



Universiteit
Leiden
The Netherlands

Archaeology and the application of artificial intelligence : case-studies on use-wear analysis of prehistoric flint tools

Dries, M.H. van den

Citation

Dries, M. H. van den. (1998, January 21). *Archaeology and the application of artificial intelligence : case-studies on use-wear analysis of prehistoric flint tools*. Retrieved from <https://hdl.handle.net/1887/13148>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/13148>

Note: To cite this publication please use the final published version (if applicable).

samenvatting (Dutch summary)

Archeologie en kunstmatige intelligentie: toepassing voor gebruikssporenanalyse op prehistorische vuurstenen werktuigen

Met behulp van kunstmatige intelligentie wordt gepoogd computers menselijke eigenschappen te laten nabootsen. Men leert ze natuurlijke talen te begrijpen, schaak te spelen, diagnoses te stellen, beursprijzen te voorspellen, beelden te herkennen, optimale routes te bedenken, etc. De mogelijkheden zijn talrijk en we komen het gebruik van kunstmatige intelligentie dan ook niet alleen tegen in verschillende takken van wetenschap, maar ook bij overheids- en zakelijke instanties waar we in het dagelijks leven regelmatig mee te maken hebben.

Sinds de jaren '80 worden ook in het archeologisch vakgebied technieken toegepast die gebruik maken van kunstmatige intelligentie, met name expertsystemen. Hun rol is tot op heden echter beperkt gebleven, waardoor ze nauwelijks geïntegreerd zijn in het archeologisch onderwijs en onderzoek (hoofdstuk 2). Naar aanleiding van de resultaten uit die beginperiode twijfelden archeologen aan het nut van deze technieken voor hun discipline. Uit studies naar de toepassingsmogelijkheden bleken de faciliteiten ontoereikend om archeologische kennis adequaat weer te geven. Verder bestond de angst voor fossilisering van kennis, dat wil zeggen dat kennis die eenmaal in een kennissysteem is opgenomen niet meer wordt vernieuwd. Kortom, archeologie leek geen geschikt kennisgebied voor dergelijke toepassingen en ze werden dan ook niet op grote schaal ingevoerd.

Het is echter de vraag of deze negatieve conclusie niet wat voorbarig is geweest. Omdat aan de ontwikkeling van de meeste programma's vooral een onderzoeksdoel ten grondslag lag, zijn de meeste toepassingen niet veel verder gekomen dan het ontwerp- of prototypestadium. Hierdoor zijn de bevindingen dan ook niet onderbouwd met evaluaties van hun praktische toepasbaarheid of met objectieve testresultaten. Het nut noch de ongeschiktheid van kunstmatige intelligentie voor de archeologie is dus nauwelijks aangetoond.

De doelstelling van de huidige studie is om na te gaan of de archeologie inderdaad een vakgebied is dat weinig profijt heeft van de toepassing van op kennis gebaseerde systemen. Hierbij was het uitgangspunt niet het bestuderen van de mogelijkheden van een techniek, maar juist het toetsen van de praktische toepasbaarheid van een systeem dat ontwikkeld is met behulp van kunstmatige intelligentietechnieken. Het voor deze studie geselecteerde archeologische specialisme

is de gebruikssporenanalyse van prehistorische vuurstenen werktuigen (hoofdstuk 4). Dit is een methode die op basis van slijtagekenmerken op werktuigen de voormalige functie ervan probeert te reconstrueren. Dergelijke reconstructies geven inzicht in de (economische) activiteiten van onze prehistorische voorouders.

De keuze voor dit specialisme is min of meer toevallig. Zij is ingegeven door de omstandigheden, niet door een afweging van verwachte resultaten. Van belang voor de keuze was namelijk dat er door de beperkte menskracht aan de Leidse Faculteit der Pre- en Protohistorie om de methode van gebruikssporenanalyse te onderwijzen, behoefte was aan educatieve ondersteuning. Een bijkomend aspect was dat de methode zelf gebaat was bij een meer objectieve aanpak. De implementatie van de kennis in een expertsysteem zou daarbij kunnen helpen. Een laatste argument was dat de in Leiden werkzame gebruikssporenspecialist bereid was haar kennis beschikbaar te stellen.

