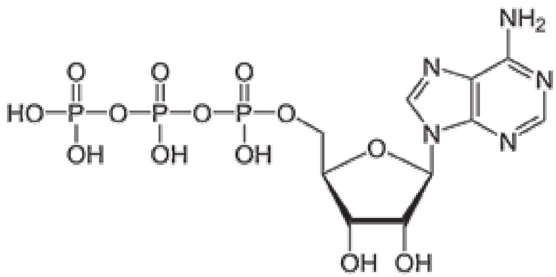
Vorhersagende Information. Beschreibt Y die Vergangenheit und X die Zukunft eines Systems, so heißt ihre gegenseitige Information auch vorhersagende Information. (6)

Die vorhersagende Information ist für Organismen aller Art von Bedeutung. Aus dem aktuellen Zustand der Welt muss eine Prognose für die (nahe) Zukunft getroffen werden, auf deren Grundlage eine (Re-)Aktion gewählt wird, deren Durchführung immer Zeit erfordert. Die Handlungsfähigkeit des Organismus hängt also entscheidend von der vorhersagenden Information ab.

Die vorhersagende Information wächst *sublinear* mit der Beobachtungszeit. Sehr umfangreiches Sammeln von Information verbessert die Vorhersage also nicht mehr wesentlich. Dadurch ist der Nachteil, den unvollkommene Organismen gegenüber optimalen haben, abgemildert. Das fördert die Artenvielfalt.

Die in (5) erkennbare Symmetrie der gegenseitigen Information erhält im Kontext von vorhersagender Information eine interessante Deutung: Es ist ebenso viel Information erforderlich, um von der Vergangenheit auf die Zukunft zu schließen, wie umgekehrt. Leider ist oft nicht bekannt, bei *welchem Teil* der insgesamt vorliegenden Information es sich um die relevante vorhersagende Information handelt, und wie diese nutzbar ist.

Im Übrigen hat das (menschliche) Auge ein vergleichbares Problem zu bewältigen: Die Retina empfängt einen Informationsfluss von ca. 10^9 bit/s, während der Sehnerv nur ca. 10^7 bit/s ans Gehirn weiterleiten kann. Das Auge steht also vor der schwierigen Aufgabe, die Information auf 1% zu reduzieren. *Offensichtlich* ist es recht effektiv darin, die für das Gehirn relevante Information herauszufischen. Dabei bedient es sich womöglich so genannter *rezeptiver Felder*, welche aus Neuronen bestehen, die in wohlunterschiedenen Schichten angeordnet sind, und die Informationen benachbarter Neuronen Schicht für Schicht zusammenführen.



Sensorische Informationsaufnahme und neuronale Weitergabe ist energetisch teuer, gemessen in der biologischen „Währung“ Adenosintriphosphat (ATP). Eine Fliege verbraucht beispielsweise allein 8% ihres ATP in ihren Photosensoren. Ein einzelner Photorezeptor benötigt ca. 7,5 10^6 ATP/bit, eine Synapse ca. 4 10^4 ATP/bit und der Mensch benötigt ca. 40% seines gesamten Energieverbrauchs, um das Spannungsgefälle an den Neuronenwänden aufrecht zu erhalten, auf dem sämtliche neuronale Signalweitergabe beruht. Information ist offenbar ein außerordentlich kostbares Gut.

Noch nicht abschließend geklärt ist die Frage, in wie weit die Evolution selbst durch informationstheoretische Modelle zu erklären ist. Die Entstehung von hoher Intelligenz in recht unterschiedlichen Entwicklungslinien (Affe, Tintenfisch, manche Vogelarten, ...) lässt vermuten, dass weniger der Zufall, als vielmehr ein grundlegendes Prinzip den evolutionären Prozess dominiert und intelligente Organismen bedingt.

Bislang konnte dies in rechnergestützten Simulationen jedoch nicht reproduziert werden. Agenten, also virtuellen Organismen, mangelt es immer noch an grundlegenden Fähigkeiten intelligenten Handelns, wie etwa der Möglichkeit zur Verallgemeinerung. Zwar gibt es einige vielversprechende Ansätze, wie beispielsweise das *empowerment*, das einem Agenten als einzige Anweisung auferlegt, seinen Handlungsspielraum und die (vorhersagende) Information über seine Umgebung zu maximieren, und das einige intelligent wirkende Verhaltensmuster hervorruft. Auch ist dies vergleichbar mit dem Bestreben von Lebewesen, eine Bestätigung ihres inneren Modells der Umwelt durch die sensorischen Reize herbeizuführen. Doch immer noch zeugen hoch spezialisierte Algorithmen, wie sie beispielsweise in Schach-Computern verwendet werden, weniger von einem Siegeszug der künstlichen Intelligenz, als viel mehr von dem menschlichen Unvermögen, einige wenige evolutionäre Prinzipien dazu zu bringen, aus sich selbst heraus schöpferisch tätig zu werden. Das kreative Denken haben uns die Computer noch nicht abnehmen können. Sie geben uns nur das zurück, was wir in sie hineinstecken.

Trotz großer Fortschritte, unter anderem auf dem Gebiet der Informationstheorie, haben wir Menschen das Leben noch immer nicht verstanden.