

Informatie theorie

Sebastiaan Polderman
nummer

Paul Sohier
0806122

20 januari 2011

Samenvatting

Inhoudsopgave

Inhoudsopgave	3
1 Informatie	4
1.1 Opdracht 1	4
1.2 Opdracht 4	5
1.3 Opdracht 5	5
2 Codesystemen voor storingsvrije omgevingen	6
2.1 Opdracht 1	6
2.2 Opdracht 4	7
2.3 Opdracht 6	7
3 Hoofdstuk 3	8
3.1 Opdracht 1	8
3.2 Opdracht 2	8
3.3 Opdracht 3	8
3.4 Opdracht 4	9
3.5 Opdracht 5	10

Hoofdstuk 1

Informatie

1.1 Opdracht 1

Een roulettespel heeft een draaischijf met 38 genummerde vakjes: 18 rode, 18 zwarte en 2 groene vakjes. Op de draaiende schijf wordt een balletje geworpen. Als de schijf tot rust komt, zal het balletje in één van de vakjes blijven liggen. Elk vakje heeft evenveel kans om het balletje te vangen. De zwarte vakjes zijn oneven genummerd van 1, 3, 5, ..., 35, de rode vakjes zijn even genummerd van 2, 4, 6, ..., 36 en de twee groene vakjes hebben de 'nummers' 0 en 00. Zodra de croupier de kleur of het nummer van het winnende vakje genoemd heeft, mag er niet meer ingezet worden.

- (a) Hoe groot is de selectieve informatie over elke kleur van het vakje?
- (b) Hoe groot is de gemiddelde informatie over de kleur van het vakje?
- (c) Hoe groot is de gemiddelde informatie over het nummer van het vakje?
- (d) Hoe groot is de conditionele entropie (equivocatie) van de kleur als het nummer van het vakje al bekend is?
- (e) Hoe groot is de conditionele entropie (Equivocatie) van het nummer als de kleur van het vakje al bekend is?
- (f) Hoe groot is de gemeenschappelijke entropie van de kleur en het nummer van het vakje?
- (g) Hoe groot is de onderlingen entropie van de kleur en het nummer van het vakje?
- (h) maak van $H(N, K)$, $H(K)$, $H(N)$, $H(N|K)$, $H(N; K)$, $H(K|N)$ een schema zoals figuur 1.3. Houd rekening met de getalswaarde van de entropieën.

1.2 Opdracht 4

Volgens bijlage D komen in normale teksten korte woorden frequenter voor dan lange woorden. In het 'Groene boekje' woordenlijst van de Nederlandse taal) blijkt dat niet te kloppen. Geef hier een verklaring voor.

1.3 Opdracht 5

Binare Coded Decimals ('BCD') is een code waarbij een getal van 0, 1, 2, ..., 99 in een 8 bit-woord (een '*byte*') gecodeerd wort. Hoeveel redundantie bevat zo'n BCD-woord?

Hoofdstuk 2

Codesystemen voor storingsvrije omgevingen

2.1 Opdracht 1

Een bron genereert een onafhankelijke rij symbolen uit het alfabet $\{0, 1\}$. De kans op een 0 is 0,9, en kans op een 1 is 0,1. De rij getallen wordt met een 'Run-Length coding' gecodeerd tot een rij met symbolen uit het alfabet $\{a, b, c, d, e, f, g, h, i\}$. Tenslotte wordt deze 'run length' codewoorden weer gecodeerd met een huffmancode.

b-code	r-code
1	a
01	b
001	c
0001	d
00001	e
000001	f
0000001	g
00000001	h
00000000	i

- (a) Bereken de entropie van de bron.
- (b) Bereken het gemiddeld aantal bronsymbolen per run-length codewoord.
- (c) Bereken het gemiddeld aantal bit van de Huffmancode per run-length woord.
- (d) Bereken het gemiddelde aantal bronsymbolen b per Huffmansymbool h .

2.2 Opdracht 4

Waarom zouden de makers van 'gzip' de blokverwijzingen beperkt hebben?

Opdracht 5 In het DNA van de bacterie *Micrococcus Lysodeiktus* hebben de basen A , C , T , G de volgende waarschijnlijkheid: $P(A) = P(T) = 29/200$ en $P(C) = P(G) = 71/200$. Bij de bacterie *E.Coli* is deze verdeling: $P(A) = P(T) = P(C) = P(G) = 1/4$. Welke bacterie zou van de twee het meest complexe organisme zijn?

2.3 Opdracht 6

Het decompressie-algoritme van het programma 'bzip2' maakt gebruik van een inversie permutatie $T(i)$ om uit de laatste kolom de bronrij terug te vinden. Wat is er fout aan de volgende redenering om de bronrij terug te vinden uit tabel 2.5, gegeven dat de rij-index van de bronrij de waarde 3 heeft?

Men kan uit de tabel 2.5 aflezen dat de onbekende bronrij moet beginnen met een 'o' (Rij 3, kolom 2) en eindigen met een 'b' (rij 3 kolom 1). Vervolgens blijkt de rij bronsymbolen uit alle andere tweetallen te zijn opgebouwd voor de tussenliggende symbolen: 'ob', 'ro', 'oo', 'rr', 'or'. Wij weten alleen niet in welke volgorde. Wel is bekend dat elk tweetal exact

Met deze gegevens kunnen wij de bronrij herstellen: Start met 'o' dan zijn er twee volgende letters mogelijk: 'b' en 'o' (vanwege 'bo' en 'oo'). De rij 'ob' heeft geen opvolger en is dus geen oplossing. De rij 'oo' heeft als uitbreiding 'oor' (Vanwege 'or'). Dit passen en meten kunnen wij herhalen tot wij de originele bronrij 'oorrob' gevonden hebben.

Hoofdstuk 3

Hoofdstuk 3

3.1 Opdracht 1

Een BSC heeft een *Binary Error Rate* van 0,3. Wat is de discrete kanaalcapaciteit van dit kanaal?

3.2 Opdracht 2

Noem 3 decodeerprincipes en hun eigenschappen.

3.3 Opdracht 3

De *Soundex codering* is een codering die bronwoorden uit een West Europese spreektaal vertaalt in codewoorden van één letter gevolgd door drie cijfers. Het voordeel van deze codering is dat de woorden die veel op elkaar lijken qua uitspraak, dezelfde code krijgen. Soundex wordt veel toegepast in spellingscontrole in woordprocessors, reisplanners en reserveringssystemen etc. Bijvoorbeeld, de namen 'brok', 'brock', 'broek' geven dezelfde code B620. Daarentegen geven de namen 'jansen', 'janssen', 'jansens' de J525.

Het soundex-algoritme werkt als volgt:

- (a) De eerste letter van het bronwoord wordt de beginletter in het Soundex codewoord;
- (b) De volgende 3 cijfers komen uit de volgende tabel. Zijn worden in volgorde opgebouwd in volgorde van de letters in het bronwoord. Als

whet bronwoord te kort is voor een volledig Soundex codewoord, wordt het Soundex codewoord aangevuld met nullen. Te lange codewoorden worden afgebroken na drie cijfers.

code	letter	uitspraakorgaan	Deze tabel geeft de cijfercodering van de letters aan. De meeste medeklinkers zijn volgens uitspraak gegroepeerd. De klinkers en de zachte medeklinkers krijgen geen cijfercode. Hoewel het Soundex-algoritme goed werkt voor West-Europese talen, is het niet voor andere spreektaalen geschikt.
1	b p f v	lippen	
2	c s k g j q x z	keel	
3	d t	tanden	
4	l	tong voor	
5	m n	neus	
6	r	tong achter geen	

a e h i o u y w

- (a) Welke aspecten maken Soundex codering geschikt voor Weste-Europese talen?
- (b) Een andere manier om alternatieve woorden te vinden is de '*Edit Distance*':
 - Spatie tussenvoegen;
 - Twee buurletters verwisselen;
 - Een letter vervangen door een andere;
 - Het verwijderen van een letter;
 - Een letter toevoegen.

Wat is het belangrijkste verschil tussen de '*Edit Distance*' en Soundex?

3.4 Opdracht 4

Gegeven een taal met 8 bronwoorden van 3 bit.

- (A)) Hoeveel redundantie moeten de codewoorden bevatten om een taal met 8 bronwoorden van ieder 3 bit intolerant voor 2 bit fouten te maken?
- (b) Welke CRC-polynoom zou in aanmerking komen om de bronwoorden te beschermen tegen 1 en 2 bit fouten?

3.5 Opdracht 5

Een geheugenloos kanaal tussen X en Y heeft de volgende eigenschappen
 $P(Y = y|X = x) = 0,5$

- (a) Bepaal de discrete kanaalcapaciteit van het kanaal in figuur 5.5?
- (b) Ontwerp een transmissiecode waarmee via dit kanaal foutloze transmissie plaats kan vinden.