



Siebter Korrespondenzbrief

Codierungstheorie

Einleitung...

1 Einführung

Eine besonders einfache Methode Fehler bei der Übertragung zu erkennen, ist die Folgende: Man sendet die zu übertragende Nachricht einfach zweimal hintereinander. Passiert beim Übertragen der Nachricht ein Fehler, so erkennt der Empfänger das daran, dass die beiden Hälften der Nachricht nicht mehr gleich sind.

Beispiel 1. Wollen wir bspw. das Wort JUNI übertragen, so senden wir stattdessen das Codewort JUNIJUNI. Passiert beim Übertragen ein einzelner Fehler (ein Buchstabe wird verfälscht), so erhält der Empfänger bspw. das Wort JULIJUNI und kann direkt sehen, dass ein Fehler passiert sein muss.

Wir können es also erkennen, wenn bei der Übertragung ein Fehler passiert - wir können diesen Fehler aber nicht korrigieren (denn wir wissen ja nicht welche der beiden Hälften des Codewortes die korrekte ist). Außerdem können wir nur dann sicher sein, Fehler zu erkennen, wenn höchstens ein Fehler passiert. Treten bei der Übertragung zwei oder mehr Fehler auf, so kann es passieren, dass der Empfänger das nicht mehr erkennen kann:

JUNIJUNI \longrightarrow JULIJULI

Aufgabe 1.

Kannst du das oben beschriebene Verfahren so anpassen, dass ein einzelner Fehler vom Empfänger nicht nur erkannt, sondern auch korrigiert werden kann?

Kannst du das Verfahren so anpassen, dass er auch zwei Fehler noch in jedem Fall erkennen kann?

Diese einfachen Codierungsverfahren funktionieren zwar grundsätzlich, haben aber einen großen Nachteil. Die zu übertragenden Nachrichten werden erheblich größer als die eigentliche Nachricht. Ein zentrales Ziel der Codierungstheorie ist es daher möglichst effiziente Codierungsverfahren zu finden, die die Nachrichten also nur wenig größer machen, aber es trotzdem erlauben Fehler zu erkennen (und evtl. sogar zu korrigieren).

Bevor wir uns auf die Suche nach solchen Verfahren machen können, müssen wir erst noch genauer festlegen, wie wir die Effizienz eines Codierungsverfahrens bestimmen wollen. Dazu werden wir ab sofort nur noch Verfahren der folgenden Form betrachten:

1. Teile die zu übertragende Nachricht in Stücke gleicher Länge ein. Diese Stücke nennen wir *Wörter*.
2. Ersetze jedes der Wörter durch ein ihm zugeordnetes *Codewort* - diese Codewörter müssen dabei ebenfalls alle eine einheitliche Länge haben.
3. Schreibe die Codewörter wieder hintereinander und erhalte so die zu übertragende Nachricht.

Beispiel 2. Wollen wir zum Beispiel die Verdopplungscodierung von oben mit einer Wortlänge von 4 anwenden, um die Nachricht **DIESISTEINETESTNACHRICHT** zu übertragen, so funktionieren die Schritte wie folgt:

1. Teile die Nachricht in Wörter der Länge 4: **DIES ISTE INET ESTN ACHR ICHT**
2. Ersetze die Wörter durch ihre Codewörter: **DIES** \rightarrow **DIESDIES**, **ISTE** \rightarrow **ISTEISTE**,
...
3. Erhalte die codierte Nachricht: **DIESDIESISTEISTEINETINETESTNESTNACHRNACHRICHTICHT**

Aufgabe 2.

Selber ausprobieren?

Für Verfahren der oben beschriebenen Form, können wir nun ganz einfach bestimmen, wie effizient sie sind:

Definition 3. Als *Informationsrate* eines bestimmten Codierungsverfahrens bezeichnet man das Verhältnis aus der Länge der Codewörter zur Länge der Nachrichtenwörter, also die Zahl

$$\frac{\text{Länge eines Codewortes}}{\text{Länge eines Nachrichtenwortes}}$$

Je kleiner diese Zahl ist, desto effizienter ist das Verfahren.

Beispiel 4. Für das Verdopplungscodierungsverfahren aus Beispiel 2 ist die Informationsrate also:

$$\frac{\text{Länge eines Codewortes}}{\text{Länge eines Nachrichtenwortes}} = \frac{4}{2} = 2$$

Eine mit diesem Verfahren codierte Nachricht benötigt also doppelt so viel Platz wie die ursprüngliche Nachricht.

Aufgabe 3.

Bestimme die Informationsrate für deine in Aufgabe 1 gefundenen Verfahren.

Als weitere Vereinfachung werden wir im restlichen Brief nur noch Nachrichten (und Codierungen) betrachten, die ausschließlich aus Zahlen bestehen. Das ist aber keine wirkliche Einschränkung, denn wie du schon im Kryptographie-Brief gesehen hast, lassen sich Buchstaben auch leicht durch Zahlen darstellen.

2 Fehlererkennung

Als erstes wollen wir uns nun mit Verfahren beschäftigen, die es uns erlauben einzelne Fehler zu erkennen. Das Verfahren soll dabei die folgende Form haben: Wir nehmen das Wort aus der Nachricht und codieren es, indem wir noch „irgendetwas“ daran anhängen. Dieses „irgendetwas“ soll es dem Empfänger dann erlauben einen möglicherweise aufgetretenen Fehler zu erkennen.

$$1847 \longrightarrow 1847___$$

Im ersten Kapitel war dieses „irgendetwas“ zum Beispiel einfach wieder das Wort selbst. Das ist aber ziemlich ineffizient, da das Wort dadurch gleich doppelt so lang wird. Besser wäre es da schon nur die Quersumme des Wortes (das ist die Summe der einzelnen Ziffern) anzuhängen. Sind die Wörter vier Ziffern lang, so kann die Quersumme höchstens $9 + 9 + 9 + 9 = 36$ sein - uns genügt also ein zweistelliges Anhängsel.

Beispiel 5. Wollen wir die Nachricht 1701 übertragen, so bilden wir die Quersumme $1 + 7 + 0 + 1 = 9$ und erhalten damit das Codewort 1701**09**. Passiert beim Übertragen nun ein Fehler, so erhält der Empfänger beispielsweise die Nachricht **8**70109. Dieser kann nun leicht erkennen, ob ein Fehler aufgetreten ist, indem er einfach wieder die Quersumme der ersten vier Ziffern bildet ($8 + 7 + 0 + 1 = 16$) und mit den letzten beiden Ziffern vergleicht. Stimmen diese beiden Zahlen nicht überein, so weiß der Empfänger, dass beim Übertragen ein Fehler passiert sein muss.

Aufgabe 4.

Du bist der Empfänger und erhältst die folgenden Codewörter (die ursprünglichen Wörter hatten die Länge 4):

$$172919 \quad - \quad 422416 \quad - \quad 939124 \quad - \quad 314109$$

Prüfe für jedes der Wörter, ob beim Übertragen ein Fehler aufgetreten ist. Falls nein, gib das ursprüngliche (uncodierte) Wort an. Falls ja, gib zwei Möglichkeiten an, wie das ursprüngliche Wort gelautet haben könnte (unter der Annahme, dass bei der Übertragung höchstens eine Ziffer verändert wurde).

Bei folgendem Codewort, ist bei der Übertragung etwas spezielles passiert - kannst du erklären was?

$$739150$$

Funktioniert dieses Verfahren auch? D.h. können wir damit wirklich jeden einzeln auftretenden Fehler erkennen? Ja, denn wann immer eine einzelne Ziffer verändert wird, so ändert sich auch die Quersumme der ganzen Zahl.

Ist das Verfahren effizient? Nun, die Informationsrate ist $\frac{\text{Länge eines Codewortes}}{\text{Länge eines Nachrichtenwortes}} = \frac{6}{4} = 1,5$ - also auf jeden Fall schonmal besser als bei dem bisherigen Verfahren.

Tatsächlich geht es aber sogar noch besser: Denn wenn nur eine einzelne Ziffer in der Nachricht verfälscht wird, dann kann sich die Quersumme doch höchstens um 9 verändern. Es genügt also völlig, wenn wir uns nur die Einerstelle der Quersumme merken (oder anders ausgedrückt: Den Rest der Quersumme bei Division durch 10).

Beispiel 6. Um mit diesem Verfahren die Nachricht 8245 zu übertragen, bilden wir also erneut die Quersumme $8 + 2 + 4 + 5 = 19$ und nehmen deren Einerstelle 9. So erhalten wir das Codewort 8245**9**.

Aufgabe 5.

Du empfängst die folgenden Codewörter - bei welchen von ihnen muss ein Fehler bei der Übertragung passiert sein?

23836 – 99999 – 47629 – 38425

Aufgabe 6. B

estimme die Informationsrate dieses Codierungsverfahrens.

Zusatzaufgabe: Kannst du das Verfahren irgendwie so verändern, dass die Informationsrate noch niedriger wird?

Rest 10

Vertauschung

Beispiel 7. ISBN

Aufgabe 7.

Bestimme Informationsrate von ISBN

Aufgabe 8.

Prüfe, ob folgende ISBN-Nummern korrekt abgetippt wurden.

Beispiel 8. Geldscheine

3 Fehlerkorrektur und Güteanalyse

Definition 9. Hamming-Abstand

Zusammenhang zwischen Hamming-Abstand und Fehlererkennung und -korrektur.

Beispiel 10. ISBN und Parity-Check

2-dimensionaler Parity-Check

Hamming-Abstand-Bestimmen? (Zumindest für $n=2$)

Aufgabe 9.

Bestimme Informationsrate (vgl. mit Abschnitt 1)

Aufgabe 10.

Konkrete Beispiele für Codierung/Decodierung

Perfekte/MDS-Codes?