

Methoden voor klasse-identificatie empirisch vergeleken

Citation for published version (APA):

Cosemans, G., Dijkstra, J., Hajek, J., Post, W., Schaefer, W. F., Vávra, F., & Melichar, J. (1992). *Methoden voor klasse-identificatie empirisch vergeleken*. (Computing centre note; Vol. 56). Technische Universiteit Eindhoven.

Document status and date:

Gepubliceerd: 01/01/1992

Document Version:

Uitgevers PDF, ook bekend als Version of Record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Technische Universiteit Eindhoven
Computing Centre Note 56

**Methoden voor klasse-identificatie
empirisch vergeleken**

Auteurs: G. Cosemans, J. Dijkstra, J. Hajek, W. Post,
W. Schaefer, F. Vávra en J. Melichar.

Mei 1992.

Bibliotheek 
Technische Universiteit
Eindhoven

Methoden voor klasse-identificatie empirisch vergeleken.

Auteurs:

Guido Cosemans, Afdeling Informatiesystemen, Studiecentrum voor Kernenergie, te Mol, België 09 32-14311801

Jan Dijkstra Calibre, Vakgroep Architectuur en Urbanistiek, Faculteit Bouwkunde, Techn. Univ. Eindhoven, Postbus 513, 5600 MB, Eindhoven 040-473399

Jan Hajek, Rekencentrum, Techn. Univ. Eindhoven, Postbus 513, 5600 MB, Eindhoven 040-474577

Wilfried Post, Vakgroep Sociaal-wetenschappelijk Informatica, Faculteit Psychologie, Univ. van Amsterdam 020-5252073

Wim Schaefer, Vakgroep Productie en Uitvoering, Faculteit Bouwkunde, Techn. Univ. Eindhoven, Postvak 8, Postbus 513, 5600 MB Eindhoven 040-472373

Frantisek Vávra & Jiri Melichar

Department of Cybernetics, Section of Artificial Intelligence and Information Theory, West Bohemian University, CS 30614 PLZEN, Czechoslovakia

Samenvatting

De auteurs hebben gezamenlijk gewerkt aan een vergelijkend onderzoek inzake klasse-identificatie methoden. De bedoelde methoden worden algemeen gerekend tot het domein van machine-learning. De onderzoeksexperimenten op het gebied van machine-learning hebben tot doel de kennisacquisitie ten behoeve van zgn. 'expert-systemen' of 'kennis-systemen' te ondersteunen of zelfs geheel te automatiseren. Een belangrijk onderdeel van kennisacquisitie vormt de klasse-identificatie. Hieronder wordt verstaan het toewijzen van een klasse (of concept) aan een object, dat beschreven is volgens een set variabele attributen.

Doel van het onderzoek was inzicht te verwerven in de prestatie verschillen tussen een aantal relevante klasse-identificatie methoden. Prestatie is in dit kader de mate, uitgedrukt in een percentage, waarin de bedoelde methoden in staat zijn een nieuw object juist te klassificeren aan de hand van een gegeven set voorbeelden (de zgn.'leer set' of 'training set'). De coördinatie van het onderzoek berustte bij Wim Schaefer.

1. Objecten van onderzoek

Doel van het onderzoek was inzicht te verwerven in de prestatie verschillen tussen klasse-identificatie methoden. Hiertoe zijn enkele relevante machine-learning methoden en statistische methoden geselecteerd. Deze zijn:

1. de 'AQ' methode, ontwikkeld door Michalski en geïmplementeerd door Wilfried Post e.a. (zie item 2).
2. de 'ID' methode, ontwikkeld door Quinlan en geïmplementeerd door Guido Cosemans in het programma 'SIPS' (zie item 3)
3. de 'Kx' methode, ontwikkeld en geïmplementeerd door Jan Hajek in het 'KnowledgeExplorer' programma (zie item 4).
4. de discriminant analyse
5. de logistische regressie analyse
6. Multi Level Classificatie ('MLC') methode (zie item 6), gemodificeerd en geïmplementeerd door Frantisek Vávra en Jiri Melichar.

