WAS HEISST HEUTE GUT CODIEREN ?

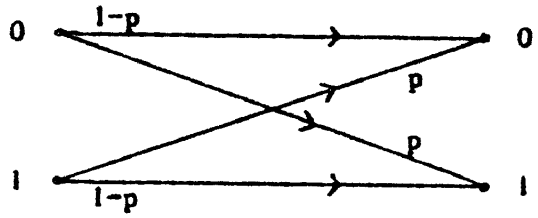
von

J.H. van Lint (Eindhoven)

1. Motivierung

Viele von Ihnen haben ohne Zweifel die wunderbaren Bilder von Mars und Saturn gesehen, die von den Mariner und Voyager Satelliten übertragen wurden. Wie geht das eigentlich?

Ein Bild wird verteilt in viele kleine Quadrate und dann wird von jedem Quadrat die Schwärzung gemessen in einer Skala von 0 bis 63, binär dargestellt mit 6 Bits. Also ergibt ein Bild eine lange Reihe von 0'en und 1'sen als Signal. Nehmen wir an, dass es ohne weiteres zur Erde gesandt wird. Das Signal ist bei Ankunft in Pasadena so schwach, dass es verstärkt werden muss. Durch thermisches Rauschen im Empfänger und andere Störungen passiert es mit einer gewissen Wahrscheinlichkeit p dass eine 0 interpretiert wird als 1 oder umgekehrt:



binärer symmetrischer Kanal

Die Wahrscheinlichkeit p ist so gross, dass ohne besondere Massnahmen die Qualität der Bilder ungeheuer schlecht sein würde. Also, was kann man machen?

Das Einfachste, dass man tun kann, ist jede 0 bzw. 1 nicht ein Mal sondern, sagen wir, fünf Mal zu senden. Wenn dann aus fünf solchen 0'en weniger als 3 falsch ankommen, wird vom Empfänger gut interpretiert, was vom Sender übertragen wurde. Wenn man in dieser Weise eine sehr grosse Genauigkeit der Uebertragung erreichen will, muss man jedes Symbol sehr oft wiederholen. Man hat den Eindruck, dass bei gegebenem p es sogar nicht möglich ist, beliebig grosse Genauigkeit zu erreichen, ohne dass man sehr viel mehr Zeit für die Uebertragung nimmt als à priori erwartet. Erstaunlicherweise ist das aber nicht der Fall! Die genaue Formulierung dieser Aussage heisst der Satz von Shannon (1948). Damit hat die Codierungstheorie angefangen.

Um den Satz zu formulieren und für die Fortsetzung brauchen wir einige Begriffe.

Alphabet $\mathbf{F}$: endliche Menge, z.B. $\{0,1\}$, oft ein Körper.

Wörter der Länge n: Elemente von $\mathbf{F}^n$.

(Hamming-)Abstand $d(\underline{x},\underline{y}) := |\{i \mid 1 \leq i \leq n, x_i \neq y_i\}|$.

(Hamming-)Gewicht $w(\underline{x}) := d(\underline{x},\underline{0})$.

Wir nehmen an, dass Information angeboten wird in k-Tupel $(a_1,\ldots,a_k) \in \mathbf{F}^k$. Eine Codierungsvorschrift ist eine Abbildung $E : \mathbf{F}^k \to \mathbf{F}^n$. Die Menge $E(\mathbf{F}^k)$, oder allgemeiner eine Teilmenge $C \subset \mathbf{F}^n$, ist der Code.

Wir stellen uns vor, dass bei der Uebertragung Rauschen (die Störung) addiert wird, also

$$\underline{c} \mapsto \underline{c} + (e_1,e_2,\ldots,e_n) .$$

Der Empfänger muss eine Abbildung E^+ haben, die für C in der Tat die Umkehrung von E ist und für nicht-Codewörter die Fehlerwahrscheinlichkeit minimiert ("maximum likelihood" Decodierung).

