Codierungstheorie
Grundlagen

Reinhold Hübl

Herbst 2022 / 2. Vorlesung

DHBW
Duale Hochschule
Baden-Württemberg
Mannheim

Grundlagen der Codierung

Die Digitalisierung von Information liefert üblicherweise eine sehr große Datenmenge. Selbst bei sorgfältiger Datenspeicherung und Datenübertragung kommt es daher immer wieder zu Fehlern beim Übertragen oder Auslesen. Ziel der Codierungstheorie ist es, sicherzustellen, dass die Information trotzdem inhaltlich korrekt übermittelt wird.
Der grundsätzliche Ansatz hierzu ist, die Nachricht mit Redundanzen so anzureichern, dass auch aus einer fehlerhaft übertragenen Nachricht noch auf deren Inhalt geschlossen werden kann.

Problemstellung

Das Problem wird dazu in zwei Teile zerlegt:

1. Fehlererkennung: Wie kann ich erkennen, dass ein empfangener Datensatz unkorrekt ist?
2. Fehlerkorrektur: Wie kann ich aus den empfangenen Daten die ursprüngliche Nachricht rekonstruieren?

Beide Teile des Problems können mithilfe der Anreicherung durch Redundanzen gelöst werden. Dabei ist es allerdings wichtig, die redundante Information sinnvoll zu wählen.

Redundanzen

Beispiel

Übertragen werden sollen Bilder, die sich nur aus den Farben weiss $= (0,0)$, hellgrau $= (0,1)$, dunkelgrau $= (1,0)$ und schwarz $= (1,1)$ zusammensetzen.
Wir betrachten zunächst folgende Anreicherung durch Redundanzen

$$\begin{array}{llll} (0,0) & \longmapsto & c_1 = (0,0,0,0,0) & (1,0) \longmapsto c_3 = (1,0,0,0,0) \\ (0,1) & \longmapsto & c_2 = (0,1,0,0,0) & (1,1) \longmapsto c_4 = (1,1,0,0,0) \end{array}$$

Frage: Erfüllt diese Anreicherung ihren Zweck?

Redundanzen

Beispiel

Übertragen werden sollen Bilder, die sich nur aus den Farben weiss $= (0,0)$, hellgrau $= (0,1)$, dunkelgrau $= (1,0)$ und schwarz $= (1,1)$ zusammensetzen.
Wir betrachten zunächst folgende Anreicherung durch Redundanzen

$$\begin{array}{llll} (0,0) & \longmapsto & c_1 = (0,0,0,0,0) & (1,0) \longmapsto c_3 = (1,0,1,0,1) \\ (0,1) & \longmapsto & c_2 = (0,1,0,1,1) & (1,1) \longmapsto c_4 = (1,1,1,1,0) \end{array}$$

Frage: Erfüllt diese Anreicherung ihren Zweck?

Redundanzen

Die erste Anreicherung bringt wenig, denn zwei Nachrichten, die sich nur an einer Stelle unterscheiden, unterscheiden sich auch nach Anreicherung mit Redundanzen nur an einer Stelle. Wird dann diese Stelle fehlerhaft übertragen, so wird eine falsche (aber zulässige) Nachricht empfangen.
Die zweite Anreicherung ist geschickter, da sich je zwei der erzeugten **Codewörter** an mindestens drei Stellen unterscheiden. Wird also eine Stelle falsch übertragen, so unterscheidet sich das entstehende Wort vom korrekten nur an einer Stelle, von allen anderen aber mindestens an zwei Stellen. Daher kann das korrekte Wort (und damit die ursprüngliche Nachricht) rekonstruiert werden.
Ziel der Anreicherung mit Redundanzen muss es also sein, Datensätze zu erzeugen, die sich an möglichst vielen Stellen unterscheiden. Die Anzahl der anzustrebenden Unterschiede hängt vom Verwendungszweck ab!

Fehlererkennung

Für viele Zwecke ist es ausreichend, zu erkennen, dass bei der Übertragung der Daten ein Fehler aufgetreten ist. Um zu erkennen, dass bei der Übertragung der Daten ein (und nur ein) Fehler aufgetreten ist, reicht es aus, wenn sich die übertragenen Datensätze an **zwei** Stellen unterscheiden. Das übliche Verfahren zur Erzeugung eines Unterschieds ist die Ergänzung eines jeden Datenblocks um eine Prüfziffer.