Een belangrijke wens van de expert was dat het systeem in de eerste plaats ingezet zou kunnen worden bij de educatie van studenten. Daarom lag het voor de hand om een expertsysteem te ontwikkelen. Deze methode was bij de start van de studie (begin van de jaren '90) de verst ontwikkelde en meest toegepaste methode. Kort daarop kwamen de neurale netwerken in het middelpunt van de belangstelling te staan. In de archeologie werd deze techniek geïntroduceerd als een superieur alternatief voor expertsystemen. Daarop is besloten om deze techniek ook in het onderzoek te betrekken, zodat een functionele vergelijking kon worden gemaakt.

Het kennisgebied: gebruikssporenanalyse

Eén van de belangrijkste problemen van de gebruikssporenanalyse die aandacht behoefde, is het subjectieve karakter. Interpretaties worden afgeleid van persoonlijke observaties, maar noch de observaties, noch de beschrijvingen van de geobserveerde slijtagekenmerken of de redeneringen zijn voldoende gebaseerd op gestandaardiseerde en objectieve procedures (hoofdstuk 4). Omdat dit probleem geruime tijd de wetenschappelijke erkenning van de methode bemoeilijkt, zijn in de loop der tijd al meerdere pogingen ondernomen om de methode een meer objectieve en kwantitatieve basis

te geven. Deze pogingen richtten zich op elk van de drie subjectieve aspecten:

- op het vervangen van de menselijke visuele observaties door optische of mechanische metingen;
- op het met behulp van experimenten opbouwen van een kwantitatief referentiekader voor de relatie tussen uitgevoerde activiteiten en het ontstaan van gebruikssporen;
- op het formaliseren en standaardiseren van het interpretatieproces.

De in de onderhavige studie gepresenteerde systemen zijn voorbeelden van deze derde benadering.

Principes van expertsystemen

Een expert- of kennissysteem is een computerprogramma dat het uitvoeren van een specialistische taak nabootst waarvoor de beschikbare kennis niet zuiver algoritmisch van aard is en waarop geen lineaire wiskundige functies van toepassing zijn. Ze zijn daarentegen gebaseerd op vuistregels en ervaringsfeiten (heuristieken). Bovendien is deze kennis op een zodanige manier opgeslagen dat niet-experts op het betreffende terrein er gebruik van kunnen maken (hoofdstuk 3). Een ander kenmerk is dat een expertstelsel uit drie gescheiden delen bestaat. Hierdoor zijn de feitelijke gegevens, de procedurele handelingen en de communicatie met de gebruiker van het systeem gescheiden. Dit maakt dat elk onderdeel afzonderlijk kan worden aangevuld of veranderd zonder dat een heel programma hoeft te worden aangepast. Hierdoor zijn ze flexibeler en onderhoudsvriendelijker dan conventionele computerprogramma's.

Omdat er verschillende soorten kennis bestaan zijn er meerdere manieren om kennis in een expertstelsel te representeren. Een veelgebruikte methode is de *als-dan*-beslissingsregel. Het ontwikkelproces is te vereenvoudigen met behulp van een zogenaamde *shell*. Dit is een leeg expertstelsel dat gedeeltelijk is voorgeprogrammeerd en dat snel tot een operationeel systeem kan worden uitgebouwd wanneer het wordt voorzien van kennis. Hierdoor is het ontwikkelen van een expertstelsel niet langer een activiteit die uitsluitend is voorbehouden aan informatica-deskundigen.

Principes van neurale netwerken

Eén van de alternatieven voor een expertstelsel is een neuraal netwerk (hoofdstuk 6). Dit is een programma dat is afgeleid van de werking van menselijke hersencellen (neuronen). Doordat deze cellen met miljarden aan elkaar zijn gekoppeld en ze signalen aan elkaar kunnen doorgeven, wordt razendsnel informatie uitgewisseld. Zo kunnen wij adequaat op omgevingsimpulsen reageren. Een neuraal netwerk simuleert dit proces. Het bestaat uit een netwerk van getallen die verdeeld zijn over een laag met invoergetallen, één of meerdere tussenliggende lagen waarin berekeningen

worden uitgevoerd, en een laag met uitvoergetallen. De verbindingen tussen de getallen in de diverse lagen hebben gewichten die de sterkte van de relatie weergeven. Alle gewichten bij elkaar bepalen de hoogte van de uitvoergetallen. In tegenstelling tot een expertstelsel wordt een neuraal netwerk niet gevoed met heuristische kennis die is weergegeven als beslissingsregels, maar met een grote hoeveelheid numerieke gegevens uit een database. Deze gegevens representeren voorbeelden van situaties (invoergegevens) waarbij een bepaalde uitkomst hoort (uitvoergegevens). Tijdens een trainingsprocedure selecteert het netwerk uit de voorgeschotelde voorbeelden zelf de gegevens die voor de betreffende taak van belang zijn. De invoergegevens worden gekoppeld aan de juiste uitkomsten en de verbindingen krijgen een gewicht. Het getrainde netwerk kan vervolgens worden ingezet om de uitkomsten van situaties te voorspellen die vergelijkbaar zijn met de geleerde voorbeelden. Doordat een neuraal netwerk geen gemodelleerde kennis nodig heeft, kan het worden ingezet voor taken die gebruikmaken van (numerieke) kennis waaruit op het eerste gezicht geen lineaire verbanden kunnen worden gehaald.

Het expertstelsel voor gebruikssporenanalyse

De belangrijkste uitgangspunten bij het ontwerp van het expertstelsel dat WAVES is genoemd (Wear Analyzing and Visualising Expert System), was dat het analyseproces van gebruikssporenonderzoek moest worden gestandaardiseerd en dat het op een gebruiksvriendelijke manier moest worden ontsloten voor studenten (hoofdstuk 5). Doordat studenten zijn in te delen in twee groepen, beginners en gevorderden, die elk gebaat zijn bij een aan hun kennis en ervaring aangepaste ondersteuning, bestaat WAVES uit twee componenten. Met de ene component kunnen beginners analyses uitvoeren, met de andere kunnen gevorderden hypothesen toetsen. De beide benaderingen zijn afzonderlijk van elkaar te gebruiken.

Het analyse-traject is bedoeld om beginnende gebruikssporenanalisten te begeleiden bij het systematisch en stapsgewijs verzamelen van de benodigde gegevens en bij het beschrijven van de waargenomen kenmerken. Verder interpreteert het deze gegevens en geeft het uitleg over de totstandkoming ervan. De tweede benadering, het traject om hypothesen te toetsen, biedt meer gevorderde analisten de mogelijkheid hun eigen interpretaties te toetsen door een *second opinion* aan WAVES te vragen. De analisten geven aan aan welk contact-materiaal ze de geobserveerde slijtagekenmerken toewijzen en WAVES vraagt hen vervolgens of de kenmerken waarvan WAVES vindt dat ze aanwezig moeten zijn om de hypothese te aanvaarden, daadwerkelijk op het werktuig zijn aangetroffen. Op die manier worden aanwijzingen verzameld die de interpretatie van de analist kunnen bevestigen of verwerpen.

Vooral in de analyse-module hebben de educatieve aspecten veel aandacht gekregen. Zo krijgt de analist onder meer informatie over de technische kanten van de methode (het prepareren van werktuigen voor de analyse, het gebruik van verschillende vergrotingen etc.) en over de diagnostische betekenis van verschillende kenmerken, en krijgt hij illustraties van sporen die horen bij een bepaald contact-materiaal. Daarnaast wordt uitgelegd hoe de interpretaties tot stand zijn gekomen, welk deel gebaseerd is op kwantitatieve en welke op subjectieve gegevens, en op grond van welke kenmerken bepaalde contact-materialen of bewegingen worden uitgesloten als de mogelijke veroorzaker van de slijtage. De hypothese-toetsingsmodule bevat minder educatieve elementen dan de analyse-module. Van de gebruikers wordt verondersteld dat ze gevorderde analisten zijn die geen prijs meer stellen op basale instructies. Wel is in deze module ook veel beeldmateriaal gebruikt. Hier helpt het de analist om zowel het contact-materiaal aan te wijzen waarvan hij of zij vermoedt dat het de slijtagesporen heeft veroorzaakt, als om de geobserveerde slijtagekenmerken te beschrijven.

De kennisacquisitie voor WAVES

Om de interpretaties van slijtagekenmerken voor de analyse-methode te kunnen formaliseren en een objectieve basis te kunnen geven is naar een manier gezocht om de beschikbare kennis te kwantificeren. Gebruikssporenexperts kunnen prima aangeven welke elementen van belang zijn voor een analyse en welke slijtage zij kenmerkend vinden voor welke contact-materialen, maar het is onmogelijk om die kennis te vertalen in interpretaties met betrouwbaarheidsfactoren. Om dit te ondervangen is voor de basis van het systeem gebruikgemaakt van de resultaten van experimenten met nagemaakte vuurstenen werktuigen. Deze gegevens zijn afkomstig uit een referentiecollectie die is opgebouwd door de gebruikssporenspecialist van de Leidse Faculteit der Pre-en Protohistorie. De collectie bestaat uit 300 experimenten die zijn uitgevoerd op 20 verschillende contact-materialen, en met 5 soorten bewegingen.

Kenmerkend voor gebruikssporen is dat ze geen absoluut diagnostische waarde hebben en dat er sprake is van elkaar overlappende patronen. Dit wil zeggen dat experimenten met verschillende contact-materialen of bewegingen vergelijkbare slijtagekenmerken kunnen veroorzaken, maar ook dat vergelijkbare experimenten met één bepaalde materiaalcategorie verschillende slijtagetypen te zien kunnen geven. Enerzijds wordt dit veroorzaakt doordat de kenmerken op een subjectieve manier worden beschreven, anderzijds doordat de variatie aan slijtagekenmerken minder groot is dan het aantal contact-materialen.

Ook bij de gebruikte referentiecollectie was sprake van overlappende patronen. Om toch een betrouwbare indicatie te kunnen krijgen van de relatie tussen de slijtagekenmerken

en de contact-materialen is per materiaalcategorie nagegaan hoe vaak een bepaald kenmerk voorkwam. Hieruit bleek dat bepaalde slijtagekenmerken meer diagnostisch zijn voor het ene materiaal dan voor het andere. Een glans met een zeer heldere en gladde textuur kwam bijvoorbeeld veel vaker voor bij plantaardige materialen (bij 86% van de experimenten met graan) dan bij dierlijke materialen (bij 46% van de experimenten met bot). Die aantallen zijn omgerekend tot percentages, die op hun beurt zijn vertaald in diagnostische waardes. Deze waardes zijn in WAVES als 'scores' in de *als-dan*-beslissingsregels verwerkt.

In de praktijk betekent dit dat het systeem op basis van elk slijtagekenmerk dat een analist/student heeft waargenomen en beschrijft een score toekent aan verschillende contact-materialen (of aan bewegingen). Als bijvoorbeeld een heldere en gladde textuur zou worden waargenomen, dan krijgt — naar analogie van het bovengenoemde voorbeeld — het contact-materiaal graan een hogere score dan bot. Op grond van een volgende kenmerk kunnen de scores precies omgekeerd zijn. De contact-materialen die tijdens de experimenten dit slijtagekenmerk niet hebben veroorzaakt krijgen ook tijdens het analyseproces geen score en worden uitgesloten van het verdere analyse-traject. Aan het eind van het analyse-traject worden alle scores voor elk contact-materiaal en voor elke beweging opgeteld. De uitkomst hiervan geeft aan dat de slijtagekenmerken op het bestudeerde werktuig veroorzaakt kunnen zijn door alle activiteiten waarvoor een score is aangegeven, maar dat ze het meest typerend zijn voor de activiteit met de hoogste score.

Voor de hypothese-toetsing is dezelfde kennis gebruikt als voor het analyse-traject, maar de redenatielijn is tegenovergesteld, dat wil zeggen 'doelgericht' (waarbij gericht data wordt verzameld om achterwaarts, d.w.z. naar een hypothese te kunnen redeneren) in plaats van 'datagericht' (waarbij op grond van de beschikbare gegevens wordt geprobeerd een hypothese op te stellen). In deze module wordt alleen gelet op de aanwezigheid of afwezigheid van bepaalde kenmerken en zijn de scores achterwege gelaten.

Het belangrijkste voordeel van het gebruik van de gegevens uit de referentiecollectie is dat het een beter inzicht geeft in de variatie aan slijtagekenmerken die de contact-materialen kunnen veroorzaken dan de literatuur. Bovendien geeft het een kwantitatieve (objectieve) basis aan de interpretaties. Een nadeel van de kennis die is afgeleid van de referentiecollectie, is dat het alleen betrekking heeft op gebruikssporen die onder ideale omstandigheden zijn verkregen. Prehistorische werktuigen vertonen vaak een complexer patroon van sporen dan de werktuigen die voor experimenten zijn gebruikt, en ze kunnen zijn aangetast door hun langdurig verblijf in de grond. Een referentiecollectie is dus niet geheel representatief voor de sporen die een analist in de praktijk kan tegenkomen. Dit probleem is voor een deel ondervangen door de

gegevens uit de referentiecollectie aan te vullen met expertise van de expert en van een aantal ervaren analisten.