De statistische methoden bedoeld onder punt 4 en punt 5, welke men kan bestempelen als 'klassiek' en 'beproefd', hebben deel uitgemaakt van het vergelijkende onderzoek om de betekenis te verduidelijken van de tot de 'post moderne' methoden behorende machine-learning. Ten behoeve van de genoemde statistische methoden zijn door J. Dijkstra routines ontwikkeld en is gebruik gemaakt van de programmatuur van SAS [9] (zie item 5).

Voor de onderlinge vergelijking van de geselecteerde methoden zijn twee sterk verschillende datasets gekozen. De eerste dataset bestaat uit 336 records, waarbij elk record bestaat uit 116 prediktoren en de klasse-variabele. Deze set van waarden in een record noemt men in het machine-learning jargon wel een 'feature vector' [1], terwijl statistici spreken van prediktoren. Zeven van de 116 prediktoren zijn nominaal met meer dan twee waarden. De hoeveelheid ontbrekende waarden in deze dataset bedraagt 59%. In de oorspronkelijke dataset had een groot deel van de ontbrekende waarden de status van zgn. 'niet-toepasbare-waarden' (not-applicable), maar voor de onderlinge vergelijkingen van de methoden zijn ontbrekende waarden en de not-applicable waarden samengevoegd tot de relatief grote groep van ontbrekende waarden. De tweede dataset bestaat uit 46 records, waarbij een record bestaat uit 7 prediktoren en de klasse-variabele. Alle prediktoren zijn voor deze dataset nominaal en er zijn geen ontbrekende waarden. Voor beide datasets

geldt dat de klasse-variabelen slechts 2 waarden kunnen hebben.

2. De 'AQ' methode.

Het voornaamste principe van het AQ-algoritme (ontwikkeld door Michalski [2]) is, dat er gezocht wordt naar een zo algemeen mogelijke conceptbeschrijving, die volledig consistent is met alle beschikbare positieve en negatieve voorbeelden. In tegenstelling tot statistische technieken, waarbij het zoeken naar kwantitatieve verbanden in een verzameling voorbeelden centraal staat, probeert AQ juist verbanden binnen individuele voorbeelden te vinden. Dit wordt bereikt door een positief voorbeeld te nemen en de kenmerken (attribuutwaarden) van dit voorbeeld te vergelijken met die van alle negatieve voorbeelden. Door op verschillende wijzen de beschrijvingen van dit positieve voorbeeld te generaliseren, worden meerdere beschrijvingen van dit voorbeeld gevonden, die voor geen enkel negatief voorbeeld opgaan.

Een enkele beschrijving is echter voldoende om als basis te dienen voor de uiteindelijke conceptbeschrijving, zodat er een selectie plaats vindt van de beste beschrijving aan de hand van een aantal criteria. Deze criteria kunnen door de gebruiker worden gespecificeerd. Een belangrijk criterium is bijvoorbeeld dat een beschrijving ook zoveel mogelijk andere positieve voorbeelden beschrijft. Een ander criterium is, dat de beschrijvingen voldoen aan een aantal preferenties, bijvoorbeeld ten aanzien van de begrijpbaarheid en de efficiency.

Als er nog positieve voorbeelden onbeschreven blijven, herhaalt het proces zich met deze overgebleven voorbeelden en alle negatieve voorbeelden tot dat uiteindelijk alle positieve voorbeelden zijn beschreven. Tenslotte worden alle gedeeltelijke concept-beschrijvingen samengevoegd tot de definitieve conceptbeschrijving.