Beispiel

Mithilfe des Paritätsprüfbits kann überprüft werden, ob beim Auslesen eines Bytes ein Fehler aufgetreten ist.

Fehlererkennung

Wir betrachten Informationsblöcke $m = (a_1, \ldots, a_5)$ bestehend aus jeweils 5 Ziffern $a_1, \ldots, a_5 \in \{0, \ldots, 9\}$ und ergänzen den Block um ein Prüfzeichen a_6, und zwar so, dass

$$6 \cdot a_1 + 5 \cdot a_2 + 4 \cdot a_3 + 3 \cdot a_4 + 2 \cdot a_5 + a_6 \equiv 0 \bmod 11$$

$$6 \cdot 1 + 5 \cdot 2 + 4 \cdot 3 + 3 \cdot 4 + 2 \cdot 5 + a_6 \equiv 0 \bmod 11$$

also muss $50 + a_6$ durch 11 teilbar sein, und damit ist $a_6 = 5$. Das angereicherte Wort ist also

$$c = (1, 2, 3, 4, 5, 5)$$

(Für den Fall, dass $a_6 = 10$ sein sollte, schreiben wir $a_6 = X$.)

Fehlererkennung

Übung

Wir betrachten Informationsblöcke $m = (a_1, \ldots, a_5)$ bestehend aus jeweils 5 Ziffern $a_1, \ldots, a_5 \in \{0, \ldots, 9\}$ und ergänzen den Block um ein Prüfzeichen a_6, und zwar so, dass

$$6 \cdot a_1 + 5 \cdot a_2 + 4 \cdot a_3 + 3 \cdot a_4 + 2 \cdot a_5 + a_6 \equiv 0 \bmod 10$$

erkannt werden?

Fehlererkennung - ISBN–10

Das klassische Verfahren eines fehlererkennenden Codes ist die Buchkennung ISBN–10. Das ist eine zehnstellige Kennung, die für jeden Buchtitel eindeutig ist. Dabei enthalten die ersten neun Stellen alle Informationen über das Buch, und das zehnte Zeichen ist eine Prüfziffer. Dieses Prüfzeichen a_{10} wird aus der Information $(a_1, a_2, \ldots, a_9)$ wie folgt berechnet:

1. Bestimme die Summe

$$s = a_1 + 2a_2 + 3a_3 + \cdots + 9a_9 = \sum_{l=1}^{9} l \cdot a_l$$

2. Dividiere s durch 11 mit Rest

$$s = q \cdot 11 + r$$

   mit einer ganzen Zahl $r \in \{0, 1, \ldots, 10\}$.

3. Falls $r = 10$ setze $a_{10} = X$, andernfalls setze $a_{10} = r$.

Fehlererkennung - ISBN–10

Beispiel

Die Information über ein Buch ist abgelegt in den neun Ziffern $3{-}528{-}27224$.

Dieses Prüfzeichen a_{10} berechnet sich dann wie folgt:

1. Bestimme die Summe

$$s = 3 + 2 \cdot 5 + 3 \cdot 2 + 4 \cdot 8 + 5 \cdot 2 + 6 \cdot 7 + 7 \cdot 2 + 8 \cdot 2 + 9 \cdot 4 = 169$$

2. Dividiere 169 durch 11 mit Rest

$$168 = 15 \cdot 11 + 4$$

3. Setze $a_{10} = 4$.

Der ISBN–10–Code dieses Buchs ist also $3{-}528{-}27224{-}4$.

Fehlererkennung - ISBN–10

Übung

Für ein Buch wird der ISBN–10–Code $3-528-17277-7$ übermittelt.
Überprüfen Sie, ob der Code korrekt sein kann.

Fehlererkennung - EAN bzw. ISBN–13

Für die Warencodierung (auch bei Büchern) hat sich die European Article Number EAN durchgesetzt. Inzwischen wird Sie auch für Bücher (als ISBN–13–Codierung) verwendet. Sie besteht aus 13 Ziffern wie folgt:

- Die 9 informationstragenden Stellen aus ISBN–10 bleiben.
- Vorangestellt werden die Ziffern 978 bzw. 979 (als Kennung für die Produktgruppe Buch/Printmedien).
- Die Prüfziffer a_{13} wird so berechnet, dass

$$a_1 + 3 \cdot a_2 + a_3 + 3 \cdot a_4 + \cdots + a_{11} + 3 \cdot a_{12} + a_{13} \equiv 0 \mod 10$$

Die ISBN–13-Artikelkennung ist dann $(a_1, a_2, \ldots, a_{12}, a_{13})$.

