

Tentamen IN2205 Kennissystemen

03 Juli 2009, 14:00–17:00

- Dit tentamen heeft 5 meerkeuzevragen in totaal goed voor 10 punten en 7 open vragen met in totaal 47 punten.
- Wat betreft de meerkeuzevragen:
 - Er is voor iedere vraag telkens maar één goed antwoord mogelijk.
- Wat betreft de open vragen:
 - Geef antwoord in correct Nederlands of Engels en schrijf leesbaar (gebruik eerst kladpapier).
 - Motiveer je antwoorden.
 - Geef geen irrelevante informatie. Dit kan leiden tot puntenaftrek.

Vraag:	6	7	8	9	10	11	12	Totaal:
Punten:	8	9	5	5	8	8	4	47

- Het gebruik van boek of aantekeningen tijdens dit tentamen is niet toegestaan.
- Het gebruik van een rekenmachine is toegestaan.
- Voordat je je antwoorden inlevert, controleer of op ieder blaadje je naam en studienummer staat en geef het aantal ingeleverde bladen aan op (tenminste) de eerste pagina.
- De tentamenstof bestaat uit hoofdstukken 1,2,3,4,6,7,8 en 9 uit *Artificial Intelligence* van Michael Negnevitsky, behalve secties 3.4, 3.5, 8.2, 9.1 en 9.7, het artikel *Swarm Smarts* van Eric Bonabeau *et. al.*, de lecture notes over Bayesian Reasoning en Swarm Intelligence en de handouts over reinforcement learning.
- Uiteraard komen in één tentamen niet alle onderwerpen aan bod. Trek daarom op basis van dit tentamen geen conclusies over stof die nooit getoetst wordt.
- Totaal aantal pagina's: 5.

Succes!

Multiple-choicevragen

1. (2 punten) Om te achterhalen of een expertsysteem een bepaalde oplossing kan realiseren met de beschikbare gegevens kunnen we gebruik maken van . . .
- A. forward chaining maar niet van backward chaining.
 - B. backward chaining maar niet van forward chaining.
 - C. forward chaining of backward chaining.**
 - D. certainty factors.

2. (2 punten) Stel we hebben de volgende regels in een expertsysteem:

1. IF a THEN b
2. IF b THEN d
3. IF b AND c THEN e
4. IF f THEN b AND c

Welke regel zal vuren als de database feiten b , c en f bevat en we gebruik maken van forward chaining waarbij de conflict resolutie methode kiest voor de meest specifieke regel?

- A. regel 1.
 - B. regel 2.
 - C. regel 3.**
 - D. regel 4.
3. (2 punten) Stel, we hebben een fuzzy verzameling van goede films met membershipfunctie $\mu_A(x)$. Hoe definiëren we de membershipfunctie van de verzameling van redelijk goede films?
- A. $\mu_A^{redelijk}(x) = \mu_A(x)^2$
 - B. $\mu_A^{redelijk}(x) = 2\mu_A(x)$
 - C. $\mu_A^{redelijk}(x) = \frac{1}{2}\mu_A(x)$
 - D. $\mu_A^{redelijk}(x) = \sqrt{\mu_A(x)}$

4. (2 punten) We spreken van *emergent* gedrag van een systeem als dat gedrag . . .

- A. niet terug te voeren is op het gedrag van de componenten van het systeem.**
- B. zonder supervisie geleerd is door het systeem.
- C. met behulp van reinforcement geleerd is door het systeem.
- D. voldoende is om de Turingtest te doorstaan.

5. (2 punten) Hoe selecteert, in een genetisch algoritme, *roulette wheel selection* chromosomen?

- A. De kans op selectie is gelijk voor elk chromosoom.
- B. De kans op selectie is proportioneel aan de fitness van het chromosoom.**
- C. Eerst worden de beste chromosomen gekozen; later een klein aantal van de minder goede.
- D. Goede chromosomen worden met slechte gepaard.

Open vragen

6. Geef voor ieder van de onderstaande vraagstukken aan welke van de in dit vak behandelde oplossingsmethoden je zou gebruiken, motiveer je antwoorden.
- (a) (2 punten) Sorteren van e-mails / documenten op basis van inhoud.

Antwoord: Dit is een vorm van clusteren. We kunnen ieder document representeren met een vector waarin bijvoorbeeld voor ieder woord wordt aangegeven hoe vaak het in het document voorkomt. Deze vectoren kunnen we clusteren met ant-based clustering of met een self-organising feature map.

- (b) (2 punten) Herkennen van een gezichtsexpressie, zoals blij, boos, neutraal.

Antwoord: Aangezien er een enorme variatie is in gezichten en gezichtsuitdrukkingen maar het wel eenvoudig is om hier voorbeelden van te verzamelen aan hand waarvan een systeem kan leren is een neuraal netwerk een goede keuze.

- (c) (2 punten) Vinden van een formule aan hand van een groot aantal metingen.

Antwoord: Genetic programming kan gegeven de benodigde basisfuncties een formule construeren. De fitnessfunctie komt neer op het berekenen van de uitkomsten met behulp van de geconstrueerde formules en deze vergelijken met de metingen.

- (d) (2 punten) Voorspellen van files aan hand van gegevens als de situatie op de weg, het weer en de dag van het jaar.

Antwoord: Het lijkt hier mogelijk enige regels op te stellen, zoals „aan het begin van de vakantie is het druk” of „als het regent is er meer verkeer op de weg”. Gegeven de invoer en de continue waarden die schuil gaan achter begrippen als ‘het weer’ en ‘druk’ is een fuzzy expertsysteem of een Bayesiaans netwerk waarschijnlijk een goede keuze. Je zou echter ook kunnen overwegen om over een langere periode data te verzamelen en dan een neuraal netwerk te trainen. Simulatie d.m.v. een ant-based systeem heb ik ook goed gerekend.

7. (a) (2 punten) Stel we hebben de fuzzy sets A en B met membershipfuncties:

$$\mu_A(x) = (0.8/1, 0.2/2, 0.1/3, 0.2/4, 0.0/5) \quad (1)$$

$$\mu_B(x) = (0.1/1, 0.2/2, 0.4/3, 0.6/4, 0.8/5) \quad (2)$$

Bereken de membershipfunctie van de fuzzy set $\neg(A \cap \neg B)$ met behulp van de standaardoperatoren.

Antwoord:

$$\mu_{\neg B} = (0.9/1, 0.8/2, 0.6/3, 0.4/4, 0.2/5) \quad (3)$$

$$\mu_{(A \cap \neg B)} = (0.8/1, 0.2/2, 0.1/3, 0.2/4, 0.0/5) \quad (4)$$

$$\mu_{\neg(A \cap \neg B)} = (0.2/1, 0.8/2, 0.9/3, 0.8/4, 1.0/5) \quad (5)$$

- (b) (5 punten) Geef met behulp van plaatjes aan hoe fuzzy inferentie met behulp van de Mamdani methode werkt. Benoem de verschillende stappen in het proces.

Antwoord: zie boek blz. 108.

- (c) (2 punten) Leg uit wat het verschil is tussen Mamdani en Sugeno fuzzy inferentie.

Antwoord: Bij Mamdani inferentie is de consequent van een regel een fuzzy set. Bij Sugeno inferentie is dit een wiskundige functie $f(\cdot)$ van de inputs. In het simpelste geval (zero-order Sugeno fuzzy model) is dit een constante.

8. (5 punten) Multilayer feedforward neural networks zijn de meest gebruikte neurale netwerken. De gewichten van een dergelijk netwerk kunnen geleerd worden met behulp van de *generalized delta rule*:

$$w_{jk}(p+1) = w_{jk}(p) + \beta \Delta w_{jk}(p) + \alpha y_j(p) \delta_k(p). \quad (6)$$

Leg uit hoe leren in een multilayer neural network werkt, geef daarbij aan waar de termen in bovenstaande leerregel voor staan.

Antwoord: Leren in een multilayer neuraal netwerk gebeurt met het backpropagation algoritme. Een voor een worden voorbeelden aan het netwerk aangeboden. Voor iedere input wordt eerst de output van het netwerk berekend. Deze output wordt vergeleken met de gewenste output. Dit levert een fout $e_k(p)$ op voor ieder output neuron k . Aan hand van deze fout wordt de error gradient $\delta_k(p)$ bepaald (door de afgeleide van de activatie functie van neuron k te vermenigvuldigen met de fout.)

De gewichten w_{jk} van neuron k worden vervolgens bijgesteld volgens bovenstaande regel. Het nieuwe gewicht hangt dus af van de huidige waarde van het gewicht, van de error gradient en van de output $y_k(p)$ van neuron k in iteratie p . De verandering in het gewicht wordt gewogen met de learning rate α , een getal tussen 0 en 1.

Aan hand van de outputwaarden in de output layer en de error gradienten in de output layer worden de error gradienten van de neuronen in de laatste verborgen laag berekend. De gewichten worden wederom bijgesteld met behulp van bovenstaande formule. Dit proces herhaalt zich tot de inputlaag bereikt is. De fout propageert dus van achter naar voren door het netwerk.

De momentum term neemt ook de vorige verandering van het gewicht $\Delta w_{jk}(p)$ mee gewogen met een constante $\beta \in [0, 1)$. Het momentum stabiliseert het leerproces. Als de veranderingen constant in dezelfde richting zijn (een eenvoudige zoekruimte) wordt het leerproces versneld, als de veranderingen in het gewicht steeds tegengesteld zijn (en de zoekruimte kennelijk veel pieken en dalen bevat) worden de leerstappen kleiner zodat het algoritme nauwkeuriger zoekt.

Dit leerproces wordt herhaald totdat de totale fout klein genoeg is.

9. (a) (2 punten) Beschrijf de structuur van een self-organising feature map (Kohonen netwerk).

Antwoord: Een self-organising feature map is een neuraal netwerk bestaande uit twee lagen. Alle neuronen in de eerste laag, de input laag, zijn verbonden met alle neuronen in de tweede laag die een 2 dimensionaal vlak geplaatst zijn (in het algemeen in N dimensies). De outputlaag vormt een afbeelding van de verdeling van de (hoogdimensionale) inputdata. De neuronen in de output laag zijn onderling verbonden („lateral connections”) om een competitie tussen die neuronen mogelijk te maken.

- (b) (3 punten) Leg uit hoe leren in een self-organising feature map werkt.

Antwoord: Een self-organising feature map gebruikt competitive learning, een vorm van unsupervised learning. Voor elke input wordt voor ieder van de neuronen in de outputlaag wordt de inputvector vergeleken met de vector die wordt gevormd door de gewichten van verbindingen naar het outputneuron. Vergelijken gaat aan hand van een afstandsmaat zoals de Euclidische afstand. Het

neuron met de kleinste afstand is de winnaar en mag zijn gewichten aanpassen volgens $\alpha[x_i - w_{ij}(p)]$ waarbij α de learning rate is, x_i een input en w_{ij} het gewicht dat hoort bij de verbinding tussen neuronen i en j . Eventueel kan een neighborhoodfunctie worden gebruikt die bepaald of en hoe de burens van het winnend neuron hun gewichten mogen aanpassen. Dit proces wordt herhaald totdat een gewenste minimumafstand is bereikt of totdat geen significante veranderingen in het netwerk meer optreden.

10. Een helpdesk heeft vier medewerkers. Zij willen allemaal ergens in de komende vier weken op vakantie, waarbij ze hun vakanties zo willen plannen dat de bezettingsgraad van de helpdesk in deze vier weken zo hoog mogelijk blijft. In deze opgave gaan we een *genetisch algoritme* gebruiken om dit vraagstuk op te lossen. Ga daarbij uit van de volgende gegevens: medewerkers Anna en Bert willen ieder 2 weken op vakantie, medewerker Cor slechts 1 week en medewerkster Daisy 4 weken.

- (a) (3 punten) Hoe kunnen we de situatie coderen als chromosoom?

Antwoord: We kunnen reeksen van 4 bits als genen gebruiken. Het chromosoom bevat voor iedere medewerker dan 1 gen. De posities in het gen staan voor de weken. Een bit is 1 als de betreffende medewerker die week op vakantie is. Omdat het aantal weken vakantie dat de medewerkers willen opnemen verschilt is niet iedere reeks van 4 bits een geldig gen, we moeten voor iedere medewerker aangeven wat geldige genen zijn (zie boek blz. 237).

- (b) (2 punten) Wat is een geschikte fitnessfunctie?

Antwoord: Uit een chromosoom kunnen we voor ieder van de vier weken aflezen hoeveel medewerkers aanwezig zijn. We zouden het kleinste van deze 4 getallen als fitness kunnen nemen.

- (c) (2 punten) Hoe ziet de cross-over operator er in dit geval uit?

Antwoord: We kunnen de standaard cross-over operator gebruiken (zolang we maar op genniveau werken, dwz. dat een reeks van 4 bits die samen een gen vormen niet gesplitst mag worden).

- (d) (1 punt) Hoe ziet de mutatie operator er in dit geval uit?