Twee voorbeelden van zo'n conceptbeschrijving zijn:

| | | | | | |
|----------------------|---|--------------|----------------------|---|-----------------|
| Concept: | = | hartinfarct | Concept: | = | hyperventilatie |
| Beschrijving: | | | Beschrijving: | | |
| hevige_pijn | = | ja | pijn | = | nee |
| & plaats | = | midden_borst | & hartkloppingen | = | ja |
| & duur | = | 30-60_min. | & leeftijd | < | 30 |
| | | of | | | |
| kortademig | = | ja | | | |
| & misselijk | = | ja | | | |
| & leeftijd | > | 60 | | | |

In de loop der tijd zijn er op verschillende wijzen verfijningen op het basis idee aangebracht. In de ene variant werd het probleem van de ruis aangepakt, in de andere probeerde men verschillende manieren van inductie uit. En in weer een andere va-

riant werd getracht gebruik te maken van achtergrond kennis.

Het AQ-experiment van deze studie is uitgevoerd met een eigen implementatie van het basis-algoritme. Zowel de representatievorm van het concept als de wijze van generalisatie zijn in het experiment zo eenvoudig mogelijk gehouden en er is dus geen gebruik gemaakt van achtergrond kennis¹.

3. De 'SIPS' methode.

De bruikbaarheid van het ID3 algoritme (van Quinlan [4]) voor praktische probleem oplossing hangt ten zeerste samen met de mogelijkheden die het omhullende programma biedt aan de gebruiker om met het algoritme en de data te spelen. De meeste commercieel beschikbare implementaties bieden terzake een gebruikscomfort dat te vergelijken is met een aardappelmesje dat in plaats van aan een handvat, aan een zwaar betonblok bevestigd is. Mits de aardappel op gepaste wijze rond te draaien, kan hij wel geschild worden, maar handig is toch wat anders.

Onvrede met bestaande implementaties was aanleiding om een eigen implementatie van ID3 te programmeren. Een gelukkige omstandigheid was dat al in een vroeg stadium van ontwikkeling een Belgische multinational wel wat zag in het gebruik van inductie als hulpmiddel bij de statistische analyse van datasets met talrijke variabelen. Door systematisch data sets van het lopende onderzoek zowel met het in ontwikkeling zijnde programma SIPS als de gewone statistische pakketten te analyseren, werd SIPS verplicht zich te ontwikkelen tot een werktuig waarmee praktisch gerichte vragen kunnen beantwoord worden uitgaande van 'real-world' datasets, waarin foutieve en ontbrekende gegevens niet te vermijden zijn.

Zoals in de AIT-89 publicatie van Cosemans e.a. [5] werd medegedeeld, bood de methode van SIPS t.o.v. de klassieke statistische benadering niet alleen voordelen qua snelheid waarmee de gezochte informatie gevonden werd (30 minuten t.g.o. meer dan één week is een typische verhouding voor "moeilijke" data sets), maar ook inzake bruikbaarheid van de gevonden informatie. In een bepaald geval kon de informatie die met SIPS gevonden werd niet worden bevestigd met gewone statistische analyses van de zelfde data set. Deze economisch belangrijke informatie werd via een aantal laboratorium experimenten op juistheid gecontroleerd. Deze experimenten bevestigden de bruikbaarheid van de door SIPS gegenereerde hy-

¹ Er werd geen negatie en interne disjunctie gebruikt; alleen de zgn. 'Dropping Condition' generalisatie werd toegepast. 'Extension against' -, 'Adding Alternative' generalisatie e.d. zijn in de experimenten buiten beschouwing gelaten. (Voor een nadere uitleg van deze termen zie [3].)

pothese.

De analyses van de data sets, die voor dit onderzoek waren geselecteerd, gebeurden zonder te weten wat de betekenis van de variabelen was. Eén van de aange naamste ervaringen bij het gebruik van inductieve technieken, namelijk de 'Aha Erlebnis' van de domeindeskundige, die dank zij de beslissingsboom plotseling een inzicht krijgt in het probleem, was hier dus niet aanwezig. De beslissingsboomen waren bovendien onoorbaar lang; deze oefening was een stimulans te meer om een algoritme te implementeren waarmee een complexe beslissingsboom via veralgemening van de paden doorheen de boom wordt omgezet in een klein aantal produktieregels [6].

