

SAS Macro zur Analyse von Diagnosemarkern mittels p-Wert-basierter CART-Analyse

Martina Malzer
IMBE
Universität Erlangen-Nürnberg
Waldstr. 6
91054 Erlangen
martina.malzer
@imbe.med.uni-erlangen.de

Annette Pfahlberg
IMBE
Universität Erlangen-Nürnberg
Waldstr. 6
91054 Erlangen
annette.pfahlberg
@rzmail.uni-erlangen.de

Janice Hegewald
IMBE
Universität Erlangen-Nürnberg
Waldstr. 6
91054 Erlangen
janice.hegewald
@imbe.imed.uni-erlangen.de

Olaf Gefeller
IMBE
Universität Erlangen-Nürnberg
Waldstr. 6
91054 Erlangen
olaf.gefeller
@rzmail.uni-erlangen.de

Zusammenfassung

Im medizinischen Umfeld dienen diagnostische Marker, die aus verschiedenen Diagnoseverfahren gewonnen werden, zur besseren Diagnostik einer Erkrankung. Ein einzelner Marker allein ist oft sehr unzuverlässig und besitzt eine geringe Aussagekraft über das Vorliegen / Nichtvorliegen einer Krankheit. Deshalb müssen oft mehrere Marker kombiniert betrachtet werden, um eine klare Abgrenzung eines Krankheitsbilds zu bekommen. Als auswertendes Verfahren kann eine p-Wert-basierte CART-Analyse (Classification And Regression Trees [1]) verwendet werden, sodass am Ende der Analyse als Resultat ein Klassifikationsbaum steht, der eine Unterstützung für eine bessere Diagnosestellung darstellen kann. Baumanalysen werden im Vergleich zu anderen statistischen Verfahren von Klinikern gerne verwendet, da sie aufgrund ihrer grafischen Darstellung leichter zu interpretieren sind. Für die p-Wert-basierte CART-Analyse existieren Kriterien für die Generierung eines Klassifikationsbaums, wie z. B. die Bildung der Splits nach dem kleinsten p-Wert. Das Verfahren konnte durch das von uns implementierte SAS Macro DIACART automatisiert werden. Die Anwendung des Macros wird anhand eines Beispiels zur Kombination verschiedener laborchemischer und radiologischer Diagnosemarker zur Diagnosesicherung innerhalb eines Studiendatensatzes über Creutzfeldt-Jakob-Verdachtsfälle aufgrund von Meldungen an das Deutsche Creutzfeldt-Jakob Surveillance System illustriert [2].

Schlüsselworte: CART, Diagnosemarker, diagnostische Tests, Klassifikation, SAS Macro.

1 Einleitung

Bei manchen Erkrankungen ist es nicht möglich eine sichere Diagnose zu Lebzeiten des Patienten zu stellen. Diese kann erst mittels aufwändiger Verfahren nach dem Tod des Patienten geschehen, z. B. durch eine Autopsie. Im medizinischen Bereich existieren als Unterstützung für die Diagnosestellung zu Patientenlebzeiten bereits verschiedene diagnostische Verfahren, aus denen die entsprechenden Diagnosemarker gewonnen werden können. Jedoch erweist sich ein einzelner Diagnosemarker für den möglichst frühzeitigen Nachweis über die Existenz einer Erkrankung oft als nicht aussagekräftig genug. Deshalb ist es sinnvoll, die Information verschiedener diagnostischer Marker so miteinander zu kombinieren, dass bezüglich einer Diagnosestellung zu Lebzeiten eine Verbesserung erzielt werden kann.

Für die gemeinsame Analyse der verschiedenen Diagnosemarker stellt eine p-Wert-basierte CART-Analyse ein geeignetes Verfahren dar. Am Ende der Baumanalyse soll ein Klassifikationsbaum stehen, der eine aussagekräftige Kombination der einzelnen Marker in sich vereint.

Für die Realisierung einer p-Wert-basierten CART-Analyse innerhalb der SAS Software haben wir ein automatisiertes SAS Macro entwickelt.

Die Funktion des SAS Macros DIACART soll später im Kontext genauer erläutert werden.

2 Diagnostische Tests

Eine Grundlage für die Kombination der verschiedenen Marker, die in die CART-Analyse unabhängig voneinander mit eingehen sollen, sind diagnostische Tests. Von einem „guten“ Diagnostetest sollen möglichst alle Kranken als krank und alle Gesunden als gesund erkannt werden. Die Genauigkeit des diagnostischen Tests wird i. d. R. über das genaue Wissen über das Vorliegen der Erkrankung, den sog. „Goldstandard“, beurteilt. Der Goldstandard ist meist ein herkömmliches Verfahren zur Diagnosestellung, mit dem das sichere Vorliegen einer Krankheit festgestellt werden kann, wie z. B. durch eine Autopsie das sichere Vorhandensein einer Creutzfeldt-Jakob-Erkrankung diagnostiziert wird.