Antwoord: Bij mutaties kunnen we geen bitflip gebruiken omdat hierdoor het aantal weken dat een medewerker op vakantie gaat veranderd kan worden. In plaats daarvan moeten we een voor de betreffende positie toegestaan gen selecteren uit de genpool.

11. Particle Swarm Optimization (PSO) kan gebruikt worden voor het trainen van neurale netwerken. Leg uit hoe dit in zijn werk gaat. Ga uit van een multilayer feedforward netwerk en de aanwezigheid van een verzameling voorbeelden waarvan bekend is welke output het netwerk zou moeten geven. Er wordt gebruik gemaakt van *local best* PSO. Ter herinnering, de regel om de snelheid van een particle aan te passen luidt:

$$v_{ij}(t+1) = v_{ij}(t) + c_1 r_{1j} [y_{ij}(t) - x_{ij}(t)] + c_2 r_{2j} [\hat{y}_{ij}(t) - x_{ij}(t)] \quad (7)$$

Behandel in je antwoord:

- (a) (2 punten) de parameters c_1 , r_{1j} , c_2 , r_{2j} in bovenstaande formule,

Antwoord: c_1 en c_2 zijn constanten die de bijdragen van cognitieve en sociale componenten wegen. Over het algemeen zijn ze ongeveer gelijk. r_{1j} en r_{2j} zijn random getallen tussen 0 en 1. Deze getallen helpen het algoritme om de zoekruimte te exploreren.

(b) (1 punt) representatie als particle,

Antwoord: Ieder particle is een vector bestaande uit de gewichten in het netwerk.

(c) (1 punt) de populatie,

Antwoord: De populatie bestaat uit N random gegenereerde particles. (Bij voorkeur goed verspreid door de hele zoekruimte.)

(d) (2 punten) fitness,

Antwoord: De fitness van een particle kunnen we bepalen door het netwerk dat correspondeert met het particle resultaten te laten berekenen voor de beschikbare data en dan te berekenen hoezeer deze afwijken van de gewenste resultaten, bijvoorbeeld door de mean-squared error te bepalen.

(e) (2 punten) aanpassen van de positie van een particle.

Antwoord: De nieuwe positie van een particle is de oude positie van het particle plus een snelheidscomponent die met bovenstaande formule berekend wordt. De snelheid van een particle hangt af van de vorige snelheid, de afstand en positie van het particle t.o.v. van zijn beste positie (cognitieve component) en zijn afstand en positie t.o.v. de beste positie van de neighborhood (sociale component). Deze beste posities worden bepaald met behulp van de fitnessfunctie.

12. (4 punten) Een biometrisch slot maakt gebruik van een aantal sensoren om vast te stellen of de persoon die voor de deur staat een bekende is die naar binnen mag. Helaas is geen van deze sensoren feilloos. De eerste sensor, een stemherkenner, herkent in 5% procent van de gevallen de stem van een onbekende als de stem van een bekende en in 10% van de gevallen herkent het de stem van een bekende niet. De tweede sensor, een irisscanner, herkent in 3% van de gevallen een onbekende ten onrechte als een bekende en in 2% van de gevallen herkent het een bekende niet. Gemiddelde is een op de tien mensen die voor de deur gaat staan een onbekende. Gegeven dat zowel de stemherkenner als de irisscanner aangeven dat het om een bekende gaat, wat is de kans dat we toch te maken hebben met een onbekende?

Antwoord: Als B aangeeft of we te maken hebben met een onbekende, S of de stemherkenner een bekende herkent en I of de irisscanner een bekende herkent, dan wordt gevraagd naar:

$$P(B = false | S = true, I = true). \quad (8)$$

Met behulp van de regel van Bayes (en de veronderstelling dat de sensoren onafhankelijk van elkaar opereren) kunnen we dit uitdrukken in termen van kansen die gegeven zijn:

$$P(B = false | S = true, I = true) = \alpha P(B = false) P(S = true | B = false) P(I = true | B = false). \quad (9)$$

Invullen geeft:

$$P(B = false | S = true, I = true) = \alpha \frac{1}{10} \times 0.05 \times 0.03 = 0.00019, \quad (10)$$

met

$$\alpha = 1 / \left(\frac{1}{10} \times 0.05 \times 0.03 + \frac{9}{10} \times 0.90 \times 0.98 \right) = 1.2595. \quad (11)$$

Einde tentamenopgaven.

Controleer voor de zekerheid of je alle vragen hebt beantwoord. Het zouden er 12 moeten zijn.