Informationstheorie: Entscheidungsgehalt und Information

Martin Lösch

3. November 2004

Zusammenfassung

Ohne Anspruch auf Vollständigkeit erheben zu wollen, faßt dieser Text die wichtigsten Begriffe zu dem Themenbereich des Entscheidungsgehalts von Zeichen und Zeichenmengen zusammen.

Sollte jemand Fehler oder Ungenauigkeiten entdecken, kann er sie mir gerne per Email unter dukemarty@gmx.de mitteilen.

1 Wozu das Ganze

Wozu treiben wir überhaupt diesen ganzen Aufwand mit der Informationstheorie? Nun, eigentlich geht es einfach um folgendes: um die gleiche Information aufzuzeichnen, gibt es immer verschiedene Möglichkeiten, oder auch *Codierungen*. Aber nicht jede Codierung hat die gleichen Eigenschaften. Einige können z.B. das Erkennen von fehlenden oder verfälschten Informationen ermöglichen (redundante Codierungen).

Wenn wir nun Informationstheorie betreiben, geht es letzten Endes um die Frage: wie sieht die "optimale" Codierung von Zeichen eines vorgegebenen Zeichenvorrates aus? (dabei ist mit optimal gemeint: Verwendung von möglichst wenig Speicher)

2 Was bedeutet der Begriff "Entscheidungsgehalt"

Wir haben das Problem, zu entscheiden, welches Zeichen eines im voraus bekannten Zeichenvorrates vorliegt. Und unser Ziel ist es, die Anzahl dafür nötiger Entscheidungen möglichst klein zu halten.

Die Idee dabei ist, in jedem "Entscheidungsschritt" die Hälfte der möglichen Zeichen auszuschließen (wobei wir schon im voraus wissen, mit welcher Wahrscheinlichkeit die einzelnen Zeichen auftreten): man kann das so darstellen, daß man in die Wurzel die Menge aller Zeichen einträgt, und dann 2 Verzweigungen macht, jede mit einer Wahrscheinlichkeit von $\frac{1}{2}$ (das ist der ideale Fall). Ans Ende dieser Verzweigungen kommt je ein neuer Knoten, in den die Zeichenmenge eingetragen wird, die man auf diesem Weg erreichen kann. Die Zeichenmenge der Wurzel wird also in genau 2 Teile geteilt derart, daß jeder dieser beiden Teile eine Auftrittswahrscheinlichkeit von 50% hat.

Auf diese Art und Weise fährt man fort, bis in jedem Blatt nur noch einelementige Zeichenmengen stehen.

Die Auftrittswahrscheinlichkeit eines jeden Zeichens ist dann dadurch wiederberechenbar, daß man die Wahrscheinlichkeiten auf dem Weg von Wurzel zum Blatt aufmultipliziert. Die Summe der Wahrscheinlichkeiten aller Blätter ergibt natürlich immer genau 1.

Und der Entscheidungsgehalt eines Zeichens ist nun einfach die Anzahl der Knoten, die auf dem Weg von der Wurzel bis zu diesem Zeichen durchlaufen werden, also die Anzahl der Entscheidungen, die man zwischen "links" und "rechts" treffen muß, bis man bei dem Zeichen angelangt ist.

3 Weitere Bedeutung des Entscheidungsgehaltes

Der Entscheidungsgehalt eines Zeichens stellt auch ein Maß dafür dar, wieviel *Information* in diesem Zeichen steckt.

Man kann sich als Beispiel vorstellen einen Strom von Daten, den man empfängt. Dieser Strom besteht zu 90% aus dem Zeichen 'a', und zu 1% aus dem Zeichen 'b'. Dann stellt natürlich das Auftreten eines 'b' eine größere Information dar als ein 'a', denn eigentlich erwartet man ja bei jedem Zeichen, das man sich ansieht, eher ein 'a' als ein 'b'.

Wenn man nun von diesen Wahrscheinlichkeiten ausgehen würde (mit noch einigen weiteren Zeichen im Zeichenvorrat), so stünde natürlich das 'a' weiter oben und das 'b' weiter unten im Baum. Der Entscheidungsgehalt wächst also ebenso wie die Information, die ein Zeichen trägt, umgekehrt proportional zur Wahrscheinlichkeit, mit der dieses Zeichen auftritt.

4 Wie berechnet man den Entscheidungsgehalt

Mathematisch definiert man dann den *Entscheidungsgehalt* als die Anzahl der Entscheidungen, die man in einem Binärbaum absteigen muß, bis man das Zeichen erreicht hat:

$$\text{Entscheidungsgehalt für das Zeichen i} := \log_2 \frac{1}{p_i}$$

Mittels dieses Entscheidungsgehaltes auf Zeichenebene kann man dann den *Entscheidungsgehalt H eines Zeichenvorrats* berechnen, indem man den Entscheidungsgehalt aller Zeichen aufaddiert, gewichtet durch die Wahrscheinlichkeit des Auftretens dieses Zeichens:

$$H = \sum_i p_i \cdot \log_2 \frac{1}{p_i} = -\sum_i p_i \cdot \log_2 p_i$$

Der Wert von H ist auch als *mittlerer Entscheidungsgehalt* eines beliebigen Zeichens aus dem Zeichenvorrat interpretierbar, d.h. jedes Zeichen im Zeichenvorrat trägt im Schnitt eine Information von H bits ([bit] ist die Einheit, die für die Information verwendet wird; sie kann (mehr oder weniger) jeden beliebigen Wert annehmen, auch Dezimalbrüche. Im Unterschied dazu sind Werte mit der Einheit [Bit] immer ganzzahlig, weil damit eine Anzahl zweiwertiger Zeichen gemessen wird).

5 Wie codiert man dann nach der Theorie

Jetzt ist nur noch die Frage offen: wie kann man denn nun von einem wie oben beschrieben aufgebauten Baum auf einen Zeichencode schließen? Ganz einfach: jede Kante steht für ein (zweiwertiges Zeichen). Jede Kante nach links wird mit einer '0' markiert, jede Kante nach rechts mit einem 'L'. Der Code für jedes Zeichen wird dadurch bestimmt, daß man von der Wurzel bis zu dem Blatt läuft, der das gesuchte Zeichen enthält, und das Zeichen jeder Kante, die man passiert hat, aufzeichnet.