
GELOOFSNETWERKEN

Kinderziekten met vlekjes

Nicholas Piël

project verslag • Universiteit van Amsterdam • 10 April, 2006



INTRODUCTIE

In deze opdracht moesten we een simpel geloofsnetwerk maken met de bedoeling om snel een diagnose te kunnen stellen bij verschillende symptomen die huid ziekten kunnen veroorzaken.

THEORIE

Regel van Bayes

Geloofsnetwerken hangen sterk samen met de theorie van Bayes die het volgende zegt. Wanneer we de product regel mogen aannemen:

$$P(A \wedge B) = P(A | B) P(B)$$

$$P(A \wedge B) = P(B | A) P(A)$$

Kunnen we dit herschrijven naar:

$$P(A | B) = P(B | A) P(A) / P(B)$$

Dit staat ook wel bekend als de regel van Bayes. Hoewel dit niet echt handig lijkt omdat je nu 3 waardes nodig hebt om $P(A | B)$ uit te kunnen rekenen lijkt het in de praktijk praktisch te zijn omdat deze drie kansen vaak al bekend zijn.

Bayesian Updating

Stel we weten:

$$P(\text{Cavity} | \text{Toothache}) = 0.8$$

$$P(\text{Cavity} | \text{Catch}) = 0.95$$

Wat is nu de kans op $P(\text{Cavity} | \text{Toothache} \wedge \text{Catch})$? We zouden dit kunnen herformuleren volgens de regel van Bayes naar:

$$P(\text{Toothache} \wedge \text{Catch} | \text{Cavity}) P(\text{Cavity}) / P(\text{Toothache} \wedge \text{Catch})$$

Dit heeft als probleem dat wanneer we bovenstaande formule willen oplossen we n^2 aantal kansen moeten hebben / genereren.

Maar, wanneer we weten dat er een conditionele onafhankelijkheid is tussen X en Y gegeven Z kunnen we gebruik maken van het volgende:

$$P(X | Y, Z) = P(X | Z)$$

Bayesian Networks

Een Baysiaans geloofsnetwerk maakt gebruik van de conditionele onafhankelijkheid tussen knopen en geeft een complete non-redundante weergave van het domein dat veel compacter is dan een ‘full joint’.

Waarbij een volledig uitgewerkt tabel 2^n nummers vereist voor n -nodes. Is dit bij een bayesiaans netwerk maar $n2^k$ voor n -nodes met maximaal k -parents.

Zo zou je bij het voorbeeld wat ik hieronder uitwerk (13 nodes, maximaal 4 parents) bij een volledige joint 8192 nummers moeten geven. Bij het bayesiaanse netwerk heb je maar 208 nodig.

Fuzzy Or

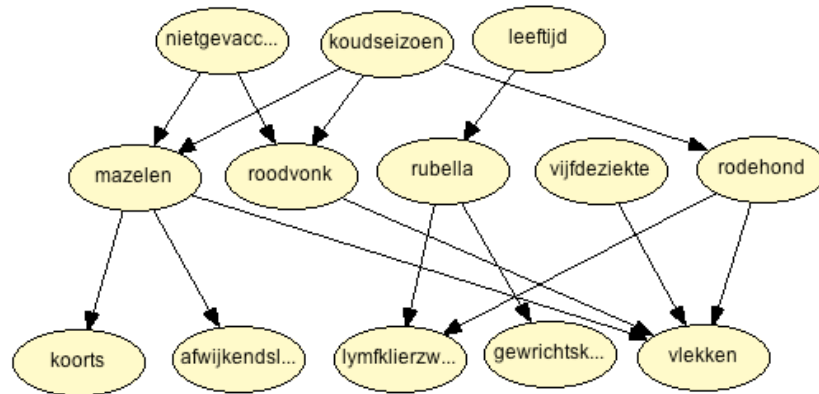
Maar zelfs dan heb je nog redelijk veel getallen nodig die je niet altijd weet. In dergelijke getallen kun je gebruik maken van de ‘noisy-or’ wanneer je de volgende twee assumpties in acht neemt:

- Elke oorzaak heeft een onafhankelijk effect
- Alles wat een bepaalde oorzaak veroorzaakt is ook weer onafhankelijke van elkaar.

Met als gevolg dat je maar een paar waardes hoeft te weten om een tabel volledig te maken.

IMPLEMENTATIE

Modelleren van het netwerk



Het netwerk

Het netwerk bestaat zeer duidelijk uit drie lagen, de mogelijke oorzaken die de kans op een ziekte verhogen, de ziekten zelf en dan de onderste laag de mogelijke symptomen.

Invullen van de tabellen

We maken gebruik van de noisy-or, onderstaande tabel is gegenereerd door Huggin met de volgende expressie:

NoisyOR(rodehond, 0.5, rubella, 0.4, true, 0.8)

Dit betekent het volgende, de kans dat alleen rodehond opgezwollen lymfklieren veroorzaakt is 0.5 (1-0.5) de kans dat Rubella dit doet is 0.6 (1-0.4). Er is ook een kans op 'lekker' gespecificeerd van 0.2. Dit omdat er waarschijnlijk ook andere oorzaken zijn voor opgezwollen lymfeklieren.

	rodehond	false		true	
	rubella	false	true	false	true
normale lymfklieren		0.8	0.32	0.4	0.16
opgezwollen lymfklieren		0.2	0.68	0.6	0.84

voorbeeld tabel 'lymfklieren'