Codierung – Fehlererkennung und Fehlerkorrektur

Codierung und Problematik

Codierung ist eine Art Präsentation, die eine Nachricht, z.B. das lateinische Alphabet in einer anderen Form darstellt. Diese andere Darstellungsform kann beispielsweise hilfreich sein bei der Übermittlung von Nachrichten über einen Kanal, da somit die codierte Nachricht leichter übertragen werden kann, dieser Aufbau lässt sich anhand Abb.1.1 darstellen.

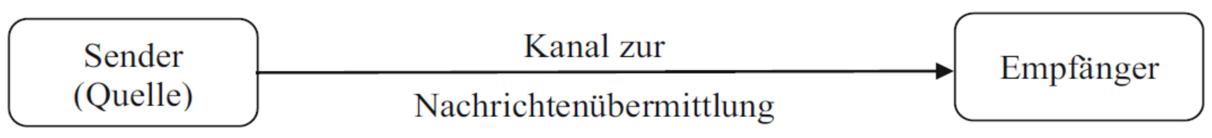

Abb.1.1: Einfaches Schema für die Nachrichtenübertragung

Ziel einer Codierung ist es, dass die von einem Sender versandte Nachricht vom Empfänger eindeutig decodiert werden kann, sodass dieser die korrekte Nachricht erhält. Das bedeutet, dass jedem Nachrichtenwort genau ein Codewort zugewiesen ist und umgekehrt auch jedem Codewort genau ein Nachrichtenwort zugewiesen ist. Dies nennt man auch injektive Abbildung.

Mathematisch betrachtet lässt sich ein Code wie folgt definieren:

Eine Menge F nennen wir ein Alphabet, mit diesem können Wörter der Länge n in F^n gebildet werden.
So ist ein Code C eine Teilmenge von F^n, also $C \subseteq F^n$.

Dabei besteht jedoch das Problem, dass bei der Übertragung von Nachrichten Fehler auftreten können. Es kann möglicherweise zu Störungen kommen, die die codierte Nachricht fehlerhaft machen. Diese sind in Abb.1.2 mit einem Rauschen dargestellt. Dadurch kann es passieren, dass der Empfänger die empfangene Code-Nachricht bzw. das Empfangssignal in das falsche ursprüngliche Nachrichtenwort zurückcodiert. Um dies zu vermeiden, gibt es verschiedene Verfahren, Fehler in Codes zu erkennen und zu korrigieren. Hierzu werden die informationstragenden Info-Bits zu einem Sendesignal erweitert. Die Problematik hierbei ist, dass die zu übertragenden Datenmengen vergrößert werden müssen. Daher ist es wichtig Möglichkeiten zu finden, möglichst viele Fehler, die bei der Übertragung auftreten könnten, zu erkennen und zu korrigieren, damit die korrekte Nachricht beim Empfänger wieder entschlüsselt bzw. decodiert werden kann und dabei die Datenmenge in möglichst geringem Maße zu erweitern.

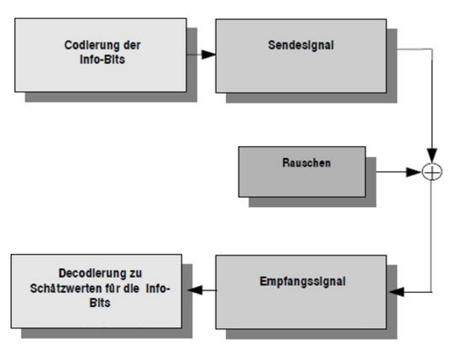

Abb.1.2: Prozess der Informationsübertragung und Fehlerbeseitigung

Beispiele für Codes

Einer der ältesten vom Menschen entwickelten Codes ist das **Morsealphabet**. Hierbei werden Buchstaben des lateinischen Alphabets mit einer Kombination aus kurzen und langen Signalen dargestellt, welche schließlich in Form von elektrischen Signalen an den Empfänger gefunkt werden können.