Fehlererkennung - ISBN–13

Beispiel

Die Information über ein Buch ist wieder abgelegt in den neun Ziffern 3−528−27224. Mit Produktgruppenkennung haben wir also die Information 978−3−528−27224.
Dieses Prüfzeichen a_{13} berechnet sich dann wie folgt:

1. Bestimme die Summe

$$s = 9 + 8 + 5 + 8 + 7 + 2 + 3 \cdot (7 + 3 + 2 + 2 + 2 + 4) = 99$$

2. a_{13} ist so zu wählen, dass $99 + a_{13}$ durch 10 teilbar ist, also $a_{13} = 1$

Der ISBN–13–Code dieses Buchs ist also 978−3−528−27224−1.

Fehlererkennung - ISBN–13

Übung

Für ein Buch wird der ISBN–13–Code $978-3-528-17217-6$ übermittelt. Überprüfen Sie, ob der Code korrekt sein kann.

Fehlerkorrektur

Unterscheiden sich zwei Codewörter nur an zwei Stellen, so ist eine Fehlerkorrektur im allgemeinen nicht möglich.
Für eine Fehlerkorrektur ist es notwendig, dass sich zwei Codewörter an mindestens drei Stellen unterscheiden, und daher ist es notwendig, dass mindestens zwei Prüfzeichen ergänzt werden.

Beispiel

Wir betrachten Informationsblöcke $m = (a_1, \ldots, a_5)$ bestehend aus jeweils 5 Ziffern $a_1, \ldots, a_5 \in \{0, \ldots, 9, X = 10\}$ und ergänzen den Block um zwei Prüfzeichen a_6 und a_7, und zwar so, dass

$$a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 \equiv 0 \mod 11$$

und so, dass

$$a_1 + 2 \cdot a_2 + 3 \cdot a_3 + 4 \cdot a_4 + 5 \cdot a_5 + 6 \cdot a_6 + 7 \cdot a_7 \equiv 0 \mod 11$$

Fehlerkorrektur

Beispiel

Ist etwa $m = (1, 2, 3, 4, 5)$, so sind a_6 und a_7 so zu wählen, dass

$$1 + 2 + 3 + 4 + 5 + a_6 + a_7 \equiv 0 \mod 11$$

und so, dass

$$1 + 2 \cdot 2 + 3 \cdot 3 + 4 \cdot 4 + 5 \cdot 5 + 6 \cdot a_6 + 7 \cdot a_7 \equiv 0 \mod 11$$

also müssen $15 + a_6 + a_7$ und $55 + 6 \cdot a_6 + 7 \cdot a_7$ durch 11 teilbar sein, und damit ist $a_6 = 5$ und $a_7 = 2$.

$$c = (1, 2, 3, 4, 5, 5, 2)$$

Fehlerkorrektur

Beispiel

Wird nun $a = (1, 4, 3, 4, 5, 5, 2)$ empfangen (tritt also ein Fehler beim Auslesen auf), so werden zunächst die beiden Prüfsummen bestimmt:

$$\mathrm{PS}_1 = 1 + 4 + 3 + 4 + 5 + 5 + 2 \mod 11 = 2$$

(damit ist schon klar, dass die Übertragung fehlerhaft war, denn $\mathrm{PS}_1 \neq 0$) und

$$\mathrm{PS}_2 = 1 + 2 \cdot 4 + 3 \cdot 3 + 4 \cdot 4 + 5 \cdot 5 + 6 \cdot 5 + 7 \cdot 2 \mod 11 = 4$$

und auch das bestätigt, dass ein Fehler vorliegt.
Der entscheidende Punkt für die Fehlerkorrektur ist nun die Gewichtung der Fehler in den beiden Prüfsummen. In die erste Summe geht der Fehler immer einmal ein, unabhängig davon, an welcher Stelle er auftritt, in die zweite Prüfsumme geht der Fehler mit einer von der Fehlerstelle abhängigen Prüfsumme in die Gewichtung ein.

