Foutenverbetering op de compact disc

Foutenverbetering op de Compact Disc*

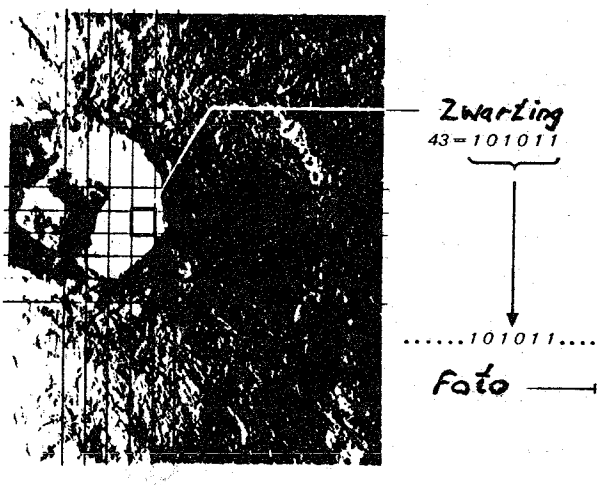
J. H. van Lint

Inleiding

Om in één uur iets te vertellen over *fouten – verbeterende codes* onder de aanname dat het onderwerp geheel nieuw is voor de meeste toehoorders is niet eenvoudig. Veel zal worden uitgelegd via voorbeelden, vaak in combinatie met oversimplificatie t.o.v. wat in de praktijk geschiedt. De situatie die we ons moeten voorstellen is de volgende.
Er wordt *informatie* aangeboden in de vorm van een zeer lange rij bestaande uit twee soorten symbolen, die we 0 en 1 noemen.
(We denken aan de punt en de streep bij Morse, al of niet een rookwolkje bij Indianensignalen, of aan geluid van twee verschillende frequenties. Meer voorbeelden volgen.) Deze informatie zal van een *zender* naar een *ontvanger* worden gebracht via een zgn. *kanaal*. De lezer die hierbij behoefte heeft aan een beeld denke aan een (gestoorde) radiozender. Helaas heeft dit kanaal de hinderlijke eigenschap dat het af en toe een door de zender aangeboden 0 (resp. 1) bij de ontvanger als een 1 (resp. 0) aflevert. We noemen dit 'fouten'. Hierbij onderscheiden we twee mogelijke situaties:

i) het optreden van fouten is een stochastisch proces: met een zekere kans p (zeg $p = 0{,}01$) treedt een fout op ('random errors');
ii) af en toe treden hele series fouten vlak bij elkaar op (dit noemen we 'burst errors').

Voor we verder gaan eerst twee voorbeelden uit de praktijk die tonen hoe zo'n rij nullen en enen kan ontstaan. Beide voorbeelden waren spectaculaire successen.
Bij de satellietfoto's die zijn gemaakt van Mars, Saturnus en Jupiter werd de foto verdeeld in zeer veel kleine vierkantjes (genaamd pixels). Van elk vierkantje wordt de zwartingsgraad bepaald, uitgedrukt in een schaal van (bijvoorbeeld) 0 t/m 63. Deze getallen worden geschreven in het tweetallig stelsel:
$43 = 101011$
$(= 1 \cdot 32 + 0 \cdot 16 + 1 \cdot 8 + 0 \cdot 4 + 1 \cdot 2 + 1.1)$.
Zó geeft één foto aanleiding tot een rij van vele miljoenen nullen en enen. Het 'kanaal' was hier de zender in de satelliet, de ruimte en de ontvanger plus versterker op aarde.