Aus der Vielzahl von Kenngrößen bezüglich der Güte eines diagnostischen Tests können z. B. Sensitivität / Spezifität und der Youden-Index, als davon abgeleitete Maßzahl, gewählt werden. Zur Erinnerung: Für die Definition der drei Gütemaße gilt:

- Sensitivität = $P(\text{positiver Test} \mid \text{krank})$
= $P(\text{positiver Test und krank}) / P(\text{krank})$
- Spezifität = $P(\text{negativer Test} \mid \text{gesund})$
= $P(\text{negativer Test und gesund}) / P(\text{gesund})$
- Youden-Index = Sensitivität + Spezifität - 1

Aus einer allgemeinen vereinfacht dargestellten 2 x 2-Kontingenztafel, die zwei kategoriale Merkmale wie Krankheitszustand und Diagnosetest in ihrer gemeinsamen Verteilung darstellt, lassen sich Sensitivität und Spezifität wie folgt berechnen:

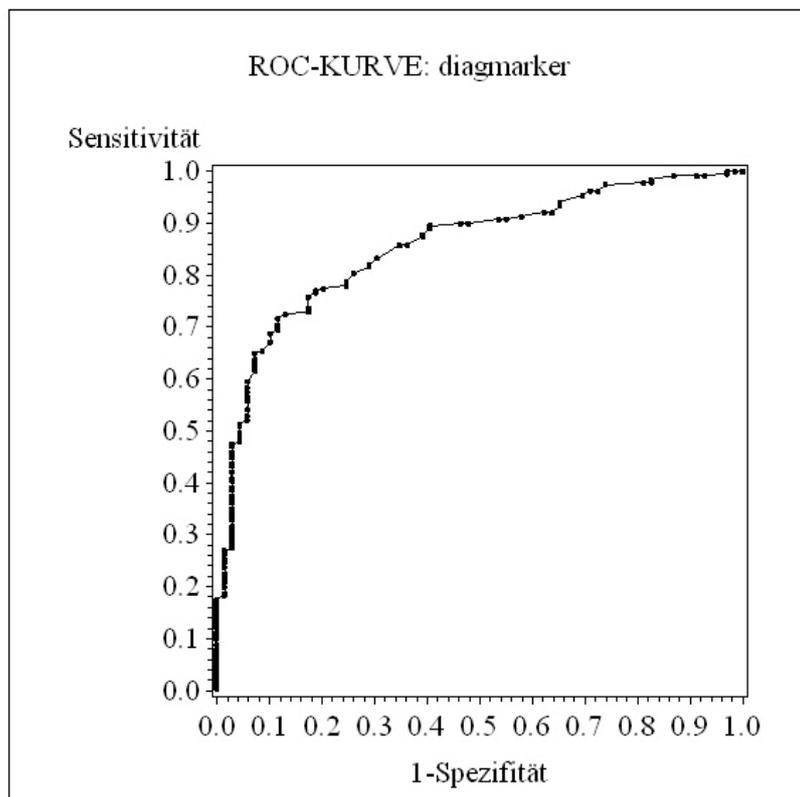
		Krankheitszustand	
		krank	nicht krank
Diagnosetest	positiv	<i>richtig positiv</i> (a)	<i>falsch positiv</i> (b)
	negativ	<i>falsch negativ</i> (c)	<i>richtig negativ</i> (d)

$$\text{Sensitivität} = \frac{a}{(a+c)}$$

$$\text{Spezifität} = \frac{d}{(b+d)}$$

Normalerweise sollten bei der Anwendung eines Tests immer Sensitivität und Spezifität gemeinsam betrachtet werden. Aber im klinischen Alltag können in Abhängigkeit des Tests zu einer Krankheit und der damit zusammenhängenden Therapie eine höhere Sensitivität und niedrigere Spezifität oder eine niedrigere Sensitivität und höhere Spezifität betrachtet werden. Es ist also auch abhängig davon, was erreicht werden soll: wenige falsch positive Ergebnisse oder eher wenige falsch negative Ergebnisse.

Der Zusammenhang zwischen Sensitivität und Spezifität in Abhängigkeit eines „Cutpoints“ zur Dichotomisierung eines stetigen Diagnosemarkers kann mittels ROC-Kurven für einen Diagnosetest dargestellt werden. Ein Beispiel hierzu aus dem Studiendatensatz:



