

CLUSTEREN van VAKKENPAKKETTEN

Piet van Blokland

Samenvatting. De algoritmes voor een manier van oplossen van het clusterprobleem worden besproken. De belangrijkste algoritmes zijn "BRANCH and BOUND" algoritmes. Bovendien worden een aantal "versnellingen" besproken.

Het clusterprobleem is ontstaan door de mammoetwet. Hierbij introduceerde de wetgever de keuzevakken. Het was de taak van de schoolleiding om de vakken zo in het rooster te zetten, dat iedere leerling zijn eigen keuzevakken kon volgen. Door het verminderde aantal leerlingen neemt de cluster-problematiek nog toe. Dit artikel zal de algoritmes bespreken aan de hand van de keuzevakken van 80 leerlingen van 5 VWO van de R.S.G. Erasmus in Almelo.

De keuze van ee leerlingen was als volgt:

- | | |
|------------------|-------------------|
| 1. Aard W. | fa du gs et ee |
| 2. Marijke A. | gs we wt bi sk ee |
| . | . |
| . | . |
| 80. Peter van D. | la fa du gs we et |

Hierbij is ee: Economie 1, et: Economie 2 enzovoorts. De computer berekent vervolgens de aantallen leerlingen die een bepaald vak gekozen hebben. De gebruiker geeft dan het aantal keren dat dat vak wordt aangeboden op en de maximale groepsgrootte.

	la	fa	du	gs	ak	we	wt	na	sk	bi	ee	et
aantal leerlingen	6	43	58	57	17	60	13	28	17	31	58	37
aantal groepen	1	2	12	2	1	2	1	1	1	2	2	2
maximale groepsg. r.	6	24	30	29	17	30	13	28	27	18	30	22

De oplossing die aan deze voorwaarden voldoet en door de computer gevonden is, is:

	na	sk	wt	we	ee	bi	du	gs	et	fa	ak	la
1.	29	0	0	0	30	0	0	0	20	0	0	0
2.	0	27	0	30	0	0	0	0	17	0	0	0
3.	0	0	13	0	0	0	0	29	0	0	17	6
4.	0	0	0	31	0	0	30	0	0	20	0	0
5.	0	0	0	0	28	13	0	28	0	0	0	0
6.	0	0	0	0	0	18	28	0	0	23	0	0

Een cluster is een verzameling vakken, die tegelijk wordt aangeboden. Het eerste cluster bestaat uit 3 vakken (na, ee, et). In totaal zijn er 6 clusters.

De Werkwijze.

Allereerst worden de pakketten gemaakt. D.w.z. alle leerlingen die dezelfde vakken hebben gekozen, worden bij elkaar gezet. Twee vakken die in een pakket zitten vormen een combinatie. In het voorbeeld zijn er 45 pakketten.

Vervolgens moeten de vakken geordend worden en wel volgens de volgende regels:

- a) Eerst die vakken voorop die slechts een keer worden aangeboden en met ieder ander vak combineren.
- b) Dan de vakken die een keer worden aangeboden, en combineren met de vakken die bij a) en b) gevonden zijn.
- c) Dan de overblijvende vakken van achteraan aanvullen zodat er zoveel mogelijk pakketten gegeven kunnen worden met de overblijvende vakken.
- d) De vakken die dan nog overblijven zo neer zetten, dat het aantal combinaties met de beginvakken zo groot mogelijk is.

Indien we bovenstaande regels toepassen, krijgen we de volgende volgorde:

na	sk	wt	we	ee	bi	du	gs	et	fa	ak	la
1	2	3	5	5	6	7	8	9	10	11	12

De volgorde van de vakken blijkt van groot belang te zijn voor de snelheid waarmee een oplossing gevonden wordt.

Daarna worden de pakketten geordend.

pakket 1:	5 4 3 2 1 (is: ee we wt sk na)	(5) pakket
pakket 2:	6 4 3 2 1	
pakket 3:	6 5 4 2 1	(6) pakket
pakket 4:	6 5 4 3 2 1	
pakket 5:	7.....	
		(7) pakket
pakket 45:	12 10 9 8 7 5	

De pakketten zijn nu lexicografisch geordend. In het vervolg worden de nummers van de vakken weer in de gewone volgorde geschreven. Dus pakket 1 is: 1 2 3 4 5.

Het probleem wordt nu in tweeën gesplitst: Eerst wordt geprobeerd een zodanige samenstelling van de clusters te vinden, zodat iedere leerling zijn eigen keuzevakken kan volgen. Indien dat eenmaal het geval is, wordt ook geprobeerd de pakketten zo over de clusters te verdelen, dat ook de groepen niet te groot worden.

