11 Codes

→ Informationsverarbeitung / -übermittlung
→ Nachricht muss standardisiert / effizient gespeichert werden
→ binäre Codierung, Binäralphabet
→ Quellencodierung
→ Nachricht soll robust über fehleranfälligen Kanal versendet werden.
→ Internet, Funk, Handy, Scanner, ...
→ mögliche Übertragung
→ Empfänger erhält nicht die korrekte Information
→ Redundanz für Fehlerkorrektur
→ Kanalcodierung
→ Nachricht kann abgehört / verändert werden
→ Nachricht verschlüsseln und signieren
→ Kryptographie

Hier: kurz Quellencodierung, dann Kanalcodierung

Quellencodierung

→ gegeben: Menge X von Zeichen
z.B. a-z, A-Z, 0-9, ␣, ., ,, !, ?

Menge A von Symbolen für Codierung,
z.B. $A = \{0,1\}$, $A = \mathbb{Z}_q$

A^k := Wörter in A der Länge k

$A^* := \bigcup_{k \geq 0} A^k$

Quellencodierung: $C : X \to A^*$

$\{C(x) \mid x \in X\}$ Codewörter

$C(X) = \text{im}(C)$ Code

Bsp Ascii-Code, Morse-Alphabet

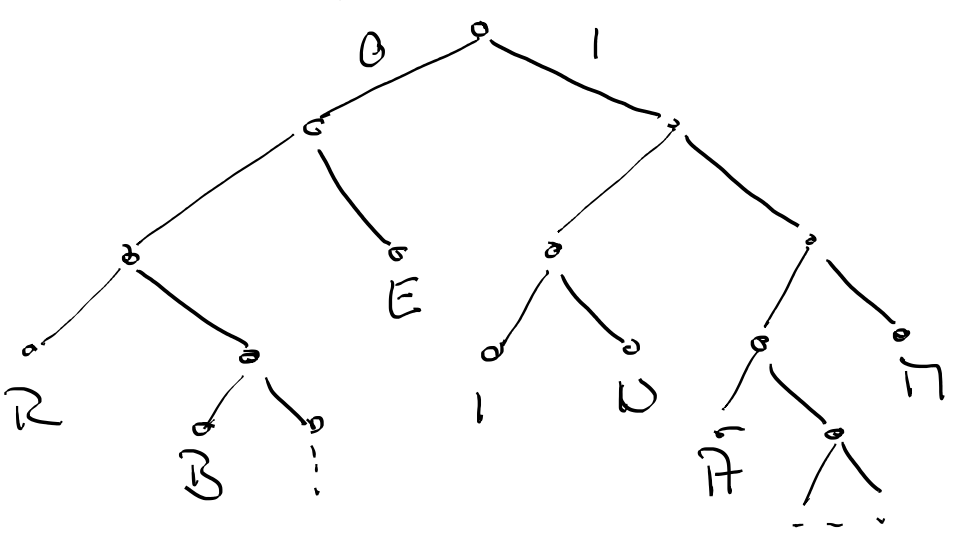
→ Codewörter müssen nicht gleiche Länge haben
→ Effizienz: kurze Wörter für häufige Zeichen.

Bsp $X = \{A, \_, Z\}$ $A = \mathbb{Z}_2$,
$A \to 01$, ... $E \to 0$, ..., $I \to 1$, ..., $S \to 11$
$011 \to EIS \mid AS$

→ Eindeutigkeit nicht gegeben, wenn Codewort Anfang (Präfix) eines anderen sein kann

Def: C ist Präfixcode $\Leftrightarrow$ kein Codewort Präfix eines anderen.

Dann: Decodierung über Baum:



Effizienz: Häufige Zeichen mit kurzem Pfad zu Wurzel $\triangleq$ Länge des Codeworts.

Sei: $l(c)$:= Länge von c

$p(c)$:= Häufigkeit von c (bzw. mit $c = C(x)$)

$\rightarrow$ wollen $L(C) := \sum_{c \in C} p(c)\, l(c)$

minimieren.

$\rightarrow L(C)$ ist die durchschnittliche Länge eines Codeworts.

Satz (Shannon, 1. Hauptsatz der Informationstheorie)

$A = \mathbb{Z}_q$, $C \subseteq A^+$ Code, $p(c)$ Häufigkeit / Wahrscheinlichkeit für $c \in C$.