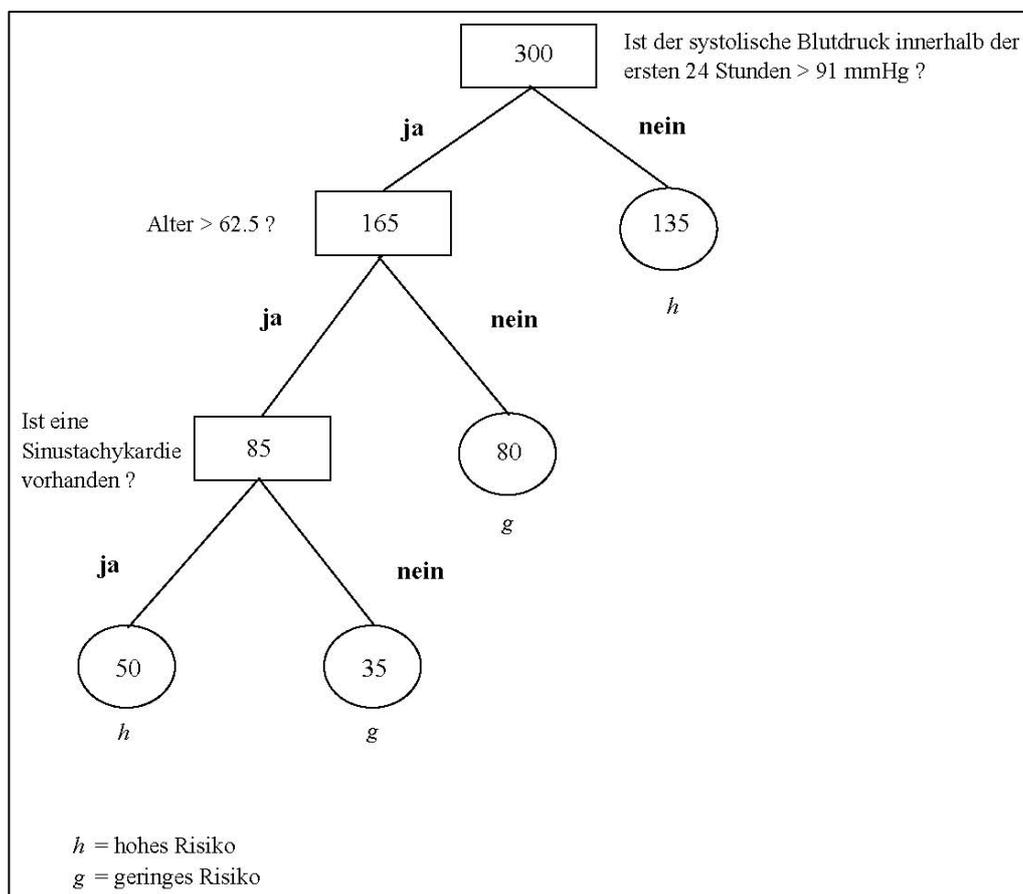
In der ROC-Kurve wurde für jeden Messwert als möglichen Cutpoint die Sensitivität und das falsch positive Testergebnis ($1 - \text{Spezifität}$) abgetragen.

Bei quantitativen Diagnosetests besteht somit die Möglichkeit mittels ROC-Kurven die quantitativen Werte zu dichotomisieren, d. h. quantitative Tests werden in dichotome überführt. Für die Findung eines optimalen Schwellenwertes für einen Test wird derjenige Cutpoint einer ROC-Kurve gewählt, der den größten Abstand zu einer gedachten Winkelhalbierenden innerhalb der beiden Achsen besitzt. An diesem gefundenen Cutpoint ist die Trennung der Kranken und Gesunden am besten sein.

Innerhalb unseres zur Verfügung gestellten Studiendatensatzes (Creutzfeldt-Jakob-Verdachtsfälle) wurden alle stetigen Merkmale (Diagnosemarker) in dieser Weise dichotomisiert bevor sie in die CART-Analyse als Variable mit eingingen.

3 CART-Analyse

CART (Classification And Regression Trees) stellt erst später nach der Weiterentwicklung von Breiman et al. [1] einen bedeutenden Algorithmus zur Entscheidungsfindung in Form von meist binären Baumstrukturen dar. Dabei werden für eine Entscheidung pro Verzweigung höchstens zwei Äste gebildet. Eine Entscheidung für die Einteilung in binäre Äste wird durch einfache Fragestellungen getroffen, die meist mit „ja“ oder „nein“ zu beantworten sind. Gerade im medizinischen Umfeld treffen Fragestellungen im Bezug auf eine durch den Arzt zu findende Diagnose zusammen, nach denen ein Patient immer weiter – durch mehr und mehr diagnostische Kriterien – klassifiziert werden kann. Er wird anhand aller Fragestellungen klassifiziert, entweder unter einem hohen Risiko für eine mögliche Erkrankung oder unter einem geringeren Risiko zu stehen. Ein Beispiel von Breiman et al. sind die Risikofaktoren für einen Herzinfarkt, welche innerhalb eines Klassifikationsbaums dargestellt werden können:



Der Baum besteht aus Ästen und Knoten. Bei den Knoten wird unterschieden zwischen dem Wurzelknoten, den Entscheidungsknoten und den Endknoten (Blätter – Knoten, die nicht mehr gesplittet werden können). Der Wurzelknoten ist der oberste Knoten im Baum. Zu jedem Entscheidungsknoten gehört eine Fragestellung, die mit „ja“ oder „nein“ beantwortet werden kann. Wird die Fragestellung mit „ja“ beantwortet, wird im linken Ast des Baumes zum nächsten Knoten gegangen. Ist die Fragestellung mit „nein“ zu beantworten, so wird im rechten Ast zum nächsten Knoten gegangen.