Het vinden van de clusters

Het probleem is nu om de volgende verdeling van 0 en 1 te vinden.

na	sk	wt	we	ee	bi	du	gs	et	fa	ak	la
1	0	0	0	1	0	0	0	1	0	0	0
0	1	0	1	0	0	0	0	1	0	0	0
0	0	1	0	0	0	0	1	0	0	1	1
0	0	0	1	0	0	1	0	0	1	1	0
0	0	0	0	1	1	0	1	0	0	0	0
0	0	0	0	0	1	1	0	0	1	0	0

Bij ieder vak is een patroon van 0 en 1. Bijvoorbeeld bij gs is dat 0 0 1 0 1 0. Dat wil zeggen dat gs in het derde en vijfde cluster wordt aangeboden.

Om deze matrix van 0 en 1 te bereiken gaan wij de patronen als volgt ordenen:

1	2	3...	6	7	8.....	21	22	23		63
1	0	0	0	1	1		0	1	1	1
0	1	0	0	1	0		0	1	1	1
0	0	1	0	0	1		0	1	0	1
0	0	0	0	0	0		0	0	1	1
0	0	0	0	0	0		1	0	0	1
0	0	0	1	0	0		1	0	0	1

De 63 patronen zijn zo geordend dat de patronen met een 1 bij elkaar staan, de patronen met twee 1 enzovoorts. Indien voor een vak twee groepen beschikbaar zijn, dan moet voor dat vak gekozen worden uit een van de patronen tussen 7 en 21. Via backtracking worden de patronen doorlopen. Deze procedure zal in de beginfase worden getoond. Stel dat de eerste 4 vakken al vast liggen en we het patroon voor ee willen bepalen:

na	sk	wt	we	ee	bi
1	0	0	1		
0	1	0	0		
0	0	1	0		
0	0	0	1		
0	0	0	0		
0	0	0	0		

Probeer alle (5) pakketten te plaatsen (1 2 3 4 5), dan is patroon 7 niet mogelijk. Immers

1	0	0	1	1
0	1	0	0	1
0	0	1	0	0
0	0	0	1	0
0	0	0	0	0

De procedure is om patroon 7 te verhogen naar 8, naar 9,, naar 13 (1 0 0 0 1 0). Dan pas is een oplossing mogelijk.

Probeer nu alle (6) patronen te plaatsen (1 2 3 4 6, 1 2 4 5 6 en 1 2 3 4 5 6):

na	sk	wt	we	ee	bi
1	0	0	1	1	
0	1	0	0	0	
0	0	1	0	0	
0	0	0	1	0	
0	0	0	0	1	
0	0	0	0	0	

Ook hier is patroon 7 (1 1 0 0 0 0) niet mogelijk. Het eerst passende patroon blijkt patroon 17 (1 0 0 0 0 1) te zijn. Indien er geen passend patroon was gevonden, had het patroon van ee met een verhoogd moeten worden. Deze procedure gaat verder totdat alle patronen passen.

Tweede fase.

Fase I heeft patronen van 0 en 1 opgeleverd zodat alle leerlingen hun vakkenpakket kunnen volgen. Nu gaat het erom of de leerlingen "redelijk" over de klassen verdeeld kunnen worden. Daartoe worden eerst alle leerlingen aan die vakken toegewezen, die zij wel moeten volgen.

Voor pakket 1(1 2 3 4 5) geldt immers dat dit pakket slechts op een manier geplaatst kan worden. Bij pakket 13 (1 4 6 7 8) moet na in het eerste cluster volgen en et in het tweede cluster. Het minimale aantal leerlingen, dat een bepaald vak moet volgen, gegeven deze clusters, staat in onderstaande tabel.

28	0	0	0	29	0	0	0	17	0	0	0
0	27	0	30	0	0	0	0	17	0	0	0
0	0	13	0	0	0	0	20	0	0	17	6
0	0	0	30	0	0	2	0	0	5	0	1
0	0	0	0	27	11	0	18	0	0	0	0
0	0	0	0	0	14	15	0	0	0	0	0

Stelling. Gegeven is de functie $E: \{1, 2, \dots, n\} \rightarrow \{0, 1\}$ en er is een i zodanig dat $E(i)=1$. Zij de functie $R: \{1, 2, \dots, n\} \rightarrow \{0, 1\}$ zodanig dat geldt $R(i) \leq E(i)$ en er is precies een i zodanig dat $R(i)=1$, dan geldt:

$$(1) \quad \sum_{i=1}^n R(i) \leq \sum_{i=1}^n E(i) - \sum_{i=1}^{n-1} E(i) * E(i+1)$$

Het bewijs van deze stelling gaat via volledige inductie naar n .

Aangezien de ongelijkheid (1) gesommeerd mag worden over alle pakketten en bovendien de ongelijkheid blijft gelden, als wij een permutatie toepassen op $1, 2, \dots, n$, kan het resultaat toegepast worden in de programma's.

Voorbeeld 3. Ook hier maken wij gebruik van het maximale aantal leerlingen, dat een vak kan volgen.

A	B
0	0
24	21
0	31
34	0
0	0

Vak A is door 51 leerlingen gekozen en de maximale klassegroote is 28. Vak B is door 47 leerlingen gekozen en de maximale klassegroote is 28. De getallen die hiernaast staan zijn weer de maximale aantallen leerlingen die een vak kunnen volgen. 4 leerlingen kunnen

tijdens het derde cluster zowel in vak A als in vak B geplaatst worden. Echter bij deze opstelling kunnen slechts

$$(24 + 21 - 4) + 28 + 28 = 97$$

leerlingen vak A en/of vak B volgen. De vakken A en B moeten echter $51 + 47 = 98$ keer gevolgd worden. Het bovenstaande is een gevolg van de volgende stelling.

Stelling. Zij E, F, R, S functies van $\{1, 2, \dots, n\} \rightarrow \{0, 1\}$ en er is een i zodanig dat $E(i)=1$ en er is een i zodanig dat $F(i)=1$. Bovendien geldt $R(i) \leq E(i)$ en $S(i) \leq F(i)$ en er is precies een i met $R(i)=1$ en er is precies een i met $S(i)=1$ en voor iedere i geldt $R(i)+S(i) \leq 1$. Dan geldt

$$\sum_{i=1}^n R(i) + S(i) \leq \sum_{i=1}^n (E(i) + F(i) - E(i) * F(i))$$

Het bewijs berust weer op een eenvoudige toepassing van volledige inductie.

Ook in het eerste deel is sprake van een aantal versnellingen.

Voorbeeld 4. Hoewel in het volgende voorbeeld iedereen zijn pakket kan volgen, is het direct duidelijk dat nooit een oplossing is met deze clusters, die ook rekening houdt met de klassegrootte. Dit valt gemakkelijk te controleren doordat het totaal aantal vakken dat gevolgd wordt bekend is, en omdat per cluster niet meer als het totaal aantal leerlingen les kan hebben:

1	0	0	1	0	1	1	0	0	1	0	0
0	1	0	0	0	0	1	0	1	0	1	0
0	0	1	0	1	0	0	1	0	1	0	1
0	0	0	1	0	0	0	1	0	0	0	0
0	0	0	0	1	0	0	0	1	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0

Nog niet genoemde pluspunten van het computer-programma:

- 1) Het aantal clusters in instelbaar. Het hoeft niet gelijk te zijn aan het aantal vakken dat leerlingen kiezen.
- 2) Een maximale klassegrootte kan tussentijds verlaagd worden.
- 3) Het niet gelijktijdig voorkomen van bepaalde vakken (bijvoorbeeld biologie en scheikunde want die worden door dezelfde docent gegeven) kan worden ingesteld
- 4) Het is gratis verkrijgbaar. Alleen de kosten van het schijfje en de porto zijn voor de gebruiker.
- 5) Het is eenvoudig om een vak aan de clustering toe te voegen. Door het toevoegen van een vak dat iedereen moet lopen neemt het aantal mogelijkheden om een oplossing te vinden toe. Eventueel kan ook het vak vr (=vrij) toegevoegd worden.

Nog onopgeloste problemen en bezwaren:

- 1) Soms is het nodig om een andere volgorde van de vakken te kiezen, dan het programma automatisch doet. Dit is voor een niet-deskundige lastig.
- 2) Soms doet het programma er wel erg lang over om een oplossing te vinden.
- 3) Als er geen oplossing is, geeft de computer geen suggesties voor wijzigingen.
- 4) Er wordt geen rekening gehouden met het aantal uren dat een vak wordt aangeboden: Bijvoorbeeld Latijn is een 6-uurs vak en Frans een 3-uurs vak.
- 5) Het programma is niet gegarandeerd foutvrij.