Dann gilt für den optimalen Code:

$$-\sum p(c) \log_q p(c) \leq \bar{L} := \min L(C) \leq -\sum p(c) \log_q p(c) + 1$$

$\Rightarrow$ ein solcher Code kann effizient konstruiert werden:

$\rightarrow$ Huffman-Code / -Algorithmus

Kanalcodierung

Def: S endliche Menge, $n > 0$. Ein Code C ist eine Teilmenge von S^n

Elemente von C heißen Codewörter

oft: $S = \mathbb{Z}_q$, speziell $S = \mathbb{Z}_2$

dann: C ist binärer Code

Def $u, v \in S$. Dann heißt

$$d(u,v) := |\{ i \mid u_i \neq v_i \}|$$ Hammingabstand.

Def $d(\cdot,\cdot)$ ist Metrik auf S

- $d(x,y) \geq 0, \quad = 0 \Leftrightarrow x = y$
- $d(x,y) = d(y,x)$
- $d(x,z) \leq d(x,y) + d(y,z)$ Dreiecksungleichung

→ Signalübermittlung:

Nachricht w → Codewort x → Empfange $y \in S^n$

Nachricht w' ← Codewort x' ←

dabei: x' mit minimalem Abstand zu y

Sei $B_\varepsilon(a) := \{x \in S^n \mid d(x,a) < \varepsilon\}$

→ Ball mit Radius ε um a

Angenommen, sei der Übertragung von y entstehen höchstens ε Fehler in y

→ wenn $B_\varepsilon(a) \cap B_\varepsilon(b) \geq 2\varepsilon + 1$ für alle $a,b \in C, a \neq b$ dann können wir korrekt decodieren

Def C ist ε-fehlerkorrigierend, wenn

$$B_\varepsilon(a) \cap B_\varepsilon(b) \geq 2\varepsilon + 1 \quad \forall\, a,b \in C,\ a \neq b$$

$$\Leftrightarrow d(a,b) \geq 2\varepsilon + 1 \quad - \text{ " } -$$

wenn $B_e(y) \cap C = \{y\}$, dann können wir zumindest ausschließen, dass Fehler gemacht wurden.

Def C ist **e-fehlerentdeckend**, wenn

$$B_e(a) \cap C = \{a\} \quad \forall a \in C$$

$$\Leftrightarrow d(a,b) \geq e+1 \quad \forall a,b \in C,\ a \neq b$$

Def C Code, dann heißt

$$d(C) := \min(d(a,b) \mid a \neq b)$$

Distanz von C

→ gesucht ist also $C \subseteq S^n$ mit

- $d(C)$ möglichst groß ↔ können viele Fehler korrigiert
- $|C|$ — " — ↔ können viele Nachrichten codiert

→ gegenläufige Ziele !

Satz S mit $|S| = q$, $d > 0$. Dann gibt es Code $C \subseteq S^n$ mit $d(C) = d$ und

$$|C| \geq \frac{q^n}{\sum_{i=0}^{d-1} \binom{n}{i} (q-1)^i}$$

Für $c \in C$: bestimme $|B_{d-1}(c)|$

→ das sind die Elemente von S^n mit $d(c,x) < d$.

- $\binom{n}{i}$ Möglichkeiten, sich an i Positionen zu unterscheiden
- Für jede solche Position: $q-1$ andere Symbole

$$\Rightarrow |B_d(c)| = \sum_{i=0}^{d-1} \binom{n}{i} (q-1)^i =: m$$

Anm: Wenn $|C| \cdot m < q^n$, dann gibt es $y \in S^n$ mit:

$$y \notin \bigcup_{c \in C} B_d(c)$$

$\Rightarrow$ können y zu C hinzufügen ⌟

Satz (Hammingschranke)

S mit $|S| = q$, $C \subseteq S^n$ e-fehlerkompensend. Dann

$$|C| \leq \frac{q^n}{\sum_{i=0}^{e} \binom{n}{i} (q-1)^i} \qquad (*)$$

⌜ Wie zuvor, nur müssen die Bälle um $c \in C$ mit allen $x \in S^n$, $d(x,c) \leq e$ disjunkt sein ⌟

Def: C heißt e-perfekt oder perfekt, wenn in $(*)$ Gleichheit gilt.