



Zum Begriff der Redundanz

Motivation

Sowohl im Alltagsgebrauch als auch in unserm informationstheoretischen Sinne ist Redundanz *überflüssige* Information.¹ Es handelt sich also um bloß scheinbare Information, solche, die wie welche aussieht, ohne tatsächlich welche zu sein.

Maß

Dies leitet uns schon mal zur folgenden Definition: Wenn \tilde{H} angibt, wieviel “wie Information aussieht” – und das ist die Länge einer Zeichenkette, eines Kodeworts –, und H der wirklich vorhandene Informationsgehalt nach Claude E. Shannons Formel $H(p_1, \dots, p_n) = -\sum_i p_i \log p_i$ ist, dann ist die überflüssige “Information”, also die, die in Wirklichkeit gar keine ist,² rein anschaulich die

$$\text{Redundanz } R_{\text{abs}} = \tilde{H} - H$$

Genauer gesagt ist dies die **absolute Redundanz** und muß mit einer Einheit, z. B. in bit, angegeben werden. Bei der Angabe der absoluten Redundanz muß immer klar sein, welcher Code die Bezugsgröße ist: Handelt es sich um ein Zeichen (ein Bit? ein Byte? ein Buchstabe? eine Ziffer?) oder um eine kodierende Zusammenfassung mehrerer solcher Zeichen, also um ein Metazeichen?

Diese Unschärfen, die häufig genug zu Mißverständnissen führen können, lassen sich umgehen, wenn stattdessen die

$$\text{relative Redundanz } R_{\text{rel}} = 1 - \frac{H}{\tilde{H}} = \frac{R_{\text{abs}}}{\tilde{H}}$$

betrachtet wird, die durch \tilde{H} normiert ist. Die relative Redundanz wird häufig auch nur **Redundanz** genannt und ist stets $\in [0, 1]$, man kann sie also bequem prozentual angeben.

Es versteht sich von selbst, daß sowohl bei der Berechnung von $R = \tilde{H} - H$ als auch von $R = 1 - H/\tilde{H}$ sich H und \tilde{H} auf denselben Sachverhalt, das heißt: auf dieselbe “Äußerung” beziehen müssen, deren Redundanz R zu ermitteln ist.

¹eine vertrauliche Wissensquelle weist die folgenden Synonyme für “Redundanz” aus: Anhäufung, Fülle, Luxus, Masse, Menge, Opulenz, Reichtum, Zuviel, Überangebot, Überfluß, Überfülle, Überfüllung, Übermaß, Überproduktion, Überreichtum, Überschuß, Überschwang, Üppigkeit

²nach Shannons strengem Begriff davon

Nominal- und Realinformation

Die Redundanz (absolute wie relative gleichermaßen) ergibt sich also stets aus den beiden Größen \tilde{H} und H :

- Die Größe

$$H$$

nennen wir **Realinformation**, also “echte”, tatsächlich vorhandene Information. Sie ergibt sich aus Shannons oben zitierter Entropieformel für Informationsquellen (und ist somit von der Art ihrer Darstellung unabhängig).

- Die Größe

$$\tilde{H}$$

hingegen gibt die nur scheinbar vorhandene, angebliche Informationsmenge an, die wir daher Nenn- oder **Nominalinformation** nennen. Sie ist identisch mit der Länge der Äußerung, d. h. des Kodewortes.

Wird also H als Erwartungswert

$$H = \mathcal{E}(I(p_i))_{i=1,\dots,n} = \sum_{i=1}^n p_i \cdot I(p_i) = - \sum_{i=1}^n p_i \log p_i$$

über die *Informationswerte* $I(p_i) = -\log p_i$ der n möglichen Äußerungen³ gebildet, so wird \tilde{H} als Erwartungswert

$$\tilde{H} = \mathcal{E}(\text{Länge}(\text{code}_i))_{i=1,\dots,n} = \sum_{i=1}^n p_i \cdot \text{Länge}(\text{code}_i)$$

über die *Kodelängen* der entsprechenden Äußerungen code_i ermittelt.

³die Claude Shannon ausschließlich in Hinblick auf ihre Auftrittswahrscheinlichkeiten p_i interessieren