

BTU Cottbus
Seminar "Mustererkennung mit syntaktischen und graphbasierten
Methoden"
Wintersemester 2006/2007

Vortrag zum Thema:

**Erkennung menschlicher Aktionen durch Inferenz
stochastischer regulärer Grammatiken**

Martin Everth
Martin.Everth@tu-cottbus.de
Studiengang: Informatik
Cottbus, den 20.01.07

Inhaltsverzeichnis

I	Motivation und Anwendungen.....	1
I.1	Motivation.....	1
I.2	Anwendungen.....	1
II	Die Hauptfunktionen des Systems zur Erkennung menschlicher Aktionen.....	3
II.1	Der Vorverarbeitungsprozess.....	4
II.2	Die stochastische Grammatik.....	6
II.2.1	Spezifikation der Grammatik.....	6
II.2.2	Erzeugung der Tabellen mit den Wahrscheinlichkeiten.....	7
II.3	Klassifikation und Erkennung der Aktion.....	8
III	Experimente und Ergebnisse.....	9
III.1	Die Daten der Experimente.....	9
III.2	Ergebnisse der Experimente.....	10
IV	Quellen.....	11

I Motivation und Anwendungen

I.1 Motivation

- Für die Entwicklung der Bewegungserkennung gibt es viele Gründe, da es ein breites Anwendungsspektrum gibt
- Hauptziel ist aber die leichtere Mensch-Maschine Interaktion
 - Maschinen sollen Gesten als Eingabe akzeptieren
 - erhöht Anzahl der möglichen Einsatzgebiete von Maschinen
 - z.B.: laute oder schmutzige Umgebungen
- KI-Forschung setzt ebenfalls auf Gesten als Kommunikationsform

I.2 Anwendungen

- Erkennung verdächtiger Bewegungen in Kauf-/ Parkhäusern
 - z.B.: das wiederholte gezielte Schauen in Autos
- Projekt „sicherer Bahnsteig“
 - