Zusammenfassung der 09. Vorlesung (14.06.2010)

Fassung vom 20.07.2010

3.4 *Der klassische Informationskanal*:

Um an einem Beispiel zu zeigen , wie die Shannon Entropie verwendet werden kann, betrachten wir einen klassischen gestörten Informationskanal. Sei $\mathbf{A} := \{a_i\}_{i=0,1,2,\ldots,(M-1)}$ ein Alphabet. Zu einem Wort, das auch unendlich lang sein kann, gehört ein Wahrscheinlichkeitsraum $(\mathbf{A}, \mathcal{A}, \mu)$, wobei $\mathcal{A}$ die Potenzmenge von $\mathbf{A}$ und $\mu(a_i)$ die Wahrscheinlichkeit für das zufällige Erkennen des Zeichens a_i ist. Die Entropie des Wortes ist $H = -\sum_{i=0}^{M-1} \mu(a_i) \log_2 \mu(a_i)$.

Der Einfachheit halber wollen wir einen Kanal betrachten, der keine Kodierung erzeugt, d.h. das gesendete und das empfangene Wort bestehen aus Zeichen des Alphabets $\mathbf{A}$, und im Idealfall stimmt das empfangene Wort mit dem gesendeten überein. Bei der Beschreibung eines solchen Kanals sind die Elementarereignisse (a_i, a_k), wobei erste Faktor ein gesendetes und der zweite Faktor das daraufhin empfangenes Zeichen angibt. Zur Beschreibung des Kanals betrachtet man daher einen Wahrscheinlichkeitsraum $(\mathbf{A} \times \mathbf{A}, \mathcal{A} \otimes \mathcal{A}, \sigma)$ mit den marginalen Wahrscheinlichkeitsräumen $(\mathbf{A}, \mathcal{A}, \mu_s)$ für das gesendete und $(\mathbf{A}, \mathcal{A}, \mu_e)$ für das empfangene Wort:

$(\mathbf{A}, \mathcal{A}, \mu_s)$ $\qquad$ $(\mathbf{A} \times \mathbf{A}, \mathcal{A} \otimes \mathcal{A}, \sigma)$ $\qquad$ $(\mathbf{A}, \mathcal{A}, \mu_e)$

Im Fall eines idealen Kanals ist $\sigma(\{(a_i, a_k)\}) = \delta_{i,k}\sigma(\{(a_i, a_i)\})$. Mit $\sigma_{ik} := \sigma(\{(a_i, a_k)\})$ seien im allgemeinen Fall

$$p_i := \mu_s(\{a_i\}) = \sigma((a_i, \mathbf{A})) = \sum_k \sigma_{ik},\ q_k := \mu_e(\{a_k\}) = \sigma((\mathbf{A}, a_k)) = \sum_i \sigma_{ik},$$

wobei $(a_i, \mathbf{A}) = \bigcup_k \{(a_i, a_k)\}$ und $(\mathbf{A}, a_k) = \bigcup_i \{(a_i, a_k)\}$ bezwichnen. Ferner seien

$$p_{ki} := \sigma((\mathbf{A}, a_k)|(a_i, \mathbf{A})) = \frac{\sigma_{ik}}{p_i}, \quad q_{ik} := \sigma(((a_i, \mathbf{A})|(\mathbf{A}, a_k)) = \frac{\sigma_{ik}}{q_k}.$$

Beim realen Kanal muss die Entropie des gesendeten Wortes, $H_s = -\sum_i p_i \log_2 p_i$, nicht mit der Entropie des empfangenen Wortes, $H_e = -\sum_i q_i \log_2 q_i$, übereinstimmen. Zum Beispiel ist $H_s = 0$, wenn das gesendete Wort nur das Zeichen a_i enthält. Die Entropie des empfangenen Wortes ist dann jedoch $H_e = H(q|a_i)$, wobei

$$H(q|a_i) := -\sum_k p_{ki} \log_2 p_{ki}$$

die **bedingte Entropie** genannt wird. Die mit $H(q|a_i)$ gemessene Information, die der Empfänger erhält, ist fur die zu übertragende Information nicht relevant. Allgemein wird der Erwartungswert von $H(q|a_i)$ **Irrelevanz** genannt:

$$\begin{aligned} H(q|p) &= -\sum_{i,k} p_i p_{ki} \log_2 p_{ki} = -\sum_{i,k} \sigma_{ik}(\log_2 \sigma_{ik} - \log_2 p_i) \\ &= -\sum_{i,k} \sigma_{ik} \log_2 \sigma_{ik} + \sum_i \left(\sum_k \sigma_{ik} \right) \log_2 p_i) = H_v - H_s, \end{aligned}$$

wobei $H_v := -\sum_{i,k} \sigma_{ki}(\log_2 \sigma_{ki}$ **Verbundentropie** genannt wird. Im Idealfall gilt $\sigma_{ik} = \delta_{ik}\sigma_{ii}$ und damit $p_i = \sigma_{ii}$. Dann ist $H_v = H_s$ und die Irrelevanz verschwindet.

Entsprechend kann im realen Fall bei Empfang des Zeichens a_k nicht darauf geschlossen werden, dass auch a_k gesendet wurde. Mit der Wahrscheinlichkeit q_{ik} könnte auch a_i gesendet worden sein. Die bedingte Entropie

$$H(p|a_k) = -\sum_i q_{ik} \log_2 q_{ik}$$

misst die Information der Zeichenreihe, die "hinter" dem Empfang des Zeichens stehen könnte, also gleich gelautet hätten. Der Erwartungswert dieser bedingten Entropie über der Verteilung der empfangenen Zeichen wird **Äquivokation** genannt:

$$H(p|q) = -\sum_{i,k} q_k q_{ik} \log_2 q_{ik} = -\sum_{i,k} \sigma_{ik}(\log_2 \sigma_{ik} - \log_2 q_k) = H_v - H_e.$$

Im Idealfall, $\sigma_{ik} = \delta_{ik}\sigma_{ii}$, verschwindet auch die Äquivokation.

Es gelten nun die folgenden Sätze:

Satz 1:

$$H_s - H(q|p) = H_e - H(p|q) = \sum_{ik} \sigma_{ik} \log_2 \frac{\sigma_{ik}}{p_i q_k} =: I.$$

I heißt die **Transinformation** (mutual information). Im Idealfall, $\sigma_{ik} = \delta_{ik}\sigma_{ii}$, gilt $I = H_s = H_e$.

Beweis (von Satz 1):

$$\begin{aligned} I &= \sum_{i,k} \sigma_{ik}(\log_2 \sigma_{ik} - \log_2 p_i - \log_2 q_k) \\ &= -H_v + H_s + H_e = H_s - H(q|p) = H_e - H(p|q). \end{aligned}$$

Satz 2:

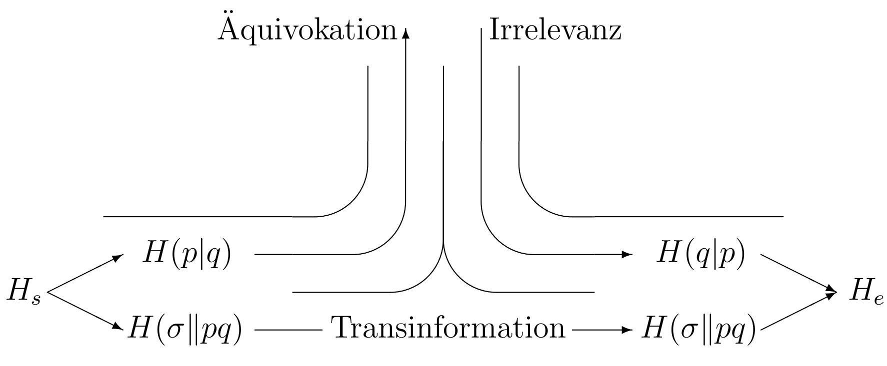
$$I \geq O, \qquad (I = 0 \quad \Leftrightarrow \quad \sigma_{ik} = p_i q_k)$$

Beweis: Die Behauptung folgt unmittelbar daraus, dass

$$I = \sum_{i,k} \sigma_{ik}(\log_2 \sigma_{ik} - \log_2 p_i q_k) = H(\sigma \| pq)$$

ist, wobei $pq = \{p_i q_k\}$ die Verteilung ist, bei der keine Korrelation zwischen gesendetem und empfagenen Zeichen besteht. $H(\sigma \| pq)$ ist die **relative Entropie** der Verteilung σ nach pq, die allgemein die Aussage des Satzes als Eigenschaft hat, wie wir im anschließenden Abschnitt zeigen werden.

Zunächst veranschaulichen wir uns das Ergebnis der Analyse des klassischen Informationskanals im Bergerschen Diagramm:



3.4 *Die relative Entropie*:

Für zwei diskrete Verteilungen $p_i \geq 0$, $\sum_{i=0}^{M-1} p_i = 1$, und $q_i \geq 0$, $\sum_{i=0}^{M-1} q_i = 1$, ist die **relative Entropie** von p nach q durch

$$H(p\|q) = \sum_{i=0}^{M-1} p_i(\log_2 p_i - \log_2 q_i) = -H(p) + \sum_{i=0}^{M-1} p_i(\log_2 q_i)$$

definiert. Die relative Entropie ist nicht symmetrisch in p und q, liefert jedoch ein Maß dafür, wie weit sich p von q unterscheidet. Dies zeigt der folgende Satz:

Satz(Kleinsche Ungleichung):

$$H(p\|q) \geq 0, \qquad (H(p\|q) = 0 \Leftrightarrow p_i = q_i)$$

Beweis: Der Beweis beruht auf der strikten Konkavität des Logarrithmus: Seien $x, y > 0$, dann gilt für $\lambda \in [0,1]$ stets

$$\lambda \log_2 x + (1-\lambda) \log_2 y \leq \log_2(\lambda x + (1-\lambda))y),$$

und die Gleichheit gilt genau dann, wenn $x = y$ oder $\lambda \in \{0,1\}$ ist. Damit folgt

$$\begin{aligned} H(p\|q) &= -\sum_{\{i|p_i\neq 0\}} p_i \log_2 \frac{q_i}{p_i} \\ &\geq -\log_2(\sum_{\{i|p_i\neq 0\}} p_i \frac{q_i}{p_i}) = -\log_2(\sum_{\{i|p_i\neq 0\}} q_i) \geq 0 \end{aligned}$$

Ist nun $H(p\|q) = 0$, dann gilt $-\log_2(\sum_{\{i|p_i\neq 0\}} q_i) = 0$, also $\sum_{\{i|p_i\neq 0\}} q_i = 1$, und damit $p_i = 0 \Leftrightarrow q_i = 0$. überdies ist

$$-\sum_{\{i|p_i\neq 0\}} p_i \log_2 \frac{q_i}{p_i} = -\log_2(\sum_{\{i|p_i\neq 0\}} p_i \frac{q_i}{p_i}) = 0.$$

Aufgrund des nachfolgenden Lemmas folgt aus der strikten Konkavität von $\log_2$

$$(\exists\, i)\ p_i = 1 \vee\ (\forall\, (i,k))\ 0 < p_i, p_k < 1 \Rightarrow x_i = x_k.$$

Gibt es ein i mit $p_i = 1$, dann ist $\frac{q_i}{p_i} = 1$ und $p_k = q_k = 0$ für $i \neq k$. Gibt es kein i mit $p_i = 1$, dann folgt aus $0 < p_i, p_k$, dass $\frac{p_i}{q_i} = \frac{p_k}{q_k} = 1$ und $q_l = 0$ falls $p_l = 0$. Damit ist6 die Kleinsche Ungleichung bewiesen.

Lemma: Sei f ein strikt konkave Funktion auf $(a,b) \subseteq \mathbf{R}$, seien $x_i \in (a,b)$, $i = 1, 2, \ldots, N$ und $p_i \geq 0$, $\sum p_i == 1$, dann gilt

$$\sum_{i=1}^{N} p_i f(x_i) \leq f(\sum_{i=1}^{N} p_i x_i)$$

und

$$\sum_{i=1}^{N} p_i f(x_i) = f(\sum_{i=1}^{N} p_i x_i). \Leftrightarrow .(\exists\, i)\ p_i = 1 \vee (\forall\, (i,k))\ 0 < p_i, p_k < 1 \Rightarrow x_i = x_k.$$

Beweis: Ist f konkav, dann gilt

$$l_y(x) := f(y) + (x-y)f'(y) \geq f(x).$$

Ist f strikt konkav, dann gilt überdies

$$f(y) + (x-y)f'(y) > f(x) \quad \Leftrightarrow \quad x \neq y, .$$

Zunächst folgert man aus der Konkavität von f für $y = \sum_k p_k x_k$, dass

$$f(\sum_k p_k x_k) - f(x_i) \geq (\sum_k p_k x_k - x_i) f'(\sum_k p_k x_k)$$

und mithin

$$\begin{aligned}
\sum_i p_i(f(\sum_k p_k x_k) - f(x_i)) &\geq \sum_i p_i(\sum_k p_k x_k - x_i) f'(\sum_k p_k x_k) \\
f(\sum_k p_k x_k) - \sum_i p_i f(x_i) &\geq (\sum_k p_k x_k - \sum_i p_i x_i) f'(\sum_k p_k x_k) x_i) \\
&\geq 0
\end{aligned}$$

Damit ist die behauptete Ungleichung bewiesen. Sei nun

$$f(\sum_k p_k x_k) - \sum_i p_i f(x_i) = 0 \quad \wedge \quad (\forall i)\ 0 \leq p_i < 1,$$

dann gibt es wenigstens ein Paar (i,k) mit $(p_i, p_k) \neq (0,0)$. Aus der Annahme $x_i \neq x_k$ folgt $x_i \neq \sum_l p_l x_l$ oder $x_i \neq \sum_l p_l x_l$. S, Sei etwa $x_i \neq \sum_l p_l x_l$, dann ist aufgrund der strikten Konkavität von f

$$f(\sum_l p_l x_l) + (x_i - \sum_l p_l x_l) f'(\sum_l p_l x_l) > f(x_i),$$

und damit

$$\sum_i p_i(f(\sum_l p_l x_l) + (x_i - \sum_l p_l x_l) f'(\sum_l p_l x_l)) = f(\sum_l p_l x_l) > \sum_i p_i f(x_i).$$

Da wir die Gleiheit vorausgesetzt haben, ist diess ein Widerspruch. Also gilt

$$(\forall\ i)\ 0 < p_i < 1 \Rightarrow x_i = \sum_l p_l x_l$$

falls kein i mit $p_i = 1$ existiert. Damit ist die Implikation der zweiten Aussage des Lemmas bewiesen. Die Reimplikation ist trivial.