Die Geschwindigkeit der Uebertragung wird gegeben durch die

$$\underline{\text{Informationsrate}} := n^{-1} \cdot {}^2\log |C| \quad , \quad \text{(im Falle } \mathbf{F} = \{0,1\}) \ .$$

Die Entropiefunktion H is definiert durch

$$H(x) := -x \cdot {}^2\log x - (1 - x) \cdot {}^2\log(1 - x) \ .$$

Bei gegebenem p gibt es für jeden Code C die Wahrscheinlichkeit P_c dass ein Wort von C nach Uebertragung und Decodierung falsch interpretiert wird.

Satz (Shannon 1948): *Es sei* $0 < R < 1 - H(p)$. *Es gibt eine Reihe* $C_1, C_2, \ldots$ *von Codes mit Informationsrate* $R_n > R$ *und mit der Eigenschaft*

$$P_{c_n} \to 0 \qquad (n \to \infty).$$

Es ist also klar, dass es Codes gibt, die man "gut" nennen darf. Das bedeutet noch nicht, dass solche Codes in der Praxis leicht zu verwenden sind. Dort gibt es noch viele andere Beschränkungen.

2. Kurze Uebersicht von verschiedenen Codes

Wir betrachten jetzt einige (Klassen von) Codes, die alle "gute" Eigenschaften haben. Für die Fehlerkorrektur ist es wichtig, dass die Minimaldistanz

$$d := \text{Min}\{ d(\underline{u},\underline{v}) \mid \underline{u} \in C, \underline{v} \in C, \underline{u} \neq \underline{v} \}$$

möglichst gross ist, weil die Korrektur von $\lfloor \frac{d-1}{2} \rfloor$ Fehler immer möglich ist.

a) Hamming-Codes (1950)

Wir betrachten $\mathbb{F}_2^n$ als Vektorraum. Es sei $n = 2^\ell - 1$. Der "Prüfmatrix" H (ℓ Reihen, n Spalten) habe alle mögliche Spalten ausser $(0,0,\ldots,0)^T$. Definiere C durch

$$\underline{c} \in C :\Leftrightarrow \underline{c}\, H^T = \underline{0} .$$

Wird ein Wort $\underline{x}$ empfangen mit einem Fehler, sagen wir an der Stelle k, so ist $\underline{x}H^T$ die k^{te} Spalte von H, und der Fehler kann sofort verbessert werden. Hier ist

$$\text{Wortlänge } n = 2^\ell - 1, \quad d = 3, \quad R = 1 - \frac{\ell}{2^\ell - 1} .$$

(Also: Informationsrate fast 1 aber nur 1 Fehler kann verbessert werden).

b) Reed-Muller Codes (1954)

Diese Codes wurden von den Mariners benutzt. Es sei $n = 2^k$. Im Raum $\mathbb{F}_2^k$ betrachten wir alle Hyperebenen. Als Wörter nehmen wir $(0,0,\ldots,0)$, $(1,1,\ldots,1)$ und die charakteristische Funktionen der Hyperebenen. Es ist jetzt leicht zu sehen, dass

$$n = 2^k, \quad d = 2^{k-1}, \quad R = (k+1)/2^k .$$

(Also: sehr viele Fehler können verbessert werden aber R wird rasch klein wenn

Es ist vielleicht interessant zu erwähnen, dass bei der Decodierung (einfach durch Vergleichen) vom Fast-Fourier-Transform Gebrauch gemacht wird.

c) BCH Codes (zyklisch) (1959)

Codierungstheorie ist besonders geeignet als Anwendung der Algebra in einer Vorlesung zu dienen. Sogar für (gute) Schüler gibt es hier Möglichkeiten.

Es sei $f(x)$ ein Teiler von $x^n - 1$. Wir identifizieren

$$(a_0, a_1, \ldots, a_{n-1}) \overset{+}{\leftrightarrow} a_0 + a_1 x + \ldots + a_{n-1} x^{n-1} =: a(x) .$$

Ein Code C wird definiert durch

$$(a_0, \ldots, a_{n-1}) \in C \Leftrightarrow a(x) \text{ ist teilbar durch } f(x) .$$

Die Theorie dieser Codes ist besonders schön. Zum Beispiel die Frage, welche Teiler $f(x)$ zu den besten Ergebnissen führen, ist interessant. Für die sogenannten BCH Codes ist $f(x)$ so gewählt, dass:

$$n \text{ ungerade} , \quad d \text{ ungerade (vorgegeben)} , \quad R \approx 1 - \frac{d \, {}^2\log n}{2n}$$

d) der Golay Code spielt eine grosse Rolle in vielen Teilen der Mathematik, z.B. Gruppentheorie (M_{23}), endliche Geometrien, Kugelpackungen, Blockpläne. Es ist

$$x^{23} - 1 = (x - 1)\, g_0(x) g_1(x) ,$$

wo g_0 und g_1 den Grad 11 haben. Wählen wir in c) $f(x) = g_0(x)$ so bekommen wir den Golay Code. Dieser hat $n = 23$, $d = 7$, $R = 12/23$.

e) Goppa Codes (1970)

Wir werden nachher sehen, dass die Klasse der 1970 von V. Goppa konstruierte Codes wirklich "gut" ist.

$$\text{Man nehme} \begin{cases} L = \{\gamma_1, \ldots, \gamma_n\} \subset \mathbb{F}_{2^m}, \\ \\ g \text{ ein Polynom mit } g(\gamma_i) \neq 0 \quad (1 \leq i \leq n), \end{cases}$$

und definiere der (L,g)-Goppa Code durch

$$(c_1, \ldots, c_n) \in C :\Leftrightarrow \sum_{i=1}^{n} \frac{c_i}{z - \gamma_i} \equiv 0 \pmod{g(z)}.$$

Für die Praxis sind die BCH und Goppa Codes darum gut, weil es Algorithmen zur Decodierung gibt, die sehr schnell sind.

f) Justesen Codes (1972)

Die Entdeckung dieser Codes war eine Sensation, weil man schon glaubte, dass eine explizite algebraische Definition nicht wirklich "gute" Codes geben könnte. Wir werden bald sehen, wie das gemeint ist. Die Definition ist zu kompliziert, um kurz zu beschreiben.

3. Was heisst "gut" ?

Definition: $A(n,d) :=$ das Maximum von $|C|$ für alle binäre Codes C mit Wortlänge n und Minimaldistanz d.

Das wichtigste Problem der kombinatorische Codierungstheorie ist die Bestimmung von (Schranken für) die Zahl $A(n,d)$. Es ist immer ein sehr schönes Kapitel in einer Vorlesung über Codierung.

Wählen wir n fest, dann muss man einen Code C der Länge n und Minimaldistanz d gut nennen, falls $|C| = A(n,d)$.

Aus dem Satz von Shannon sehen wir, dass wir eine Reihe von Codes betrachten sollen mit $n \to \infty$. Ist p fest, dann wird die Anzahl der Fehler pro Wort (im Mittel) wie pn wachsen, und deswegen soll $d > 2pn$ gefordert werden. Das erklärt, warum die Funktion

$$\alpha(\delta) := \overline{\lim_{n\to\infty}}\ n^{-1} \cdot {}^2\log A(n,\delta n)$$

wichtig ist. Es gibt eine Aussage über die grösste Informationsrate bei festem Verhältnis zwischen d und n.

Betrachten wir zuerst eine untere Schranke für $\alpha(\delta)$.

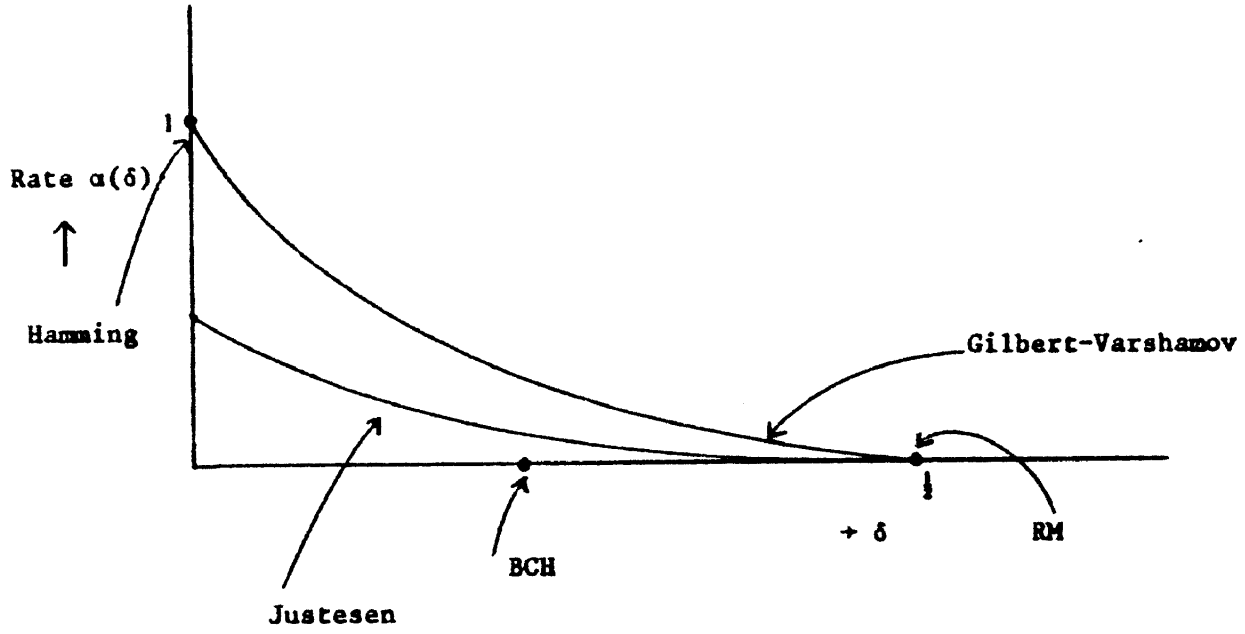
Satz (Gilbert-Varshamov): *Es ist*

$$\alpha(\delta) \geq 1 - H(\delta) .$$

Beweis: Durch eine leichte Rechnung folgt dies aus der triviale Behauptung

$$A(n,d) \cdot \sum_{i=0}^{d-1} \binom{n}{i} \geq 2^n . \qquad \square$$

Die Figur zeigt die bis jetzt behandelten Codes:



Es gibt eine Reihe von Goppa Codes, die die Gilbert-Varshamov Schranke erreicht. In diesem Sinne sind diese Codes gut. Leider ist es bis jetzt nicht möglich, die Reihe explizit zu definieren.

Es soll inzwischen klar geworden sein, dass eine Reihe von Codes gut heissen soll, wenn diese Reihe die Gilbert-Varshamov Schranke erreicht. In diesem Sinne sind die Justesen Codes noch nicht gut, aber wie die Figur zeigt sind diese Codes besser als was vorher bekannt war.