4. De 'Kx' methode.

De 'Kx' methode zoals geïmplementeerd in KnowledgeXplorer is een zelfstandige ontwikkeling van Jan Hajek. In het algemeen kan men stellen dat Kx steunt op de drie 'machtige en beproefde pilaren' van de informatica, namelijk: twee- & meerwaardige logica + statistiek + niet-al-te-triviale informatietheorie [7, 8]. De kunst van het ontwerpen van Kx was een synthese van deze drie wetenschappelijke elementen op zodanige wijze, dat er een synergetische samenwerking ontstond. De onderliggende infrastructuur (b.v. nieuwe virtually coded hashers) maakt de benodigde berekeningen überhaupt uitvoerbaar op een PC.

Kx is zodanig geïmplementeerd, dat het kan werken op een standaard (kleine) MS-DOS compatibele PC met 512 K, eventueel zonder harde schijf. Hiermee wordt duidelijk gemaakt welke hoeveelheid energie is gestoken in de 'soft-ware engineering'. Kx is een geavanceerd en goed getest prototype, maar nog net niet een beslissingspakketje voor de huisvrouw of huisman.

Een korte impressie van het geïmplementeerde programma luidt als volgt:

User I/O: Geen vuurwerk op het scherm; wel goede en snelle resultaten zelfs op de goedkoopste hardware. Keuze uit een compacte of uitgebreide rapportage op het scherm en op een schijf. Volautomatische evaluatie van de sensitiviteit, specificiteit en van de totale hit rate in % van de genomen beslissingen op de test-cases dataset.

Correctheid: wordt runtime bewaakt door tientallen ASSERTies en CHECKs.

Snelheid: optimaal dankzij de verwerking uitsluitend in het snelle geheugen, dus zonder disk-overlays.

Door de grenzen aan de groei van de code, zoals die worden opgelegd door de huidige Turbo Pascal op de goedkoopste hardware en door de (on)menselijke complexiteit van de code zijn er drie versies van Kx:

1. KxX alias White Box genereert de expliciete beslissingsregels (geen bomen) voor rules-driven expert systems. Alhoewel historisch de eerstgeborene, is KxX op dit ogenblik minder ontwikkeld dan zijn jongere broertje KxY.
2. KxY alias Black Box beslist d.m.v. impliciete Bayes inductie. KxY kan draaien met een snelle algoritme, welke zich door de menselijke tuning gemakkelijk laat beïnvloeden of met een dure self-optimizing algoritme welke zich niet al te veel laat begeleiden, afleiden of misleiden door de knowledge engineer.
3. KxZ alias Grey Box is een iets gestripte Black Box plus een mogelijkheid om de training set op een informatietheoretisch verantwoorde manier te analyseren en reduceren. De analyse is bijna optimaal en de reductie is (bijna) maximaal; absolute perfectie zou tot een combinatorische explosie leiden (het beruchte NP probleem). Zelfs zonder dat de parameters optimaal waren 'getuned' gaf KxY de hoogste prestaties.

5. De statistische methoden.

De gevolgde statistische benadering is in eerste instantie die van de discriminantanalyse [9]. Bij een trainingsset van voorbeelden bestaande uit een criteriumvariabele en verklarende variabelen (prediktoren), levert deze methode een beslissingsregel (... de functie) gebaseerd op deze trainingsset, die ons in staat stelt een nieuwe voorbeeld met een zo groot mogelijke kans toe te wijzen aan de juiste klasse. Hierbij zij opgemerkt dat ingeval van ontbrekende waarden bij een of meer variabelen het betreffende voorbeeld niet in de analyse wordt meegenomen. Bij veel ontbrekende waarden kan het aantal voorbeelden dusdanig klein worden (kleiner dan het aantal verklarende variabelen), dat deze methode niet toegepast mag worden. Bovendien wil men de dataset volledig houden om vergelijking met andere data-analyse technieken zo eerlijk mogelijk te doen. Een gangbare en eenvoudige methode om hieraan tegemoet te komen is het toewijzen van gemiddelden per groep aan de ontbrekende waarden.