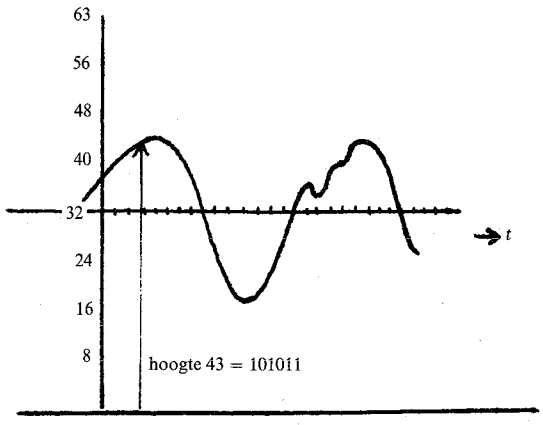


Figuur 1

Het tweede voorbeeld is de *compact disc* (volledig: Compact Disc Audio System). De informatie (muziek) is op de plaat aangebracht als een lange spiraal bestaande uit twee soorten objecten, nl. wél een putje ($0{,}24\,\mu m^2$, $0{,}12\mu m$ diep) of niet zo'n putje (= een 'land'). Hierbij kan men een putje een 1 noemen, een land een 0. Een disc bevat ongeveer 5.10^9 bits (een bit is 0 of 1).
Het lezen van de informatie (afspelen van de plaat) geschiedt door een laser (een slechts in 't Nederlands optredende woordspeling!). De diepte van de putjes is zodanig dat door interferentie bij de putjes veel minder licht wordt gereflecteerd dan bij de

* voordracht gehouden op de 12de gemeenschappelijke studiedag NVvW-VVWL; 28.3.87

landen en zo meet het apparaat of er wel of niet een putje is. Behalve dat de plaat bij het lezen niet slijt is nog één van de voordelen van het optisch lezen het feit dat de informatie beschermd is door een ongeveer 1 millimeter dikke doorzichtige laag waardoor de laser weinig last heeft van stof op de plaat en kleine oppervlakte-beschadigingen. Desondanks treden er fouten op bij het lezen. Hoe men deze fouten kan opsporen en verbeteren, met als gevolg een prachtige kwaliteit muziek zonder de hinderlijke tikjes die we van beschadigde grammofoonplaten kennen, is het onderwerp van deze voordracht.
Hoe komt de rij nullen en enen tot stand? Men moet zich een continu signaal voorstellen als in figuur 2 (hetgeen bijvoorbeeld de ingangsspanning van een audioversterker voorstelt als functie van de tijd), hier in een schaal 0-63.
Om de τ seconden (in de praktijk is $\tau = 1/44100$, d.w.z. 'sampling frequency' = 44.1 kHz) wordt de sterkte van het signaal gemeten (bij stereo zelfs twee tegelijk) in een schaal die resulteert in een voorstelling met 16 bits. Zo ontstaat ook hier een lange rij nullen en enen. We merken op dat bij het spelen van de plaat meer dan een *miljoen* nullen en enen per seconde worden gelezen. Allerlei toevalligheden, beschadigingen e.d. kunnen wat wij boven een random error noemden veroorzaken terwijl krassen en vlekken aanleiding zijn tot burst errors.
De lezer die meer wil weten over technische aspecten van de CD, foutenverbetering enz. verwijzen we naar [1], [2].

