

IMIB

Prototypen und Prototypikalitätsmaße zur Diagnose von Dysmorphiesyndromen

Rainer Schmidt und Tina Waligora

Institut für Medizinische Informatik und
Biometrie
Universität Rostock

Rainer Schmidt

September 2006

Inhalt

- Was sind Dysmorphiesyndrome?
 - Probleme ihrer Diagnose

- Material

- Das System zur Diagnose
 - Prototypikalitätsmaße
 - Adaptation

- Resultate
 - Erläuterungen

- Diskussion

Was sind Dysmorphiesyndrome?

- „Syndrome morphologischer Mißbildung“
- Nicht zufällige Kombination verschiedener physischer Mißbildungen
- Eine frühzeitige Diagnose erhöht die Erfolgsaussichten einer Lang-Zeit-Therapie

Probleme bei der Diagnose von Dysmorphie-Syndromen

- Über 200 Syndrome sind bisher bekannt, neue werden immer wieder „entdeckt“
- Unser System enthält über 800 Merkmale
- Ein Syndrom wird durch 40 bis 130 Merkmale beschrieben
- Selbst Experten der pädiatrischen Genetik sehen nur wenige Syndrome während ihres Lebens

Material

- Große Fallbasis der pädiatrischen Genetik der Universitätsklinik in München plus seltene Syndrome aus Fachzeitschrift
- Neben Einzelfällen werden über 200 Prototypen verwendet
- Jeder Prototyp beschreibt ein bestimmtes Syndrom durch seine typischen Merkmale

Prototypen

- $\frac{2}{3}$ der Prototypen stammen aus der Fachliteratur oder wurden durch Experten definiert
- Der Rest wurde halb-automatisch generiert
- Prototypen enthalten weniger Merkmale als die Einzelfälle, nämlich nur die (proto-) typischen

Das Diagnosesystem

- Benutzer wählt die Merkmale zur Beschreibung des aktuellen Falls aus
- Benutzer wählt ein Prototypikalitätsmaß aus
- System wendet das Maß sequentiell auf alle Prototypen (Syndrome) an

Prototypikalitätsmaße

- Die Suche nach Einzelfällen (wie sonst beim CBR) ist hier sinnlos!
- Statt dessen: Suche nach Prototypen.
- CBR Retrieval Methoden wie z.B. Nearest Neighbour Suche sind hier unangebracht.
- Statt dessen: Relativ alte sequentielle Prototypikalitätsmaße: Tversky und Mervis / Rosch.

Tversky:

$$P(F,k) = \sum_k f(M(k) + F) - \sum_k f(M(k) - F) - \sum_k f(F - M(k))$$

- Ähnlichkeit zwischen Anfragefall und einem Prototyp
 - Von der Zahl der übereinstimmenden Merkmale ($M(k) + F$) wird
 - die Zahl der Merkmale des Prototyps, die nicht im Anfragefall vorkommen ($M(k) - F$) und
 - die Zahl der Merkmale des Anfragefalls, die nicht im Prototyp vorkommen ($F - M(k)$) abgezogen

Mervis und Rosch:

$$P(F, k) = \frac{\sum_{m(F)} g_m(k) \cdot f(M(k) + F)}{\sum_{m(F)} f(M(k) + F) + \sum_{m(F)} f(M(k) - F)}$$

- Zwei Unterschiede zum Maß von Tversky:
 - Mervis und Rosch berücksichtigen die Merkmale nicht, die im Anfragefall vorkommen aber nicht im Prototyp
 - Die gemeinsamen Merkmale können unterschiedlich gewichtet werden: $g_m(k)$

- Nach Anwendung des Prototypikalitätsmaßes werden alle Prototypen bezüglich ihrer Ähnlichkeit zum Anfragefall sortiert.
- Da Prototypen untereinander sehr ähnlich sein können und da unser Programm der Vorauswahl dient, werden dem Benutzer die ähnlichsten 20 Prototypen präsentiert.