Fehlerkorrektur

Beispiel

Als Verhältnis der beiden Prüfsummen (wenn wir $\mathrm{mod}\ 11$ rechnen):

$$\begin{aligned}
&\text{Fehler an der Stelle 1:} && \text{Prüfsumme 2} &&\equiv&\ 1 \cdot \text{ Prüfsumme 1}\\
&\text{Fehler an der Stelle 2:} && \text{Prüfsumme 2} &&\equiv&\ 2 \cdot \text{ Prüfsumme 1}\\
&\text{Fehler an der Stelle 3:} && \text{Prüfsumme 2} &&\equiv&\ 3 \cdot \text{ Prüfsumme 1}\\
&\text{Fehler an der Stelle 4:} && \text{Prüfsumme 2} &&\equiv&\ 4 \cdot \text{ Prüfsumme 1}\\
&\text{Fehler an der Stelle 5:} && \text{Prüfsumme 2} &&\equiv&\ 5 \cdot \text{ Prüfsumme 1}\\
&\text{Fehler an der Stelle 6:} && \text{Prüfsumme 2} &&\equiv&\ 6 \cdot \text{ Prüfsumme 1}\\
&\text{Fehler an der Stelle 7:} && \text{Prüfsumme 2} &&\equiv&\ 7 \cdot \text{ Prüfsumme 1}
\end{aligned}$$

In unserem Beispiel haben wir

$$\mathrm{PS}_2 = 4 = 2 \cdot 2 = 2 \cdot \mathrm{PS}_1$$

also war der Fehler an der zweiten Stelle.

Fehlerkorrektur

Beispiel

Wir wissen also, dass der Fehler an der Stelle 2 ist, können daher annehmen, dass

$$a_1 = 1,\ a_3 = 3,\ a_4 = 4,\ a_5 = 5,\ a_6 = 5,\ a_7 = 2$$

korrekt sind. Prüfgleichung zu

$$1 + x + 3 + 4 + 5 + 5 + 2 \equiv 0 \mod 11$$

$$20 + x \equiv 0 \mod 11$$

Hierfür ist dann auch die zweite Prüfsummenbedingung erfüllt.

Grundbegriffe der Codierung

Grundbegriffe aus der Codierungstheorie

- Grundbausteine der Nachrichten sind **Buchstaben** aus einem fixierten **Alphabet** $\mathbb{A}$ (etwa $\mathbb{A} = \{0, 1\}$), also einer endlichen Menge $\mathbb{A}$ mit q Elementen.
- Die Buchstaben werden zu Gruppen (**Wörtern**) einer fixierten Länge k zusammengefasst und bilden den Nachrichtenraum $M = \mathbb{A}^k$ (etwa $M = \{0, 1\}^2$). Versendet werden sollen Folgen von Wörtern.
- Die **Nachrichtenwörter** $m \in M$ werden nach fixierten Regeln mit Redundanzen angereichert und in **Codewörter** c einer festen Länge n (über dem selben Alphabet) umgewandelt (etwa $m \in \{0, 1\}^2 \longmapsto c \in \{0, 1\}^5$).

Grundbegriffe der Codierung

Diese Vorschrift heißt **Codierung** und die Menge C der angereicherten Nachrichten $c \in \mathbb{A}^n$ bezeichnet man als Code $C \subseteq \mathbb{A}^n$.

Definition

Ein q–**adischer Code** C ist eine nicht–leere Teilmenge $C \subseteq \mathbb{A}^n$.

Die Zahl $k = \log_q(|C|)$ heißt **logarithmische Kardinalität** von C und n heißt **Blocklänge** von C.

Die Zahl $R = \frac{k}{n}$ heißt **Informationsrate** von C.

Wir nennen C dann einen $[n, k]_q$–Code oder $[n, k]$–Code, falls $\mathbb{A}$ klar ist.

Grundbegriffe der Codierung

Definition

Die **Hammingmetrik** d auf $\mathbb{A}^n$ ist gegeben durch

$$d((a_1, \ldots, a_n), (b_1, \ldots, b_n)) = |\{i \in \{1, \ldots, n\} | \, a_i \neq b_i\}|$$

Bemerkung

Die Hammingmetrik von zwei n-Tupeln a und b bezeichnet die Anzahl der Stellen, an denen sich a und b unterscheiden.

Definition

Ist C ein $[n, k]_q$–Code, so heißt

$$d(C) := \min\{d(a, b) | a, b \in C, a \neq b\}$$

heißt **Minimalabstand** oder **Zuverlässigkeit** von C.

Grundbegriffe der Codierung

Bemerkung

Der Minimalabstand $d(C)$ gibt an, an wie vielen Stellen sich zwei Codewörter mindestens unterscheiden.
Je größer $d(C)$ desto mehr Fehler kann C erkennen bzw. korrigieren.