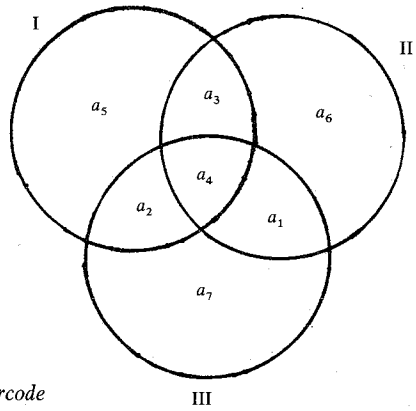


Figuur 2

Foutenverbetering

Een simpele manier om fouten te verbeteren maakt gebruik van een oud didactisch principe: als 't gehoor iets niet begrepen heeft, dan *herhaalt* men de bewering! Als we in plaats van een 0 (resp. 1) 00000 zenden (resp. 11111) dan mogen van deze vijf symbolen er twee door het kanaal veranderd worden zonder dat dit de ontvanger in moeilijkheden brengt. Hij laat bij ieder vijftal de meerderheid beslissen of het 0 of 1 moet zijn. Een aardig vraagstuk voor een les over waarschijnlijkheidsrekening is het volgende. We zenden eerst 43 = 101011 over een kanaal met foutenkans $p = 0{,}01$. Wat is de kans op goed overkomen? Nu gebruiken we de boven genoemde 'herhalingscode'. Wat is nu de kans op goed overkomen? De verbetering is geweldig maar de tol die we betalen ook! Het duurt nu *vijf* keer zo lang om een boodschap over te brengen. We zullen hiervoor een maat invoeren. In ons voorbeeld zeggen we dat de *informatiesnelheid* gelijk is aan $\frac{1}{5}$. Iedere vijf bits die we ontvangen geven slechts één bit informatie. De rest is 'redundantie' die ons helpt bij het verbeteren van fouten maar verder nutteloos is.
We bespreken nu meteen de hoofdprincipes van *codering* voor foutenverbetering. Verdeel de informatiestroom in 'blokken' van elk k bits. Via het zgn. coderingsalgoritme wordt een k-tal afgebeeld op een n-tal bits ($n > k$).
In dit geval is de informatiesnelheid k/n. Hoe dichter bij 1 hoe liever! Bij de CD is deze snelheid 0,75. Hoe we deze afbeelding van $\{0,1\}^n$ moeten construeren is één van de hoofdonderwerpen in de coderingstheorie.
Wat we moeten nastreven is eenvoudig uit te leggen m.b.v. onze eigen taal. Als we een lang Nederlands woord lezen met één of enkele drukfouten er in (bijvoorbeeld: wiqkunzige) dan kunnen we i.h.a. wel zien welke letters fout zijn en ze verbeteren. Als het lukt dan is dat omdat wij slechts één Nederlands woord kennen dat lijkt op wat er gedrukt staat. Beschouw nu twee rijtjes $\underline{x} = (x_1, x_2, \ldots, x_n)$ en $\underline{y} = (y_1, y_2, \ldots, y_n)$ uit $\{0,1\}$. We definiëren de *afstand* $d(\underline{x}, \underline{y})$ van deze rijtjes door $d(\underline{x}, \underline{y}) := |\{i \mid x_i \neq y_i\}|$, met andere woorden: $d(\underline{x}, \underline{y})$ is het aantal plaatsen waarin de rijtjes x en y verschillen.
De rijtjes $\underline{x}$ die kunnen ontstaan uit een aangeboden k-tal informatiebits noemen we *codewoorden*.

Als ieder tweetal verschillende codewoorden afstand $\geq 2e + 1$ heeft, dan kunnen we ons tot e fouten per verzonden woord permitteren zonder dat de ontvanger in moeilijkheden komt.
We spreken dan van een *e-fouten – verbeterende code.*
Het voorafgaande formuleren we nu iets strenger. We kiezen een *alfabet* van q symbolen; in de praktijk is dit steeds een eindig lichaam (veld) $\mathbb{F}_q$ [3], [4], (voor de CD is gekozen voor $\mathbb{F}_{28}$, een lichaam waarin iedere letter overeenkomt met een rijtje van 8 bits). Een $[n, k]$ code C is een k-dimensionale lineaire deelruimte van de vectorruimte $(\mathbb{F}_q)^n$, waarin we de afstand net zo definiëren als boven, nu met x_i en y_i symbolen uit $\mathbb{F}_q$. De *minimum afstand* van C is het minimum van $d(x, y)$ over alle paren verschillende vectoren (= woorden) uit C. Is $d = 2e + 1$ dan is C e-fouten verbeterend. Een extra voordeel hierbij is dat dicht bij elkaar liggende fouten in de bits slechts één of twee symbolen beïnvloeden (die immers uit acht bits bestaan).



Figuur 3 Kindercode

Als illustratie van het voorafgaande behandelen we de 'kindercode' van McEliece en geven daarna de abstracte formulering van hetzelfde idee. Op een school moeten de leerlingen *vier* vragen beantwoorden met ja of nee, de antwoorden op een briefje schrijven en inleveren. De briefjes worden naar de leraar gebracht door een onaangename jongen, die één van de meisjes dwarszit door op weg naar de leraar één van haar antwoorden snel en stiekem te veranderen! Als wapen hiertegen spreekt het (slimme!) meisje het volgende met de leraar af: (zowaar een nuttige toepassing van een Venn-diagram).