Het neurale netwerk voor gebruikssporenanalyse

Naar de mening van sommige archeologen zijn neurale netwerken superieur aan expertsystemen wat betreft hun functionaliteit en sociale acceptatie. Daarom is in deze studie niet alleen gekeken naar de mogelijkheden die een expert-systeem biedt voor de toepassing die in deze studie centraal staat, maar is ook nagegaan wat een neuraal netwerk in dit geval kan betekenen (hoofdstuk 6). Hiervoor is WARP (Wear Analysing and Recognizing neural network Prototype) ontwikkeld.

Om een eerlijke vergelijking te kunnen maken, is WARP gevoed met dezelfde gegevens als WAVES. Het netwerk is uitsluitend getraind om de glans op werktuigen te interpreteren, niet de overige slijtagekenmerken zoals krassen, afronding en afsplinteringen. Dit heeft twee redenen. Ten eerste bevatte de referentiecollectie meer gegevens over dit type slijtagekenmerk dan over de andere, ten tweede bevond ook WAVES zich in een zodanig ontwikkelingsstadium dat het uitsluitend kennis bevatte over glanspatronen. Door de kennis van WARP tot dit aspect te beperken, kon een eerlijke vergelijking van de functionaliteit worden gemaakt. De structuur van WARP is uiterst eenvoudig. Het bestaat uit een invoerlaag, een verborgen laag en een uitvoerlaag. De invoerlaag bestaat uit 31 neuronen die de verschillende slijtagekenmerken representeren, de uitvoerlaag bestaat uit 15 neuronen die de contact-materialen vertegenwoordigen. De gebruiker selecteert de 5 neuronen van de invoerlaag die overeenkomen met de kenmerken die hij of zij op een werktuig observeert. Op grond van deze 5 kenmerken geeft WARP aan welke van de 15 contact-materialen waarschijnlijk het slijtagepatroon heeft veroorzaakt. Eenvoudige neurale netwerken zoals WARP kunnen echter niet aangeven hoe deze interpretatie tot stand is gekomen.

De trainingsset voor WARP bestond uit 160 voorbeelden van glanspatronen en bijbehorende contact-materialen. Na diverse trainingsrondes en allerhande aanpassingen van de netwerkstructuur bleek het prototype niet in staat om alle voorbeelden te leren. De onbewerkte gegevens bevatten te veel ruis en met elkaar conflicterende voorbeelden om het trainingsproces optimaal te kunnen laten verlopen. Aangezien het gemiddeld aantal fouten niet kon worden teruggebracht tot nul, is het trainingstraject als voltooid beschouwd toen de foutenmarge van een testset op 9% lag en dit niet verder verbeterde. Aangezien WARP was bedoeld als een experimenteel model dat alleen voor een vergelijking van twee benaderingen zou worden gebruikt en niet als operationeel systeem, zijn geen pogingen ondernomen om de functionaliteit te verhogen door bijvoorbeeld het aantal trainingsvoorbeelden uit te breiden of de bestaande voorbeelden te verbeteren.

Testresultaten

Een tussentijdse versie van WAVES en het prototype van WARP zijn aan een test onderworpen om de verrichtingen en prestaties van beide systemen te vergelijken. De test bestond uit het analyseren van 16 experimenteel gebruikte vuurstenen werktuigen en 10 archeologische exemplaren van de neolithische (Lineair Bandkeramische) vindplaats Elsloo. De gebruikssporen op deze laatste groep werktuigen zijn voorafgaand aan de test geïnterpreteerd door een ervaren gebruikssporenanalist. Hierdoor konden de antwoorden van de beide programma's worden vergeleken met traditionele interpretaties.