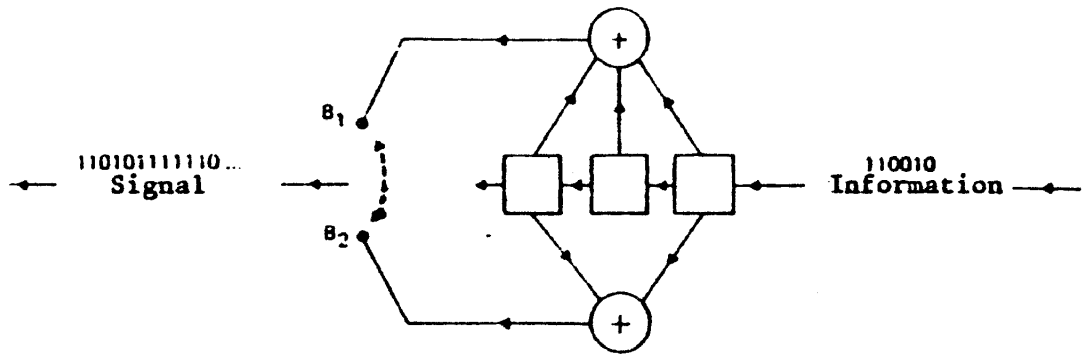
4. Konvolutionelle Codes (gut im Praxis)

Leider für die Mathematiker sind viele von den im Praxis verwendete Codes bis jetzt mathematisch viel weniger interessant. Voyager z.B. benutzte Codierung mit sogenannten konvolutionellen Codes. Wir zeigen kurz, wie solche Codes definiert werden.

Die Information betrachten wir jetzt als eine unendliche Reihe von 0'en und 1'sen. Für die mathematische Beschreibung nehmen wir die Bits als Koeffizienten einer Potenzreihe:

$$I(x) := i_0 + i_1 x + i_2 x^2 + \ldots$$

Das kanonische Beispiel zeigt die folgende Figur



Die Information läuft von rechts durch ein Schieberegister. Zur Zeitpunkt t sind i_{t-2}, i_{t-1}, i_t im Register und an den Ausgänge B_1, B_2 erscheint $i_{t-2} + i_{t-1} + i_t$ bzw. $i_{t-2} + i_t$. Wir beschreiben das durch

$$b_1(x) = (1 + x + x^2)\ I(x)\ ,$$
$$b_2(x) = (1 + x^2)\ \ I(x)\ .$$

Der Empfänger bekommt die gestörten Signale

$$r_1(x) = b_1(x) + n_1(x)\ ,$$
$$r_2(x) = b_2(x) + n_2(x)\ .$$

Er muss aus

$$(1 + x^2)r_1(x) + (1 + x + x^2)r_2(x) = (1 + x^2)n_1(x) + (1 + x + x^2)n_2(x),$$

was also von der Imformation unabhängig ist, eine Schätzung der Störungen $n_1(x), n_2(x)$ machen. Alle bis jetzt benutzte Methoden kommen darauf hin, dass man viele Möglichkeiten vergleicht und die wahrscheinlichste wählt.

Es wird noch viel gearbeitet in dieser Richtung um schnellere Decodierverfahren zu finden. Wichtig ist auch hier um gute Codes zu haben. Dazu muss man die Polynome $p_1(x)$ (oben $1 + x + x^2$) und $p_2(x)$ (oben $1 + x^2$) geeignet wählen. Die jetzt im Praxis benutzte konvolutionelle Codes sind einfach durch Probieren gefunden! Hier liegen noch grosse Möglichkeiten für mathematische Untersuchungen.

LITERATUR

J. Duske, H. Jürgensen, Codierungstheorie, Reihe Informatik B, Bibliographisches Institut, Mannheim, 1977 (nicht einfach, für Mathematiker und Ingenieure)

V. Pless, Introduction to the Theory of Error Correcting Codes, Wiley Interscience, New York, 1981 (für Studenten der Elektrotechnik, Informatik, Mathematik im 4. Semester)

J. Swoboda, Codierung zur Fehlerkorrektur und Fehlererkennung, R. Oldenbourg Verlag, München, 1973 (sehr einfache Einführung)

R.H. Schulz, Kodierung, ein Weg zur Behandlung binärer Strukturen im Unterricht, Didaktik der Mathematik 1 (1973), 70-80

J.H. van Lint, Mathematik und die Satellitenbilder der Planeten, Tetra 1982, Heft 4

J.H. van Lint, Introduction to Coding Theory, GTM 86, Springer Verlag, 1982.