De vier antwoorden a_1 t/m a_4 worden geplaatst in de figuur. Daarna komen er nog drie keer 'ja' of 'nee' op a_5 t/m a_7 en wel zó dat elk der cirkels een *even* aantal keren ja en nee heeft. Stel dat de vervelende jongen a_1 verandert.
De pariteit van 'ja' in de cirkels I, II, III is daarna even, oneven, oneven, d.w.z. II en III zijn verkeerd. Daar a_1 het enige antwoord is dat in II én in III maar niet in I ligt, weet de leraar dat a_1 is veranderd. Hij *verbetert* de 'fout'. Nu, hetzelfde in de taal van de algebra.
Laat C de 4-dimensionale deelruimte zijn van $(\mathbb{F}_2)^7$ opgespannen door de rijen van de matrix G, met

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Voeg aan de 'informatie' $\underline{a} = (a_1, a_2, a_3, a_4)$ toe het 'codewoord' $(a_1, a_2, a_3, a_4, a_5, a_6, a_7) = \underline{a}\,G \in (\mathbb{F}_2)^7$. Zo ontstaan 16 mogelijke codewoorden met onderlinge afstand ≥ 3. Dus is deze code 1 – fout – verbeterend. De lezer dient nu zelf na te gaan dat hier opnieuw de 'kindercode' is beschreven, nu in z'n officiële gedaante nl. als de [7,4] *Hamming code.*
Oefening Ga na dat als de jongen kans ziet i.p.v. één antwoord te veranderen er twee uit te gommen, ook dan het meisje er géén schade van ondervindt.

De Singleton grens

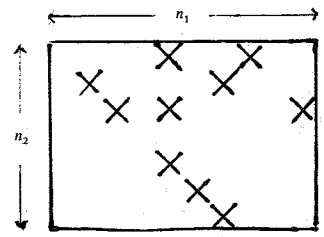
We geven nu een voorbeeld van een eenvoudig (maar leuk) stukje wiskunde uit de coderingstheorie, nl. de pessimistische kant: wat is zeker niet haalbaar. We kiezen weer een alfabet met q letters en beschouwen een code C met woorden van n letters en onderlinge afstand $\geq d$. Hoe veel woorden kan C dan hebben? Welnu, maak een lijst van deze woorden (het aantal noemen we $|C|$). Van alle woorden laten we de laatste $d - 1$ letters weg. Nog steeds zijn de (kortere) woorden verschillend! Onze conclusie is dat $|C| \leq q^{n-d+1}$. Stel nu dat de code C een k-dimensionale lineaire deelruimte is van $\mathbb{F}_q^n$. Dan is $|C| = q^k$. Daarmee is dan bewezen dat $k \leq n - d + 1$. Bij gegeven n en d is dit het beste dat men kan hopen te bereiken.

Reed – Solomon codes

Eén van de belangrijke ingrediënten van het foutenverbeterende systeem van de CD is een zgn. Reed – Solomon code. Het principe is met enige goede wil zelfs wel uit te leggen aan middelbare scholieren met gebruik van reële getallen of als men de leerlingen kan laten slikken dat er zoiets is als een eindig lichaam $\mathbb{F}_q$ ('rekenen met de symbolen is mogelijk'). We beschouwen nu informatie in de vorm van een zeer lange rij *letters* = elementen van $\mathbb{F}_q$. We hakken deze rij in blokken van steeds k letters. De RS code is een k-dimensionale code in $(\mathbb{F}_q)^n$, d.w.z. dat we moeten vertellen hoe een k-tal $(a_0, a_1, \ldots, a_{k-1}) \in (\mathbb{F}_q)^k$ wordt omgezet in een codewoord $(A_0, A_1, \ldots, A_{n-1}) \in (\mathbb{F}_q)^n$. We maken eerst de veelterm $a(x) = a_0 + a_1 x + \ldots\ a_{k-1} x^{k-1}$. Laat $\mathbb{F}_q = \{\alpha_0, \alpha_1, \ldots, \alpha_{q-1}\}$.
Definieer $A_i := a(\alpha_i)$, $(0 \leq i \leq q-1)$. Wat kunnen we van de afstand van deze code zeggen? Welnu, als het rijtje $\underline{a} = (a_0, a_1, \ldots, a_{k-1})$ aanleiding is tot het codewoord $\underline{A}$, en het rijtje $\underline{b} = (b_0, b_1, \ldots, b_{k-1})$ aanleiding tot het codewoord $\underline{B}$, dan is de afstand van $\underline{A}$ tot $\underline{B}$ gelijk aan het aantal indices i zó dat $A_i \neq B_i$. Dit is echter hetzelfde als het aantal elementen α_i van $\mathbb{F}_q$ zó dat $a(\alpha_i) \neq b(\alpha_i)$.
De veelterm $a(x) - b(x)$ met graad $\leq (k-1)$ kan niet meer dan k-1 nulpunten hebben. Daaruit volgt dat zeker $n - (k-1)$ keer geldt $A_i \neq B_i$. Deze code heeft dus minimum afstand $d \geq n - k + 1$. De Singleton grens vertelt ons dat $d \leq n - k + 1$ en dus is $d = n - k + 1$.