Bij de analyse van de experimentele stukken gaven WAVES en WARP respectievelijk in 56,3% en in 43,8% van de gevallen een juiste interpretatie. Bij de archeologische stukken waren die percentages 50% en 80%. WARP scoorde bij de tweede groep dus opmerkelijk beter. WAVES kon geen antwoord geven in de gevallen die afweken van de ingevoerde kennis. WARP kon dat wel, omdat het antwoord van een neuraal netwerk altijd een voorspelling is die gebaseerd is op de overeenkomsten met de geleerde voorbeelden. Zelfs als er weinig overeenkomsten zijn, wordt een voorspelling gedaan. Deze aanpak kan zowel gunstig als ongunstig uitwerken. In het geval van WARP werden bij de archeologische stukken weinig fouten gemaakt (2 van de 10), maar van de 16 experimentele stukken werden er 9 (56,3%) fout geïnterpreteerd. In totaal werden 11 (42,3%) foute antwoorden gegeven. WAVES gaf alleen bij de experimentele stukken één foute interpretatie, maar kon in totaal in 42,3% (11 van de 26) van de gevallen geen antwoord geven.

Toch waren de resultaten bemoedigend. De ontbrekende of foute interpretaties werden namelijk veroorzaakt door de geringe hoeveelheid kennis die de beide systemen op dat moment tot hun beschikking hadden. Duidelijk was wel dat de verschillende benaderingen onder verschillende omstandigheden een ander resultaat te zien gaven: WARP kon beter interpoleren (nieuwe situaties interpreteren die sterk lijken op de voorbeeld-situaties), WAVES beter extrapoleren (nieuwe situaties interpreteren die enigszins afwijken van de voorbeeld-situaties).

WAVES is op basis van de bovenstaande resultaten verder ontwikkeld. De uiteindelijke versie, die ook kennis bevat van de overige slijtagekenmerken, is opnieuw aan een test onderworpen. Hierbij stond niet zozeer de evaluatie van de kennis centraal, als wel het toetsen van de prestaties wanneer het wordt gebruikt door analisten en studenten. De test vond plaats in de vorm van een *blind test*. Dit wil zeggen dat de samensteller van de test weet welke contact-materialen de sporen hebben veroorzaakt, doordat ze door middel van experimenten zijn verkregen, maar de deelnemers niet. Zo kan met zekerheid worden nagegaan of de deelnemers met behulp van WAVES kunnen achterhalen welke contact-

materialen de sporen hebben veroorzaakt. In dit geval werden dus niet de analyse-resultaten van de deelnemers getest, maar die van WAVES. Ter controle werden de interpretaties die het systeem ontleende aan de beschrijvingen van de analisten vergeleken met de antwoorden die het had ontleend aan de beschrijvingen van de gebruikssporenexpert die werkzaam is aan de Faculteit der Pre- en Protohistorie. Vier analisten van verschillende leerscholen en van verschillende kennisniveaus analyseerden elk 15 werktuigen. Eén van de deelnemers had nog nooit eerder gebruikssporen bestudeerd. De werktuigen waren voor uiteenlopende experimenten gebruikt en vertoonden een grote variatie aan gebruikssporen. Op basis van de verschillende kenmerken gaf WAVES telkens interpretaties voor drie variabelen: de hardheid van het bewerkte materiaal, het contact-materiaal en de uitgevoerde beweging.

Alle beschrijvingen van de deelnemers en de daaruit door WAVES afgeleide interpretaties werden vergeleken met de controle-set. Eén van de verrassingen was dat de beginnening opmerkelijk goed presteerde. Het gemiddelde percentage correcte interpretaties van alle deelnemers was 55,8%, dat van de blanco student 56,8%. Daarnaast vielen de grote verschillen tussen de behaalde resultaten van de analisten op: het hoogst behaalde individuele percentage correcte interpretaties was 80%, het laagste 30%. Het best behaalde resultaat van de blanco student was dat 9 van de 15 (60%) interpretaties van de hardheids categorie van het bewerkte materiaal correct waren. Het laagste percentage was 53,3% (de interpretatie van de beweging op basis van de macroscopische slijtagekenmerken).