Een ander aspect is dat ervan uitgegaan wordt dat de verklarende variabelen scalaire waarden bevat. Bij het voorkomen van nominale waarden worden daartoe dichotome dummy variabelen gecreëerd. Indien de criteriumvariabele in de datasets

een dichotome variabele is, wordt tevens de logistische regressie methode [9] toegepast. Press en Wilson [10] beschreven de relatieve voordelen van beide methoden, waarbij de logistische regressie in geval van een classificatie 'tool' de voorkeur verdiende.

Uit de experimenten blijkt discriminantanalyse beter te voldoen dan logistische regressie. De resultaten wijken echter maar nauwelijks af van de apriori kans, hetgeen deze methode tot een minder aanvaardbare benaderingswijze maakt.

6. Multi Level Classification (MLC) methode.

De MLC methode is gebaseerd op het klassieke hypothese toetsen door middel van de verhouding van 'likelihood'-functies. Het maakt gebruik van het fundamentele idee van sequentiële tests: als een beslissing niet geaccepteerd kan worden, dan moet de volgende meting worden uitgevoerd. Wij hebben dit idee enigszins gemodificeerd: indien een beslissing niet geaccepteerd kan worden voor het gekozen probabilistische model, dan moet een ander model worden gekozen waarvoor een beslissing wel mogelijk is. De methode vereist twee classificatie klassen [11]. Vanuit de optiek van applicaties is het redelijk om twee typen fouten te hanteren (dit is eveneens gebruikelijk voor het hypothese toetsen). Type 1- fout luidt: een patroon is geklassificeerd als klasse 2, terwijl het behoort tot klasse 1. Type 2 - fout luidt: een patroon is geklassificeerd als klasse 1, terwijl het behoort tot klasse 2. De eis dat beide fouten geminimaliseerd moeten worden leidt tot gevallen waarover geen besluit kan worden gevormd.

Als een classificatieregel wordt de score gebruikt die gelijk is aan het logaritme van de verhouding van de 'likelihood'-functies. In een training set worden beslissingslimieten vastgesteld zodat de waarschijnlijkheid van beide fouten niet wordt overschreden. Dit leidt echter tot een onbeslisbaar gebied. In de volgende stap worden nieuwe waarschijnlijkheden van gebeurtenissen in klassen opnieuw vastgesteld voor die gevallen waarvoor de vorige score in het onbeslisbare gebied viel. Verder worden nieuwe logaritmische scores en nieuwe grenzen vastgesteld, zodat de totale classificatie fouten niet de gegeven grenzen overschrijden. Deze wijze wordt voortgezet totdat de training set effectieve schattingen geeft van waarschijnlijkheden of totdat de vastgestelde grenzen voor waarschijnlijkheden onredelijk zijn.

De methode kan op verschillende manieren gemodificeerd worden. In deze vergelijkende studie was de volgende variant gebruikt. Op het eerste niveau was slechts

een limiet vastgesteld zodat de som van geschatte waarschijnlijkheden van beide fouten minimaal zou zijn. Vervolgens werd de limiet 'wijder' gemaakt op een manier dat binnen dit interval 50 % van de foute beslissingen zou liggen. Voor dit interval werd een nieuwe score berekend en - overeenkomstig het eerste niveau - opnieuw met slechts één beslissingsgrens. Op dit punt werd het proces gestopt.

Uit het voorgaande volgt dat de beschreven methode een statistische methode is en daarom wordt tenminste gedeeltelijk de onafhankelijkheid van de attributen verondersteld. Daarom geloven wij dat deze methode effectief kan zijn in die situaties, waarin syntactische methoden (afhankelijkheid wordt vaak verondersteld) falen en vica versa.

Daar de methode in de eerste plaats een volledige trainingset behoeft, wordt het ontbreken van een attribuutwaarde beschouwd als een nieuwe (fictieve) waarde van het attribuut wanneer een discriminatie functie wordt berekend. Voor de keuze van de attributen en voor het toekennen van hun onderscheidende vermogens is Kullback's maat voor informatie gebruikt [12].