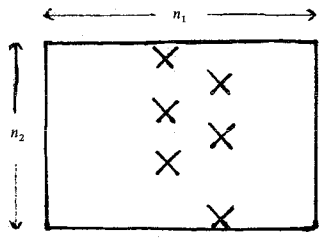
Product Codes

We geven nu nog in het kort een idee van één van de coderingsprincipes die een grote rol spelen bij de CD, nl. het gebruik maken van twee codes die op de een of andere manier *samenwerken*. Bij de CD zijn deze samenwerkende codes allebei (ingekorte) RS codes.
Het eenvoudigste voorbeeld zijn de zgn. product codes. Er is weer een alfabet $\mathbb{F}_q$ gekozen en we beschikken over twee lineaire codes C_1 en C_2 met parameters $[n_1, k_1]$ resp. $[n_2, k_2]$.
De aangeboden informatie splitsen we in blokken van $k_1 k_2$ letters die we gebruiken om een rechthoek met k_1 kolommen en k_2 rijen te vullen. Eerst wordt iedere rij via het coderingsalgoritme van C_1 omgezet in een rij van n_1 symbolen. De nieuwe rechthoek heeft afmetingen k_2 bij n_1. Daarna wordt iedere kolom via het coderingsalgoritme van C_2 omgezet in een codewoord van C_2 zodat tenslotte een n_2 bij n_1 rechthoek ontstaat. Deze rechthoek wordt nu *niet* rij voor rij uitgelezen en verzonden maar in een andere volgorde, die zo is gekozen dat een burst error in de oorspronkelijke rechthoek wordt verspreid over diverse rijen en kolomen. Via een zeer eenvoudig voorbeeld tonen we een groot voordeel van deze methode.
Stel dat C_1 en C_2 beide minimum afstand 3 hebben. De product code heeft dan afstand 9 (oefening voor de lezer). We verwachten dus 4 fouten in een ontvangen woord te kunnen verbeteren en mogen a priori niet op meer hopen. In figuur 4 geven de kruisjes aan waar de fouten in een ontvangen woord zich bevinden.



Figuur 4

Er zijn tien fouten! Dat ziet er niet zo best uit. We laten de decodeerprocedure van de code C_2 op de kolommen los. Er zijn vijf kolommen met slechts één fout en in elk van die kolommen wordt deze fout dus verbeterd. Laten we aannemen dat de fouten in kolom 5 toevallig een codewoord vormen. Die blijven dan staan. In kolom 7 is het nog erger. De procedure maakt er een fout bij met als resultaat figuur 5.



Figuur 5

De decodeerprocedure van de code C_1 gaat nu op de rijen werken en daar in ons voorbeeld iedere rij niet meer dan één fout heeft, worden in deze tweede slag alle fouten verbeterd.
In de praktijk is de situatie veel ingewikkelder maar het idee dat er achter zit is hierboven weergegeven. Tenslotte noemen we nog één van de trucs die bij muziek mogelijk zijn. De decodeerprocedure kan falen of wellicht zó veel 'fouten' verbeterd hebben dat, enige twijfel bestaat aan de juistheid van deze verbetering. In beide gevallen kan het resultaat worden voorzien van een waarschuwingsteken: 'onbetrouwbaar'. Als bij de reconstructie van het signaal (digitaalanaloog conversie; de omkering van figuur 2) een waarde optreedt die als onbetrouwbaar is bestempeld, dan wordt op die plaats geïnterpoleerd. Zo worden vele niet-verbeterbare foutenpatronen toch nog *gemaskeerd!*