De oorzaak van de verschillen bleek te liggen in de grote variatie in de beschrijvingen van de analisten. Ondanks het feit dat het systeem de analisten richtlijnen geeft voor het beschrijven van hun observaties, weken hun beschrijvingen in veel gevallen af van elkaar en van die van de expert (de controle-set). Het viel daarbij op dat in sommige gevallen de beschrijvingen van de beginnening objectiever waren dan die van de gevorderde analisten. De gevorderde analisten werden gehinderd door hun persoonlijke interpretaties. Ze pasten hun beschrijving daaraan aan, zodat ze — onbewust — het antwoord van WAVES probeerden te beïnvloeden. De beginnening maakte systematisch een aantal beginnersfouten, maar volgde nauwlettender de aanwijzingen van het systeem op. Het feit dat het systeem niet de juiste interpretaties gaf, betekende in dit geval dus niet dat de kennis in het systeem niet toereikend was, maar dat de invoer van de gegevens voor verbetering vatbaar was.

Over het algemeen waren de bevindingen van de deelnemers aan de test positief. Men erkende unaniem het gemak van een dergelijk systeem, maar de meningen waren verdeeld over de wenselijkheid van het op grote schaal inzetten van expertsystemen voor educatieve doeleinden. Ten slotte werden

de resultaten vergeleken met die van alle eerder gepubliceerde *blindtests*, die tot doel hadden de functionaliteit te evalueren van de gebruikssporenanalyse als methode. Hieruit bleek dat de met WAVES behaalde resultaten op zijn minst vergelijkbaar zijn met de prestaties van menselijke deskundigen. De hoogste positieve score die andere groepen analisten gemiddeld hadden behaald met het interpreteren van het bewerkte contact-materiaal was 49,0%, de laagste positieve score was 33,7%. In de test met WAVES was de gemiddelde positieve score 46,7%. Deze was dus bijna net zo hoog als de beste gemiddelde score van de menselijke experts. De positieve scores ten aanzien van de interpretatie van de uitgevoerde beweging gaven het zelfde beeld te zien: het hoogste gemiddelde was 82,6%, het laagste 69,4% en het met WAVES behaalde gemiddelde was 80%.

Conclusies

Op grond van de tests die met WAVES en WARP zijn uitgevoerd (hoofdstuk 7) kan worden geconcludeerd dat beide systemen vergelijkbaar met menselijke deskundigen presteren. Met een verbetering van de gebruikte technieken en een uitbreiding van de kennis kunnen de prestaties wellicht nog verder verbeterd worden. Een voordeel van het automatiseren van dit aspect van de gebruikssporenanalyse is, dat het meer objectieve interpretaties oplevert en dus beter controleerbare en vergelijkbare resultaten. Toch heeft de methode daarmee nog geen optimale objectieve basis gekregen. Het waarnemen en beschrijven van de slijtagekenmerken door de analisten blijft subjectief. Zo lang ook dit niet wordt verbeterd, blijven de resultaten van een analyse sterk afhangen van de inbreng van de analist. Het kwantificeren van de observaties verdient dan ook meer aandacht dan het op dit moment krijgt. Gezien de ontwikkelingen in beeldverwerkingstechnieken zijn ook op dit gebied toepassingen voor gebruikssporenanalyse denkbaar. Uiteindelijk zal de gebruikssporenmethode alleen een volledig objectief karakter kunnen krijgen wanneer een combinatie van verschillende technieken wordt toegepast.

Met betrekking tot de toepasbaarheid van WAVES kan worden geconcludeerd dat het vooral een hulpmiddel is om beginnende analisten de beginselen van de gebruikssporenanalyse te leren. WAVES zet de dingen voor hen op een rij, geeft de complexiteit van de methode weer, laat hen nadenken over wat ze doen en biedt de gelegenheid om te oefenen. Op een aantal universiteiten wordt het inmiddels voor dat doel gebruikt door archeologiestudenten. Toch is het een instrument dat onder begeleiding van een docent moet worden gebruikt. Het is niet bedoeld om foutieve observaties en beschrijvingen te corrigeren en heeft van een aantal materiaal-categorieën nog slechts een beperkte hoeveelheid kennis. WARP bleek eveneens adequate interpretaties te kunnen geven, zeker gezien de beperkte hoeveelheid kennis die voor

het trainingsproces beschikbaar was. De verwachting is dan ook dat met een flinke uitbreiding van het aantal trainingsvoorbeelden en met een aantal andere aanpassingen de resultaten nog verder kunnen verbeteren. Wel is het minder geschikt voor de educatieve doeleinden die het primaire doel van deze studie waren. Het programma is voor de gebruiker weinig transparant en kan bijvoorbeeld niet aangeven waarop een interpretatie is gebaseerd. Ook hier zijn wel verbeteringen te bewerkstelligen, maar die zijn niet uitgetoetst omdat het buiten het doel van deze *case study* viel om naast WAVES ook nog een volledig operationeel neurale netwerk te bouwen. WARP is alleen bedoeld geweest als een prototype.