Beispiel

Der Code

$$C = \{(0,0,0,0,0),(1,0,1,0,1),(0,1,0,1,1),(1,1,1,1,0)\}$$

hat Minimalabstand $d(C) = 3$

Grundbegriffe der Codierung

Ist C ein $[n,k]_q$–Code mit Minimalabstand $d = d(C)$, so heißt

$$t = \lfloor \frac{d-1}{2} \rfloor$$

die **Fehlerkorrekturschranke** von C.

Bemerkung

Ein Code C mit Fehlerkorrekturschranke t ist in der Lage, bis zu $2t$ Fehler zu erkennen und bis zu t Fehler in einem Wort zu korrigieren.

Definition

Ein **Körper** ist eine nicht–leere Menge K mit zwei ausgezeichneten Elementen 0 und 1, wobei $0 \neq 1$, und mit zwei inneren Verknüpfungen (also Abbildungen) $+$ und $\cdot$

$$\begin{aligned} + &: K \times K \longrightarrow K, \qquad (a,b) \longmapsto a+b \\ \cdot &: K \times K \longrightarrow K, \qquad (a,b) \longmapsto a \cdot b \end{aligned}$$

so dass gilt

1. $(K,+)$ ist eine kommutative Gruppe mit neutralem Element 0.
2. $(K \setminus \{0\}, \cdot)$ ist eine kommutative Gruppe mit neutralem Element 1.
3. Es gilt das Distributivgesetz, dh. für alle $a,b,c \in K$ gilt

$$a \cdot (b+c) = a \cdot b + a \cdot c$$

Beispiele

Beispiel

$\mathbb{R}$, $\mathbb{Q}$ und $\mathbb{C}$ sind Körper

Beispiel

Die Menge $\mathbb{F} = \{0, 1\}$ mit

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

ist ein Körper. Dieser Körper wird auch mit $\mathbb{F}_2$ bezeichnet.

Definition

Ein endlicher Körper K ist ein Körper $(K, +, \cdot)$ mit $|K| < \infty$.

Beispiele

Beispiel

Die Menge $\mathbb{Z}_n$ oder $\mathbb{Z}/n\mathbb{Z}$ ist der Menge der Äquivalenzklassen ganzer Zahlen modulo n. Er kann beschrieben werden durch die Repräsentanten $0, 1, \ldots, n-1$. Auf diesen Repräsentanten definieren wir eine Addition $+$ und eine Multiplikation $\cdot$ explizit wie folgt:

$$\begin{aligned} x + y &= (x + y) \mod n \\ x \cdot y &= x \cdot y \mod n \end{aligned}$$

dh. wir addieren bzw. multiplizieren die Repräsentanten zunächst in $\mathbb{Z}$, dividieren das Ergebnis mit Rest durch n und nehmen diesen Rest als Ergebnis der Addition bzw. der Multiplikation.

Dadurch wir $\mathbb{Z}/n\mathbb{Z}$ zum (kommutativen) Ring.

Beispiele

Beispiel

In $\mathbb{Z}/12\mathbb{Z}$ gilt

- $3 + 4 = 7$.
- $9 + 10 = 7$.
- $4 + 8 = 0$, dh. $8 = -4$.
- $2 \cdot 3 = 6$.
- $5 \cdot 6 = 6$.
- $3 \cdot 4 = 0$.

Folgerung

$\mathbb{Z}/12\mathbb{Z}$ ist kein Körper.

Beispiele

Beispiel

Es ist $\mathbb{Z}/2\mathbb{Z} = \mathbb{F}_2$ der Körper von oben mit zwei Elementen.

Übung

Ist auch $\mathbb{Z}/15\mathbb{Z}$ ein Körper?

Primkörper

Der Ring $\mathbb{Z}/n\mathbb{Z}$ kann also ein Körper sein, muss aber nicht.

Satz

Genau dann ist $\mathbb{Z}/n\mathbb{Z}$ ein Körper, wenn n eine Primzahl ist.

Bezeichnung

Ist p eine Primzahl, so schreiben wir $\mathbb{F}_p$ für $\mathbb{Z}/p\mathbb{Z}$ und nennen $\mathbb{F}_p$ den **Primkörper** mit p Elementen.

Beispiel

$\mathbb{F}_2$, $\mathbb{F}_3$ und $\mathbb{F}_5$ sind Körper.