Im obigen Beispiel des Klassifikationsbaums ist das Zielmerkmal das Auftreten eines Herzinfarkts („ja“/„nein“). Die erste Fragestellung dazu lautet:

Ist der systolische Blutdruck innerhalb der ersten 24 Stunden > 91 mmHg?

Wird die Frage mit „ja“ beantwortet, so wird durch den linken Ast bis zum nächsten Knoten gegangen. Die anderen Fragen (Merkmale) werden solange gestellt, bis die Endknoten des Baumes erreicht sind. In dieser letzten Baumebene kann die eigentliche Klassifikation stattfinden, es ist abzulesen wie viele unter einem hohen Risiko bzw. einem geringen Risiko für das Auftreten eines Herzinfarkts stehen.

Für weiteres Interesse an CART soll an dieser Stelle auf die Literatur [1] verwiesen werden.

Für die statistische Auswertung anhand des Macros DIACART unter der Kombination von Diagnosemarkern soll eine p-Wert-basierte CART-Analyse durchgeführt werden. Die CART-Analyse erfolgt also auf der Basis der Ergebnisse statistischer Tests. Das bedeutet, dass die p-Werte für den vom Programm zu generierenden Klassifikationsbaum aus einer Teststatistik hervorgehen sollen. Aufgrund der zugrunde liegenden Struktur einer 2 x 2-Kontingenztafel handelt es in unserem Fall um die Teststatistiken des Chi-Quadrat-Tests bzw. des exakten Test von Fisher. Als Ergebnis der Analyse erhält man einen einzigen Klassifikationsbaum zur Stellung der Diagnose.

Das kategorielle Zielmerkmal ist das Vorliegen / Nichtvorliegen einer Krankheit und alle stetigen Diagnosemarker gehen nach Dichotomisierung als binäre Merkmale in das Analyseverfahren mit ein. So besitzen diese innerhalb der Baumanalyse die gleichen Anzahlen von Merkmalsausprägungen, sodass pro Variable und Knoten immer nur ein Split erfolgen kann. Dadurch ist eine binäre Struktur des Klassifikationsbaums garantiert. Durch die vorausgegangene Dichotomisierung aller verwendeten stetigen Merkmale liegt für das statistische Testen (Teststatistik) zur Findung des kleinsten p-Werts jeweils eine Vierfeldertafel vor, wobei jeder zu testende Marker (Variable) mit einem Freiheitsgrad von 1 in die jeweilige Teststatistik eingeht. Deshalb besteht hier die Option, dass als Splitkriterium für die Knotengenerierung des Klassifikationsbaums entweder der größte aus den Teststatistiken resultierende kritische Wert für χ^2 oder der kleinste zugehörige P-Wert gewählt wird. Die Splits erfolgen mittels p_{\min} . Dabei darf der P-Wert ein vorgegebenes Signifikanzniveau α nicht überschreiten. Es muss also für jeden errechneten P-Wert gelten: $p < \alpha$. Für jeden Split wird durch einen Vergleich der P-Werte der kleinste P-Wert als optimales Splitkriterium herausgesucht.

Für das automatisierte Macro wurde auf jeder Ebene der exakte Test nach Fisher als statistischer Test für das Finden der kleinsten P-Werte ausgewählt – denn je weiter man in einem Klassifikationsbaum den Blättern näher gelangt, desto kleiner werden die Stichprobenumfänge in den Knoten (n). Eigentlich könnte man bei großen n zuerst den Chi-Quadrat-Test (im oberen Baum) durchführen und Fisher's exakten Test erst wenn man weiter im Baum Richtung Endknoten gelangt. Da es keine Einschränkung für die Rechenzeit bedeutet, den exakten Test nach Fisher zu benutzen und der Programmieraufwand dadurch geringer ist, wurde auf allen Ebenen der exakte Test nach Fisher gerechnet.

Als Stoppkriterium für den Klassifikationsbaum kann die Anzahl der Beobachtungen (n) in den Knoten herangezogen werden. Hier soll gelten: $n \leq g$. Ein Grenzwert g kann beliebig gewählt werden. Für das SAS Macro DIACART wurde als Stoppkriterium $n \leq 10$ festgesetzt.

Dieses Analyseverfahren bietet folgende Vorteile:

- Leicht interpretierbare Grafik (Baumstruktur)
- Interaktion von Variablen wird aufgedeckt (Zusammenwirken der kombinierten Marker)
- Baumgröße (Wachstum) ist bekannt
- Hohe Flexibilität in Bezug auf die Datenmenge

4 SAS Macro DIACART

Für die Analyse der Kombination von Diagnosemarkern als Unterstützung bei einer frühzeitigen Diagnosestellung für eine Erkrankung steht ein SAS Macro zur Verfügung. Für die Überführung der für die Analyse ausgewählten Methodik in eine programmtechnische Anwendung dient als Kern hauptsächlich die SAS-Prozedur FREQ. Die Bildung der einzelnen Baumknoten durch Splitten erfolgt durch mehrmalige Ausführung von PROC FREQ jeder abhängigen Variable (Marker) mit der Zielvariable (krank / nicht krank). Die Beobachtungen mit fehlenden Werten (Missings) in den bei der CART-Analyse berücksichtigten Variablen wurden von der Analyse ausgeschlossen. Die Funktion des Macros wurde anhand eines Studiendatensatzes mit über 728 Creutzfeldt-Jakob-Verdachtsfälle getestet. Das SAS Macro soll nun beispielhaft mittels dreier beliebig ausgewählten Marker illustriert werden.

Allgemeine Syntax des Aufrufs des Macros DIACART:

```
%DIACART(datei => Name der auszuwertenden Datei,
          markervar=Liste der Variablen (Marker),
          alpha=Signifikanzniveau für die Berechnung der Splits
              (Default=0.05),
          kzustand=Variable für ODS-Output über den Klassifikationsbaum für
              alle erkrankten Individuen (kzustand=1) oder alle gesunden
              Individuen (kzustand=0); (Default=1)
          ziel=Zielvariable, die Kodierung muss vorher festgelegt werden: 1=ja
              (krank), 0=nein (nicht krank)
          );
```

Beispiel für den Aufruf des Macros DIACART:

```
%DIACART(datei => befund.tree,
          markervar=diagmarker1 diagmarker2 diagmarker3,
          ziel=krank
```

); =>Übernahme der Defaulteinstellungen für alpha und kzustand!
 Pro Split wird jede Markervariable verglichen mit der Zielvariable krank.
 Für die Variable diagmarker1 wurde folgender Output erhalten:

The FREQ Procedure

Table of diagmarker1 by krank

diagmarker1		krank		
Frequency		1	0	Total
Percent				
Row Pct				
Col Pct				
1	442	32	474	
	60.71	4.40	65.11	
	93.25	6.75		
	89.11	13.79		
0	54	200	254	
	7.42	27.47	34.89	
	21.26	78.74		
	10.89	86.21		
Total	496	232	728	
	68.13	31.87	100.00	

Statistics for Table of diagmarker1 by krank

Statistic	DF	Value	Prob
Chi-Square	1	394.7365	<.0001
Likelihood Ratio Chi-Square	1	414.1408	<.0001
Continuity Adj. Chi-Square	1	391.4279	<.0001
Mantel-Haenszel Chi-Square	1	394.1943	<.0001
Phi Coefficient		0.7364	
Contingency Coefficient		0.5929	
Cramer's V		0.7364	

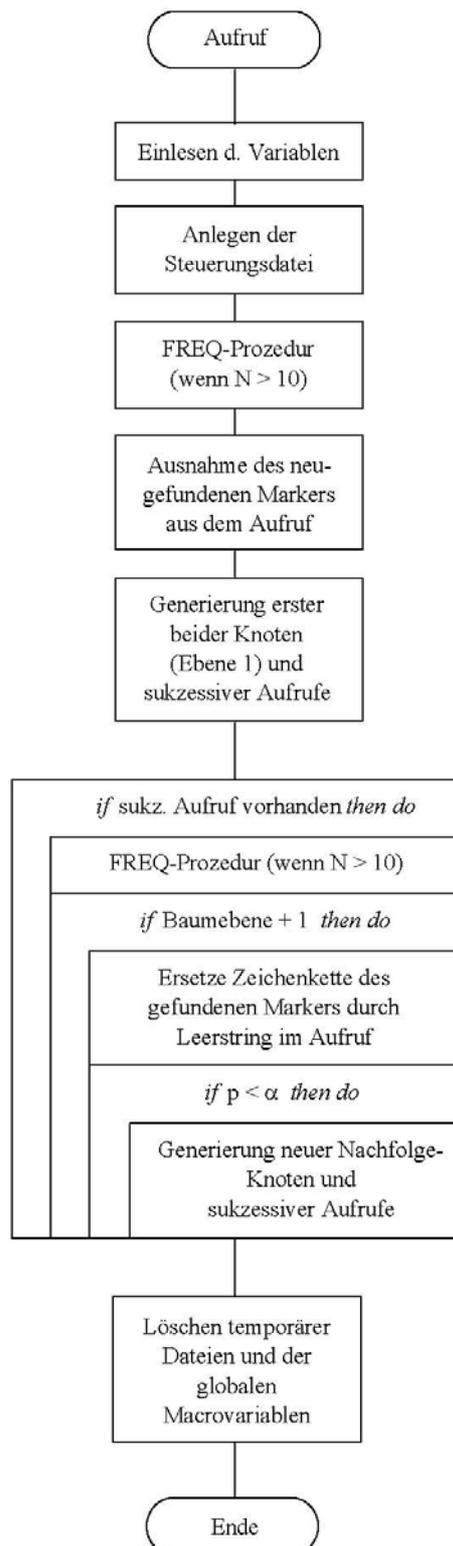
Fisher's Exact Test

Cell (1,1) Frequency (F)	442
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	1.682E-91
Table Probability (P)	8.036E-90
Two-sided Pr <= P	1.682E-91

Sample Size = 728

Ein in der Liste enthaltener Marker (s. Macroaufruf) kann pro noch zu generierendem Knoten nur solange berücksichtigt werden, bis dieser im jeweiligen Ast noch nicht verwendet wurde. Danach soll er nicht mehr für die Analyse berücksichtigt werden, da pro Variable und Teilbaum immer nur ein Split erfolgen soll. Es wird solange gesplittet, bis das Signifikanzniveau überschritten ist bzw. die Anzahl der Beobachtungen im jeweiligen Knoten weniger als 10 beträgt.

Zur Verdeutlichung des Macros DIACART soll folgendes Flussdiagramm beitragen:



Das Macro besteht aus einem Hauptmacro (DIACART) und sieben Untermacos (*wort*, *knoten*, *pwert*, *filter*, *name*, *ausgabe* und *loesche*). Nach Aufruf des Macros DIACART werden die Datensatzvariablen (Marker) aus der Variablenliste (*markervar=*) durch das Untermacro *wort* in Macrovariablen eingelesen. Das Untermacro *knoten* legt die Reihenfolge zur Knotengenerierung fest. Es werden zur Orientierung für den Durchlauf der Knoten abhängig von der Anzahl der Markervariablen die möglichen Permutationen aus 0 für negativ und 1 für positiv – also die Folge 0, 1, 00 usw. – berechnet. Für den ersten Split wird mithilfe des Untermacos *pwert* der kleinste P-Wert gefunden. Durch Untermacro *filter* wird garantiert, dass ein Marker nicht mehr als einmal pro Ebene mitlaufen kann. Für die Information, welcher Marker bereits einmal in die Analyse eingeschlossen wurde, und die Generierung von sogenannten „Voraufrufen“ (sukzessive Aufrufe) als Information für die Grundlage der nächsten Splits (wenn $n > 10$) dient Untermacro *name*. Nach Durchlauf des gesamten Analyseverfahrens werden die Informationen über die Knoten und deren Kennzahlen per ODS (Output Delivery System) als eine RTF-Datei ausgegeben (Untermacro *ausgabe*). Für die Findung des ersten Splits werden die Untermacos aus didaktischen Gründen vom Hauptmacro DIACART sequentiell aufgerufen und ausgeführt. Alle weiteren Splits werden innerhalb einer Schleife (s. Flussdiagramm) gebildet, solange bestimmte Bedingungen erfüllt sind. Zum Schluss werden alle während der Analyse erzeugten temporären Dateien und globalen Macrovariablen gelöscht (Untermacro *loesche*).

Der Anwender erhält nun folgenden RTF-Output (Untermacro *ausgabe*), welches nach der Ausführung des Programmes erzeugt wird:

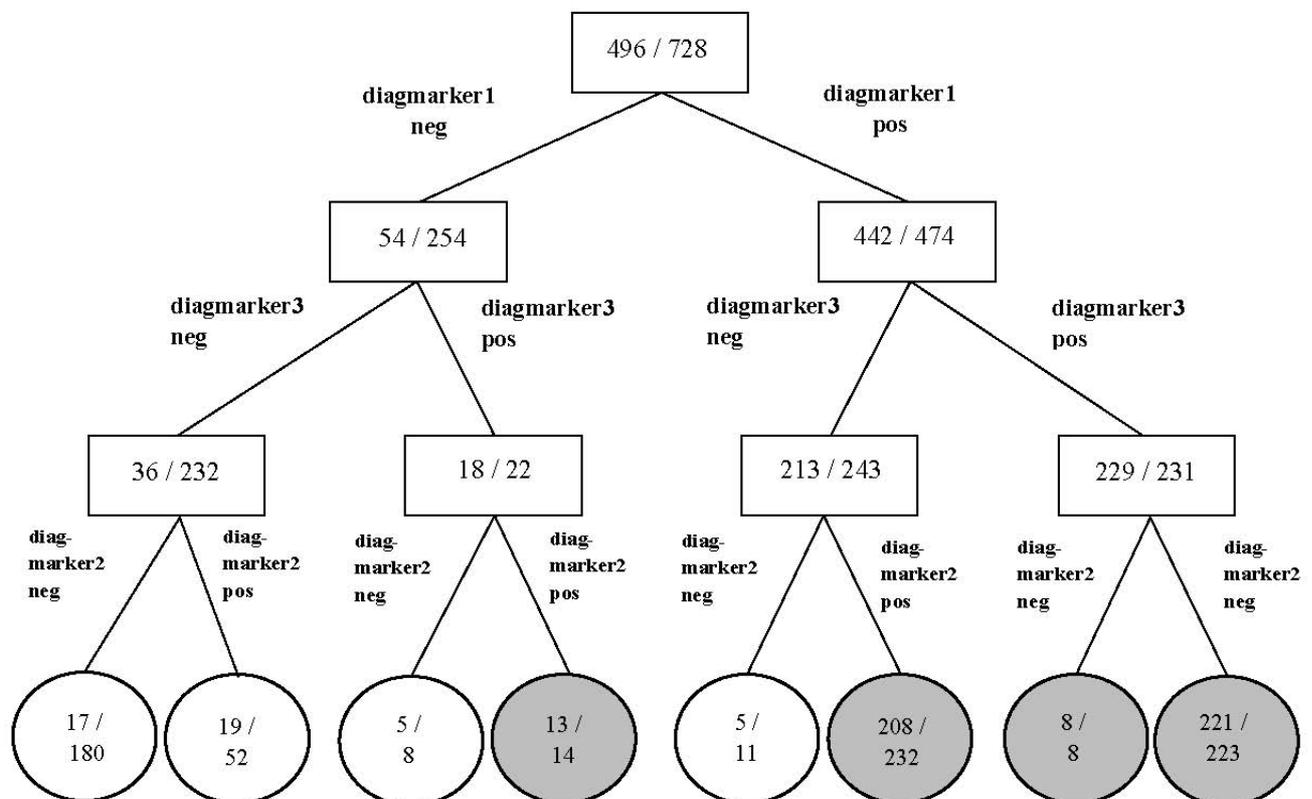
Ergebnisse

Gesamte Fälle	N	Anteil Fälle
496	728	0.681

Knoten	Anzahl Erkrankter	Anzahl Getesteter	Anteil getesteter Fälle	Klassifikation
diagmarker10	54	254	0.213	nicht krank
diagmarker11	442	474	0.932	krank
diagmarker300	36	232	0.155	nicht krank
diagmarker301	18	22	0.818	krank
diagmarker310	213	243	0.877	krank
diagmarker311	229	231	0.991	krank
diagmarker2000	17	180	0.094	nicht krank
diagmarker2001	19	52	0.365	nicht krank
diagmarker2010	5	8	0.625	nicht krank
diagmarker2011	13	14	0.929	krank
diagmarker2100	5	11	0.455	nicht krank
diagmarker2101	208	232	0.897	krank
diagmarker2110	8	8	1.000	krank
diagmarker2111	221	223	0.991	krank

In der Spalte Knoten der RTF-Ausgabe sind die auf jeder Ebene gefundenen Diagnosemarker (diagmarker1 bis diagmarker3) aufgelistet. An deren Variablennamen wurden zur besseren Orientierung als ergänzende Information die jeweils zugehörigen Zahlenkombinationen aus 0 und 1 angehängt. So wird z. B. im ersten Knoten diagmarker10 gefunden - diagmarker1 mit angehängter 0 für ein negatives Testergebnis (1 wäre positiv). Werden die Knoten innerhalb der einzelnen Ebenen betrachtet, so ist bekannt, dass eine angehängte Ziffer für die erste Ebene steht, eine zweite Zahl für die Ebene zwei und folglich eine dritte Ziffer für drei Ebenen.

Zur besseren Übersicht und Interpretation lässt sich die Ausgabe des Macros DIACART als ein Klassifikationsbaum darstellen:



Für eine Interpretation sind die Blätter des Klassifikationsbaums zu betrachten. Hier findet die eigentliche Klassifikation statt, wobei auch eine Aussage über die Trennkraft des Baumes gemacht werden kann. In den Blättern spiegeln sich die verschiedenen Kombinationen der einzelnen Marker wieder. Für eine Möglichkeit der Interpretation können jeweils die Anteile in den Blattknoten mit dem Anteil aus der Wurzel verglichen werden (hier: Anteil der Kranken am Gesamtkollektiv). Im Baum lässt sich in der Baumwurzel ein Anteil der Kranken von $\frac{496}{728}$ beobachten – dies entspricht einem Anteil von 0.681.

Dieser Anteil wird als a priori Verteilung für die Erkrankten genutzt, um die Blätter zu klassifizieren. Der errechnete Zahlenwert bildet also einen Vergleichsmaßstab mit dem

die Zahlenwerte aus den Anteilen in den Blattknoten verglichen werden sollen. Das Ziel eines Klassifikationsbaums ist die Gruppen (abhängig von der Zielvariablen krank/nicht krank) optimal voneinander zu trennen – also die Trennung in eine Gruppe Kranke und eine Gruppe Gesunde. Der Anteil der gesamt erkrankten Individuen von 0.681 wird den Werten aus den Blättern gegenübergestellt. Die Werte der Blätter müssen dabei über dem Vergleichswert von 0.681 liegen, damit ein Blattknoten als krank definiert werden kann. Betrachtet man beispielsweise die Kombination rechts außen: diagmarker1 pos und diagmarker3 pos und diagmarker2 pos so ergibt sich hier ein Anteil von 0.991. Es ist erkennbar, dass der eben errechnete Anteil des „positiven“ Blattknotens (positiv bedeutet: Beobachtungen mit positiven Testergebnissen gegeben sie sind krank) weit über dem Anteil der gesamt Erkrankten liegt – somit kann dieses Blatt als krank identifiziert werden.