7. Resultaten en slotbeschouwing.

In tabel 1 zijn de resultaten afgebeeld van ons onderzoek. De prestaties, de percentages van correcte beslissingen in de kolommen A t/m G, zijn als volgt tot stand gekomen. De grote dataset bestaande uit 336 voorbeelden is aselekt gesplitst in twee partities:

- (i) 80 % van het aantal records, gebruikt als leerset
- (ii) 20 % van het aantal records, gebruikt als testset

Zo is een kruisvalidatie opgezet door vijf maal twee partities te creëren van respectievelijk vijf leersets en vijf testsets. In de kolommen A t/m E staan de prestaties voor iedere leerset met de daarbij behorende testset. In kolom F staan de gemiddelde waarden van de voorgaande vijf kolommen.

De gemiddelde prestaties mbt. de tweede dataset staan in kolom G. Deze kleinere dataset, bestaande uit 46 records is aselekt in twee partities gedeeld van elk 23 records. De partities fungeerden respectievelijk als leerset en trainingset en omgekeerd als trainingset en leerset.

| | A | | B | | C | | D | | E | | F | G |
|---------------------|------|------|------|------|------|------|------|------|------|------|------|------|
| | ok | ? | ok | ? | ok | ? | ok | ? | ok | ? | ~% | ~% |
| 0 apriori | 58.2 | — | 58.2 | — | 58.2 | — | 58.2 | — | 58.2 | — | 58.2 | 63.0 |
| 1 AQ | 52.2 | 09.0 | 59.7 | 10.4 | 50.7 | 10.4 | 49.2 | 10.4 | 56.7 | 09.0 | 59.4 | 91.9 |
| 2 SIPS | 58.2 | 01.5 | 61.2 | 13.4 | 65.7 | 13.4 | 64.1 | 20.8 | 56.7 | 23.9 | 69.6 | 82.0 |
| 3 Kx | 71.6 | — | 80.6 | — | 77.6 | — | 79.1 | — | 77.6 | — | 77.3 | 93.5 |
| 4 Discrimin. | 47.8 | — | 62.3 | — | 59.7 | — | 59.7 | — | 56.7 | — | 57.2 | 73.7 |
| 5 Log. regr. | 40.3 | — | 47.8 | — | 41.8 | — | 43.3 | — | 46.3 | — | 43.9 | * |
| 6 MLC | 68.7 | — | 73.1 | — | 64.2 | — | 67.2 | — | 65.7 | — | 67.7 | 84.8 |

Tabel 1. Overzicht van de prestaties van de vergeleken methoden.

Op regel '0' staan de apriori kansen genoteerd. Dit zijn de frequenties van de meest voorkomende klassen in de testsets. Zowel bij de resultaten van 'SIPS' als 'AQ' bestaat er een groep niet-geclassificeerde gevallen (kolommen '?') naast de groepen correct-geclassificeerd (kolommen 'ok') en fout-geclassificeerd. Ten behoeve van de vergelijking met de andere methoden in dit onderzoek, is 58.2 % van het aantal niet-geclassificeerde gevallen, overeenkomstig de apriori kans, toegerekend aan de groep correct-geclassificeerd en verwerkt bij de gemiddelden van kolom F.

De hier gebruikte techniek van evaluatie, de kruisvalidatie, is slechts een van de manieren om de geselecteerde methoden op hun waarde te schatten. Een andere wijze van evalueren kan bijvoorbeeld plaats vinden door de geleerde concept-beschrijvingen meer genuanceerd te bestuderen. Hierbij zouden 'real-world' experts de geleerde beschrijvingen individueel moeten beoordelen. De gevonden klasse (concept) zou dan niet met 'goed' of 'fout' moeten worden beoordeeld (1 óf 0), zoals bij kruisvalidatie, maar de beschrijving zou kwalitatief moeten worden beoordeeld met expressies als: 'geeft de essentie goed weer', 'mist een belangrijk kenmerk', 'sommige kenmerken zijn onzinnig', of 'lijkt helemaal nergens op'.