Slot

Ik heb gepoogd om aan de hand van een bekend recent product een idee te geven van wat fouten – verbeterende codes zijn en hoe ze toegepast worden. Coding theory is een fascinerend vak waarin hulpmiddelen uit allerlei delen van de wiskunde een rol spelen. Desondanks is het mogelijk er onderwerpen in te vinden die zich lenen voor behandeling op het niveau van middelbaar onderwijs en soms (zoals ik heb aangetoond) zelfs lager onderwijs.

Literatuur

1 'Compact Disc Digital Audio', Philips Technisch Tijdschrift, Jaargang 40 (1981/2) no. 9. p. 265-296.
2 J. B. H. Peek, Communications Aspects of the Compact Disc Digital Audio System, IEEE Communication Magazine, Vol 23. No. 2, p. 7-15.
3 R. Lidl and H. Niederreiter, Finite Fields, Addison-Wesley, Reading (Mass.), 1983.
4 Discrete Wiskunde II, Collegedictaat T.U.E. 1987 (J. H. van Lint). (Hierin staat alles wat voor dit artikel nodig is over eindige lichamen op de eerste 5 bladzijden; voor belangstellenden is op aanvraag een kopie beschikbaar.)

Verschenen

S.L.O., *Leermiddelengids Wiskunde/Rekenen/Informatica,* ƒ12,75.
De Centrale Registratie Leermiddelen geeft jaarlijks voor het voortgezet onderwijs overzichten uit per vakgebied. De bovengenoemde gids geeft leraren wiskunde etc. een volledig overzicht van wat er aan methoden, leerboeken en educatieve software verkrijgbaar is.
Een titel- en auteursregister maken de gids op verscheidene manieren toegankelijk.

Berry c.s., *Mathematical Modelling Courses,* Ellis Horwood, 36.50, 281 blz. en Berry c.s., *Mathematical Modelling Methodology, Model and Micros,* Ellis Horwood, 38.50, 318 blz.
Deze twee boeken van dezelfde auteursteams beschrijven het opzetten van cursussen Wiskundige Modellen, waarbij het eerstgenoemde boek de meer theoretische aspecten behandelt terwijl het tweede meer de praktische kant benadert. Veel aandacht wordt besteed aan de werkelijkheidswaarde van de te ontwikkelen modellen.

Dirickx, Baas, Dorhout, *Operationele Research,* Adademic Service, ƒ60,–, 373 blz.
De belangrijkste methoden en technieken uit de OR, zoals (niet-) lineaire programmering, netwerkproblemen en geheeltallige programmering komen in dit boek aan de orde. Aan de hand van praktijkvoorbeelden wordt uitvoerig ingegaan op de modelmatige aspecten. Per hoofdstuk is een flink aantal opgaven opgenomen.

G. Tsu-der-Chou, *dBase III Handboek,* Academic Service, ƒ68,–, 364 blz.
Dit boek wil een leerboek zijn voor gebruikers van dit populaire databasepakket en is daarmee een uitbreiding op de door de dealer geleverde documentatie. D.m.v. vele voorbeelden en een uitgewerkt praktijkvoorbeeld wordt de lezer geleerd een op persoonlijke wensen toegesneden datasysteem te ontwerpen. Het boek is o.a. voorzien van een lijst met alle dBase III functies.

E. Verhulst, *MODULA-2,* Academic Service, ƒ68,–, 390 blz.
In dit boek wordt op de eerste plaats de programmeertaal MODULA-2 behandeld. Daarnaast wordt veel aandacht besteed aan het ontwerpen van softwaresystemen met behulp van modulen. Daarbij wordt gebruik gemaakt van een als standaard voorgestelde module bibliotheek.
Enige voorkennis van het ontwerpen van algoritmen en de taal PASCAL wordt verondersteld.