Präsentation der ähnlichsten Prototypen

Most similar prototypes:

Names of the prototypes	Similarities
<input type="checkbox"/> SHPRINTZEN-SYNDROM	0.49
<input type="checkbox"/> LENZ-SYNDROM	0.36
<input type="checkbox"/> BOERJESON-FORSSMAN-LEHMANN-S.	0.34
<input type="checkbox"/> STURGE-WEBER-SYNDROM	0.32
<input type="checkbox"/> LEOPARD-SYNDROM	0.31
<input type="checkbox"/> BRANCHIO-OCULO-FAZIAL-SYNDROM	0.3
<input type="checkbox"/> COCKAYNE-SYNDROM	0.29
<input type="checkbox"/> WILLIAMS-BEUREN-SYNDROM	0.28

Adaption:

- Benutzer kann sich dafür entscheiden, auf die sortierten Prototypen zusätzlich Adaptionsregeln anwenden zu lassen.
- Adaptionsregeln bestehen aus (i.a. zwei) Merkmalskombinationen, die jeweils ein bestimmtes Syndrome bestärken oder zurückweisen
- Die sortierte Liste der Prototypen ist zu modifizieren.

Präsentation der ähnlichsten Prototypen - Nach Anwendung von Adaptionsregeln

Names of the prototypes	Similarities	Applied rule(s)
PROBABLE prototypes after application of the adaptation rules:		
<input type="checkbox"/> LENZ-SYNDROM	0.36	<input type="checkbox"/> REGEL-6
<input type="checkbox"/> DUBOWITZ-SYNDROM	0.24	<input type="checkbox"/> REGEL-9
Prototypes, no adaptation rules could be applied:		
<input type="checkbox"/> SHPRINTZEN-SYNDROM	0.49	
<input type="checkbox"/> BOERJESON-FORSSMAN-LEHMANN-S.	0.34	
<input type="checkbox"/> STURGE-WEBER-SYNDROM	0.32	
<input type="checkbox"/> LEOPARD-SYNDROM	0.31	
<input type="checkbox"/> BRANCHIO-OCULO-FAZIAL-SYNDROM	0.3	
<input type="checkbox"/> COCKAYNE-SYNDROM	0.29	

Ergebnisse

- Vergleich der Prototypikalitätsmaße von Tversky und von Mervis / Rosch bezüglich 100 zufällig ausgewählten Fällen

Richtiges Syndrom	Mervis und Rosch	Tversky
An erster Stelle	29	40
Unter den ersten 3	57	57
Unter den ersten 10	76	69

- Das Maß von Tversky liefert geringfügig bessere Ergebnisse

Ergebnisse

- Zusätzlich: Anwendung einer ersten Menge von Adaptionenregeln

Richtiges Syndrom	Mervis und Rosch	Tversky
An erster Stelle	32	42
Unter den ersten 3	59	59
Unter den ersten 10	77	71

Ergebnisse

- Zusätzlich: Anwendung einer erweiterten Menge von Adaptionsregeln

Richtiges Syndrom	Mervis und Rosch	Tversky
An erster Stelle	36	44
Unter den ersten 3	65	64
Unter den ersten 10	77	73

Erläuterungen

- Die Diagnose ist umso schwieriger, je seltener das Syndrom ist, weil dann
 - fast keine Information in der Literatur zu finden ist und
 - nur sehr wenige solche Fälle in der Fallbasis sind.
- Für bekanntere Syndrome sind deshalb auch die Ergebnisse besser.
- Die 100 Testfälle wurden deshalb nicht rein zufällig ausgewählt, sondern unter der Bedingung, dass jedes Syndrome darin nur einmal vorkommen kann.

Diskussion

- Bisher ist nur eine sehr geringe Verbesserung der Ergebnisse durch Adaptionsregeln zu beobachten.
- Diese Regeln müssen aufwendig mit Hilfe von Experten gewonnen werden.
- Deshalb ist ihre Zahl bisher recht gering.
- Je besser ein Dysmorphiesyndrom bekannt ist, desto leichter können Adaptionsregeln gewonnen werden.

IMIB

Danke für Ihre
Aufmerksamkeit

Fragen ? ? ?