Um herauszufinden bzw. zu prüfen, welche Güte nun das kombinierte Diagnoseverfahren besitzt können Gesamtsensitivität und –spezifität, sowie der Youden-Index ausgerechnet werden. Dazu können die „kranken“ und „gesunden“ Blattknoten herangezogen werden. Für eine Beurteilung müssen die Kombinationen der Tests (Blätter) in eine Variable zusammengefasst werden, die die Ausprägungen krank und gesund besitzen soll.

Dies lässt sich in einer Vierfeldertafel darstellen:

Table of var by krank

var(Variable für Testkombination aus CART)			
krank			
Frequency	ja	nein	Total
krank	450	27	477
gesund	46	205	251
Total	496	232	728

Aus dieser 2x2-Kontingenztafel ergeben sich nun die Gesamtgütemaße:

- Sensitivität = 90.7%
- Spezifität = 88.4%
- Youden = 0.79

Vergleicht man nun die Gütemaße aus dem kombinierten Testen mit denen eines einzelnen Diagnostetests (z. B. diagmarker1: Sensitivität=89.1%, Spezifität=86.2%, Youden=0.75), so kann festgestellt werden, dass eine deutliche Verbesserung mittels der Testkombination erreicht wurde.

Würde dieser Klassifikationsbaum als eine Unterstützung für das klinische Umfeld verwendet werden, so müssten die diagnostischen Tests bei positiven Testergebnissen genau in dieser Reihenfolge durchgeführt werden (diagmarker1 und diagmarker3 und

diagmarker2). So kann ein als gesamt positiv befundenes Individuum in diesem Fall als unter einem hohen Risiko für eine Erkrankung stehend klassifiziert werden.

Die Fehlklassifikation beträgt hierbei 10% ((falsch positive + falsch negative) / N).

Natürlich könnte auch 0.5 als Cut für die Klassifikation krank / nicht krank gewählt werden, damit würde aber hier die Sensitivität des kombinierten Testens stärker gewichtet werden, da mehr Kranke vorhanden sind. In diesem Fall wird ein zusätzlicher „kranke“ Knoten erhalten ($5/8 = 0.625 > 0.5!$). Die Vierfeldertafel würde sich dann folgendermaßen ändern:

Table of var by krank

var(Variable für Testkombination aus CART)			
krank			
Frequency	ja	nein	Total
krank	455	30	485
gesund	41	202	243
Total	496	232	728

Für die Gütemaße ergeben sich: Sensitivität=91.7%, Spezifität=87.06%, Youden=0.788 (vgl. mit Gütemaße aus vorheriger Vierfeldertafel!).

Die Fehlklassifikation beträgt 9,7% (falsch positive: 30, falsch negative: 41).

Wenn wir als Verteilung der Kranken 0.5 gewählt hätten, wäre unsere Fehlklassifikationsrate kleiner gewesen, jedoch hätten wir Sensitivität und Spezifität nicht mehr als gleichrangige Validitätsmaße für das diagnostische Vorgehen betrachtet. Da dies in der Regel nicht der typischen Anwendungssituation diagnostischer Tests entspricht, haben wir als Standard im Macro DIACART die a priori Verteilung der Kranken als Vergleichsmaßstab für die Klassifikation der Blätter benutzt.

5 Ausblick

Bisher weist in unserem Beispiel die Anwendung des Macros DIACART eine Fehlerklassifikationsrate von 10% für die Diagnosestellung auf. Diese naive Quantifizierung der Fehlerrate ist in der Regel deutlich zu optimistisch, da sie auf derselben Datenbasis berechnet wird, die zur Entwicklung des Klassifikationsbaums verwendet wurde. In einer zukünftigen Version wird das Macro daher eine zusätzliche Schätzung der Fehlklassifikationsrate auf der Basis der Anwendung von Resampling-Techniken, z.B. Kreuzvalidierung, beinhalten, um eine realistischere Abschätzung der Güte des Klassifikationsbaums zu ermöglichen.

Zusätzlich soll eine bessere grafische Aufbereitung des resultierenden Klassifikationsbaums die Anwendungsfreundlichkeit erhöhen und zur besseren Interpretation der Ergebnisse beitragen.

Literatur

- [1] Breiman L., Friedman J. H., Olshen R. A., Stone C. J. (1984): *Classification And Regression Trees*. Wadsworth Inc., Monterey – California, USA.
- [2] Weber T., Poser S., Zerr I., Felgenhauer K., Otto M., Wiltfang J., Kretzschmar H. A. (1999): *Laborchemische Verfahren in der Differentialdiagnose der Creutzfeldt-Jakob-Krankheit*. Deutsches Ärzteblatt 96, A-3097, B-2494, C-2248.