Uit de vergelijking van de twee benaderingen kan worden geconcludeerd dat ze beide voor- en nadelen hebben. Een keuze voor één van beide moet afhangen van het doel dat men met een toepassing voor ogen heeft en van de beschikbare kennis en gegevens. Wanneer men bijvoorbeeld een beter inzicht wil krijgen in de relatie tussen een fenomeen en de omgevingsvariabelen, of de statistische waarschijnlijkheid van een interpretatie wil kennen, dan biedt een neurale netwerk meer mogelijkheden dan een expertsysteem. Omgekeerd is voor het stap voor stap nabootsen van een procedure een expertsysteem geschikter dan een neurale netwerk.

Aanbevelingen

De bevindingen uit deze studie zijn een duidelijke reden om eerdere conclusies, dat technieken die gebaseerd zijn op kunstmatige intelligentie een beperkte toepassing hebben voor de archeologie, te nuanceren. Ze kunnen wel degelijk van nut zijn voor de archeologie en een aanzienlijke toegevoegde waarde leveren (hoofdstuk 8). Het grootste pluspunt van dergelijke methoden is dat ze de mogelijkheid bieden om kostbare kennis en unieke expertise vast te leggen en te exploiteren als de menselijke expert niet (meer) beschikbaar is. Deze democratisering van kennis is zowel van belang voor het wetenschappelijk onderzoek als voor het onderwijs. Complexe en subjectieve kennis kan worden gestructureerd, geformaliseerd, geëvalueerd en vervolgens worden uitgebouwd. Zo kunnen archeologische methoden objectiever en consistenter worden gebruikt en worden de resultaten beter vergelijkbaar.

Het inzetten van een expertsysteem of neurale netwerk voor een bepaalde taak kan ook de betrokken experts en studenten voordelen bieden. Wanneer eenvoudige en routinematige klussen door een geautomatiseerd systeem worden overgenomen, kan de expert zijn capaciteit beter inzetten voor moeilijke en vaak meer interessante problemen, zodat nieuwe kennis kan worden opgedaan. Bovendien kan de expert door het inzetten van een geautomatiseerde assistent zijn onderwijscapaciteit vergroten. Het voordeel voor studenten is dat ze de leerstof in hun eigen tempo kunnen bestuderen. Bovendien kan dit op een moment waarop het hen uitkomt. Een op kennis gebaseerd systeem is immers dag en nacht beschikbaar, is onvermoeibaar en zal keer op keer dezelfde taak geduldig en consequent uitvoeren.

In de archeologie is een keur aan potentiële toepassingen denkbaar en realiseerbaar die al dan niet voor educatieve doeleinden kunnen worden ingezet. Concrete voorbeelden zijn specialistische analyses van vondstcategorieën, relatieve dateringstechnieken, classificaties etc. Met name nu de commercialisering in de archeologie zijn intrede doet, kunnen dergelijke kennisystemen en neurale netwerken bijdragen aan een optimalisering van de exploitatie van expertise — bijvoorbeeld bij het analyseren van een groeiende hoeveelheid opgravingsgegevens — en dus aan een grotere efficiëntie. Om echter de rol van op kennis gebaseerde technieken in de archeologie daadwerkelijk te kunnen vergroten, verdient het aanbeveling om, op grond van de kennis die we hebben van de mogelijkheden van dergelijke systemen, onze houding ten aanzien van kunstmatige intelligentie te wijzigen. Zo zullen we — zowel systeemontwikkelaars als gebruikers — een andere betekenis en invulling aan de term *functionaliteit* moeten geven: we moeten af van het idee dat op kennis gebaseerde technieken op zijn minst op het specialistische kennisniveau van een menselijke expert moeten functioneren voordat we ze het predikaat ‘bruikbaar hulpmiddel’ kunnen geven. Juist als we ze inzetten voor taken die de expert als eenvoudig en routinematig ervaart, wordt niet alleen de inzetbaarheid van de basale specialistische kennis vergroot, maar kan ook de ware expertise van de deskundige beter worden benut. Zo wordt een op kennis gebaseerd systeem geen concurrent, maar een waardevolle assistent.