Theorie SIA-Arbeit

ENTROPIE EINES WIRTSCHAFTSSYSTEMS

Daniel Meiborg

Januar 2023

Inhaltsverzeichnis

You should call it entropy [...] no one really knows what entropy really is, so in a debate you will always have the advantage.

John von Neumann zu Claude Shannon, *Scientific American Vol. 225 No. 3, (1971)*

Information

In der Informatik wird der Informationsgehalt, auch genannt Shannon-Information, als Menge an “Überraschung” eines Ereignisses definiert. Formal ist der Informationsgehalt eines Ereignisses als $I(x) = -\log_2(p(x))$ (in Bit) definiert. Hierbei ist zu beachten, dass diese Größe nichts mit dem im Sprachgebrauch verwendeten Begriff von Information zu tun haben muss. Der Informationsgehalt sagt nichts darüber aus, wie nützlich die Nachricht ist. Stattdessen kann diese Größe als die natürliche Grenze für die Kompression einer Nachricht verstanden werden, oder als das Minimum an Ja-Nein Fragen, um das Ergebnis zu bestimmen (bei nicht-natürlichen Zahlen die durchschnittliche Anzahl).

Ein Beispiel: Eine typische faire Münze hat eine Wahrscheinlichkeit von $1/2$ für Kopf und $1/2$ für Zahl. Der Informationsgehalt für beide Ereignise ist also jeweils $-\log_2(1/2) = 1$ Bit. Bei einer nicht-fairen Münze mit einer Wahrscheinlichkeit von $1/3$ für Kopf und $2/3$ für Zahl ist der Informationsgehalt für Kopf $-\log_2(1/3) = 1,585$ Bit und für Zahl $-\log_2(2/3) = 0,585$ Bit.

Eigenschaften des Informationsgehalts

Der Informationsgehalt ist eine streng monoton sinkende Funktion der Wahrscheinlichkeit, welche für $(0, 1]$ definiert ist. Der Informationsgehalt von einem Ereignis, welches völlig sicher ist, also eine Wahrscheinlichkeit von 1 hat, ist 0. Der Informationsgehalt von einem Ereignis, welches nicht vorkommen kann, wäre ∞.

Der Informationsgehalt von zwei unabhängigen Ereignissen ist die Summe der beiden Informationsgehalte.

Entropie

Arten und Anwendungen von Entropie

Entropie findet sich auf viele verschiedene Arten der Wissenschaft wieder. In der Physik findet man sie in Gestalt der quantenmechanisch definierten Von-Neumann-Entropie, in den Temperaturdifferenzen verschiedener Systeme und in der statistischen Mechanik, von denen jede eine etwas andere Definition verwendet. In der Informatik wird hauptsächlich die Entropie nach Claude Shannon verwendet. Für die Zwecke dieser Arbeit wird ausschließlich die Shannon-Entropie verwendet und ist synonym mit Entropie zu verstehen.

Formale Definition

Die Entropie ist der durchschnittliche Informationsgehalt einer zufälligen Variable. Sie ist definiert als $H(X) = \sum_{x \in X} p(x) \cdot I(x) = -\sum_{x \in X} p(x) \cdot \log_2 p(x)$ (in Bit), wobei $p(x)$ die Wahrscheinlichkeitsfunktion der zufälligen Variable ist.

Anschauliche Erläuterung

Entropie wird oft als Maß für die Unordnung eines Systems beschrieben. Die Shannon-Entropie ist allerdings etwas abstrakter: Sie beschreibt die Verteilung von Wahrscheinlichkeiten.

Die Entropie ist eng mit der statistischen Mechanik verknüpft. In dieser hat ein Gas, welches gleichmäßig verteilt ist, eine hohe Entropie, während ein Zustand, in dem alle Gasmoleküle an einer Seite des Raumes sind, eine niedrige Entropie hat. Das hat den Grund, dass es deutlich mehr Zustände i.e. eine höhere Wahrscheinlichkeit gibt, dass der erste Zustand eintritt. Dieses Verhältnis entspricht allerdings eher der Information der Zustände als der Entropie des Systems.

Nehmen wir wieder das Beispiel mit der Münze: Bei der fairen Münze ist die Entropie $\frac{1+1}{2} = 1$ Bit. Bei der nicht-fairen Münze ist die Entropie $\frac{1{,}585 \cdot 1/3 + 0{,}585 \cdot 2/3}{2} \approx 0,459$ Bit. Die Entropie ist also niedriger als bei der Einheitsverteilung. Diese Tatsache gilt für alle Wahrscheinlichkeiten i.e. die

Einheitsverteilung hat immer die höchste Entropie. Die niedrigste Entropie haben Verteilungen, be denen nur ein einziger Zustand möglich ist.

Bedingte Entropie

Sei $H(X|Y) = \sum_{x,y} p(y)p(x|y)\log_2 p(x|y)$ die bedingte Entropie von X unter der Bedingung Y. $H(X|Y)$ ist nie größer als $H(X)$ und nur gleich, wenn Y vollkommen unabhängig von X ist. Das lässt sich anschaulich so erklären, dass zusätzliche Information (Y) nicht Wissen über X zerstören kann, sondern höchstens vollkommen irrelevant sein kann.