Ein weiteres Beispiel ist der **IBAN-Code**, welcher sich wie folgt zusammensetzt:

Länderkennung +	Prüfziffer +	Bankleitzahl +	Kontonummer =	**IBAN**
DE	21	3012 0400	0000 0152 28	

Theoretischer Hintergrund

Hammingabstand und Mindestabstand von Codewörtern

- Hammingabstand: Anzahl der Stellen, in denen sich zwei Codewörter unterscheiden
  - Bsp.: „Hase" ⟷ „Nase" ⟶ Hammingabstand: 1
  - 010110 ⟷ 011100 ⟶ Hammingabstand: 2
- Mindestabstand: minimaler Hammingabstand eines Codes

Wahrscheinlichkeit für fehlerhafte Übertragung

- Nachricht „Hase" ⟶ Länge $n = 4$
- Wahrscheinlichkeit für korrekte Übermittlung eines Buchstaben: $p > \frac{1}{2}$
- Mithilfe der Binomialverteilung können folgende Schlussfolgerungen vollzogen werden:

$$p^3 \cdot (1-p) > \left(\frac{1}{2}\right)^2 \cdot p \cdot (1-p)$$

$$\left(\frac{1}{2}\right)^2 \cdot p \cdot (1-p) > (1-p)^2 \cdot p \cdot (1-p)$$

$$(1-p)^2 \cdot p \cdot (1-p) = p \cdot (1-p)^3$$

$$p^3 \cdot (1-p) > p \cdot (1-p)^3$$

- Ergebnis: 3 Fehler sind unwahrscheinlicher als 1 Fehler!

Berechnung der Fehlererkennungs- und korrekturrate

Fehlererkennungsrate: $\boldsymbol{d_{min}} - 1 = \boldsymbol{t_{erk}}$

Fehlerkorrekturrate: $\frac{\boldsymbol{d_{min}-1}}{\boldsymbol{2}} = \boldsymbol{t_{kor}}$

Methoden der Fehlererkennung und Fehlerkorrektur

Wiederholungscode

- Nachrichtenwort „Hase" ⟶ Informationsstellen $k = 4$
- *dreifach* senden (Mindestabstand: 3)

empfangene Nachricht: *HaseNaseHase*

mögliches Ursprungswort	Hammingabstand
HaseHaseHase	1
NaseNaseNase	2

→ geringerer Hammingabstand ist wahrscheinlicher

empfangene Nachricht: *HaseNaseVase*

mögliches Ursprungswort	Hammingabstand
HaseHaseHase	2
NaseNaseNase	2
VaseVaseVase	2

→ keine eindeutige Zuordnung möglich: gleicher Hammingabstand **!**

Codewort: *Hase* ⟶ *HaseHaseHase*

→ Ergebnis:

erkennbare Fehler:	2
korrigierbare Fehler:	1
Codewortlänge:	12

Hamming-Code

- weitere, effizientere Codierungsvariante
- Nachrichtenwort „0110" ⟶ Informationsstellen: $k = 4$
- Ergänzung durch 3 Prüfbits

$a_1 a_2 a_3 a_4\ \boldsymbol{a_5 a_6 a_7}$

4 Informationsstellen + **3 Prüfziffern**

$$a_5 \equiv a_1 + a_3 + a_4 (mod 2)$$

$$a_6 \equiv a_1 + a_2 + a_4 (mod 2)$$

$$a_7 \equiv a_1 + a_2 + a_3 (mod 2)$$

Codewort: 0110 ⟶ 0110 **110**

- Fehlerkorrekturrate:

$$\sum_{i=1}^{r} \binom{7}{i} \leq 2^3 - 1$$

- Mindestabstand: $2 \cdot 1 + 1 = d_{\min} = 3$

→ Ergebnis:

erkennbare Fehler:	2
korrigierbare Fehler:	1
Codewortlänge:	7

Sarah Haunhorst