De inhoud van een gevonden beschrijving kan namelijk aanknopingspunten bevatten voor nieuwe kennis eliciteringssessies met een expert, ondanks, of juist dankzij het feit dat de beschrijving niet volledig correct is. Belangrijk is dan wel dat de beschrijvingen inzichtelijk zijn afgebeeld. 'AQ', 'SIPS' en 'KxX' (maar niet 'KxY') sluiten wat dat betreft nauw aan bij de gebruikte representatie vormen van regelgebaseerde kennissystemen.

Tot slot nog een enkele opmerking over de geldigheid van het onderzoek. Een vergelijkbaar experiment met klasse-identificatietechnieken is eerder uitgevoerd door

Post en Van Someren [13] en leidde tot resultaten, die in een aantal opzichten afwijken van de hier gepresenteerde resultaten. Hoewel de opzet van het onderhavige onderzoek zo consistent mogelijk was, mag men slechts een relatieve waardering toekennen voor de geselecteerde methoden. De status quo van de methoden draagt in veel opzichten immers nog het predikaat 'dynamisch'. Algemeen geldt, dat indien er sprake is van relatief veel nominale variabelen en van relatief veel ruis (zoals zeker door ontbrekende waarden en misschien ook door inconsistente waarnemingen), de statistische methoden een lage prestatie leveren in verhouding tot 'machine-learning' technieken. Het behoeft echter nog veel experimenteel onderzoek om meer genuanceerd vast te kunnen stellen welke techniek voor welke dataverzameling het meest geëigend is.

Bibliografie:

- [1] Forsyth, R., Rada, R.: 'Machine Learning, Applications in Expert Systems and Information Retrieval'. Ellis Horwood Ltd, England, 1986
- [2] Michalski, R.S., Mozetic, I., Hong, J., and Lavrac, N.: 'The AQ15 inductive learning system: an overview and experiments', Report UIUCDCS-R-86-1260, Department of Computer Science, 1986b, University of Illinois at Urbana-Champaign.
- [3] Michalski, R.S.: 'A theory and methodology of inductive learning' in: Michalski, R.S. et al., editors, 'Machine Learning. An Artificial Intelligence Approach', Palo Alto, CA, 1983
- [4] Quinlan J.R. : 'Learning Efficient Classification Procedures and their Application to Chess End Games', in 'Machine Learning'. Michalski and Carbonell Eds., 1984, Springer Verlag, pp.463-482
- [5] Cosemans G., Kretzschmar J., Samyn J.: 'Inductieve Analyse van Productieprocessen met SIPS', in AIT Proceedings 1989. Publ. NGL, Stichting Informatica Congressen, Amsterdam
- [6] Quinlan J.R.: 'Simplifying Decision Trees', in 'Knowledge Acquisition for Knowledge-based Systems', Volume I. Gaines and Boose Eds., 1988, Academic Press, pp.241-254
- [7] Shannon C.E., Weaver W., 'Mathematical Theory of Communication', Univ. of Illinois Press, 1949
- [8] Wozencraft J.M., Jacobs I.M., 'Principles of Communication Engineering', 1967, John Wiley
- [9] SAS/STAT, SAS Institute Inc.
- [10] Press, S.J. and Wilson, S.: 'Choosing between logistic regression and discriminant analysis', 1978, J. Americ. Statist. Assoc. 73, 699-705
- [11] Vávra F. Classification of a non-numerical pattern. In Automatizace (Czech), Nr. 6. Prague; 1989
- [12] Kullback S. Information Theory and Statistics. New York: John Wiley , 1960
- [13] Post W., Van Someren M.: 'Een vergelijkende evaluatie van enkele technieken voor inductief leren', in Proceedings AI Toepassingen 1988, pp.451-459