Relative Entropie

$D(p||r) = \sum_x p(x)\log_2 \frac{p(x)}{r(x)}$ ist mit p und r zwei Wahrscheinlichkeitsverteilungen für x die relative Entropie, auch Kullback-Leibler Divergenz genannt. Die relative Entropie ist ein Maß für die Unterschiedlichkeit zweier Wahrscheinlichkeitsverteilungen. Sie ist nicht symmetrisch, d.h. $D(p||r) = D(r||p)$ ist nicht gegeben. Sie kann als Menge an Information interpretiert werden, die man von p erhält, wenn r gegeben ist.

Markov-Prozesse

Für diese Zwecke sollen lediglich endliche diskrete Markov-Prozesse betrachtet werden, i.e. Prozesse, bei denen es nur endlich viele Zustände gibt und die Zeit in diskreten Schritten abläuft.

Definition

Ein Markov-Prozess, auch Markov Chain genannt, ist ein stochastischer Prozess, der die Markov-Eigenschaft besitzt. Er besteht aus einer Menge an Zuständen und einer Wahrscheinlichkeitsmatrix, die die Übergangswahrscheinlichkeiten zwischen den Zuständen beschreibt. Die Markov-Eigenschaft besagt, dass die Übergangswahrscheinlichkeiten ausschließlich von dem aktuellen Zustand abhängen. Sei A die Übergangsmatrix mit A_{ij} als Wahrscheinlichkeit, dass Zustand j in Zustand i übergeht, und $\vec{P}(t)$ die Wahrscheinlichkeiten der Zustände zu Zeitpunkt t in Vektorform. Dann lautet die sogenannte Mastergleichung $\vec{P}(t+1) = A\vec{P}(t)$.

Stationäre Verteilung

Die stationäre Wahrscheinlichkeitsverteilung eines Markov-Prozesses ist die Verteilung $\vec{P}$, die erfüllt, dass $\vec{P} = A\vec{P}$. Die stationäre Verteilung ist ein Gleichgewichtszustand des Prozesses. $\vec{P}$ ist also ein Eigenvektor von A mit Eigenwert 1.

Zweiter Hauptsatz der Thermodynamik

Für den zweiten Hauptsatz der Thermodynamik existiert keine eindeutige Formulierung. Alle aber enthalten die folgende Kernaussage:

> *Entropie kann in einem geschlossenem thermodynamischen System nicht abnehmen.*

Diese Aussage lässt sich bis zu einem gewissen Grad allgemein auf stochastische Prozesse zurückführen. Die Beweise dafür wurden in diesem Paper geführt [1], aus dem ich die wichtigsten Aussagen und Beweise skizzieren werde. Die Notation X_n bezeichnet den Zustand des Systems zu Zeitpunkt n und X_n^k die Menge aller Zustände vom Zeitpunkt n bis k.

$H(X_n)$ steigt monoton

$H(X_n)$ steigt monoton für alle endlichen diskreten Markov-Prozesse wenn die stationäre Verteilung uniform ist.

Sei m die Anzahl an möglichen Zuständen, μ_n eine Wahrscheinlichkeitsverteilung zu Zeitpunkt n und μ die stationäre Verteilung, in diesem Fall die Einheitsverteilung. Dann ist

$$D(\mu_n||\mu) = \sum_x \mu_n(x) \cdot log_2(\tfrac{\mu_n(x)}{1/m}) = -H(\mu_n) + log(m)$$

Da $D(\mu_n||\mu)$ monoton sinkt [1], steigt $H(\mu_n)$ monoton.

Dieser Satz lässt sich wie folgt verstehen: Angenommen, man hat einen Markov-Prozess, keine Informationen über den vorherrschenden Zustand (allerdings kennt man die prinzipiell möglichen Zustände und die Übergangswahrscheinlichkeiten) und führt man den Prozess für eine gewisse Zeit aus. Wenn man danach nicht genauer sagen kann, welche Zustände wie wahrscheinlich sind, gilt der zweite Hauptsatz der Thermodynamik.

$H(X_0|X_{-\infty}^{-n})$ steigt monoton

Die bedingte Entropie $H(X_0|X_{-\infty}^{-n}) = H(X_0|X_{-n}, X_{-(n+1)}, ...)$ der Gegenwart mit der gegebenen Verganenheit steigt monoton für alle stochastischen Prozesse.

Man kann $H(X_0|X_{-\infty}^{-n})$ auch wie folgt darstellen:

$$H(X_0|X_{-\infty}^{-n}) = H(X_0|X_{-\infty}^{-(n+1)}, X_{-n})$$

Zusätzliche Information senkt Entropie (Bedingte Entropie). Also gilt

$$H(X_0|X_{-\infty}^{-(n+1)}, X_{-n}) \leq H(X_0|X_{-\infty}^{-(n+1)})$$

und damit

$$H(X_0|X_{-\infty}^{-n}) \leq H(X_0|X_{-\infty}^{-(n+1)}).$$

Anschaulich bedeutet das, dass wenn man die Information über die Vergangenheit X_{-n} ignoriert, man dadurch nur Information über die Gegenwart verlieren kann.

Quellen

[1] T. M. Cover and J. Halliwell, “Which processes satisfy the second law,” *Cambridge University Press New York, NY*, pp. 98–107, 1994.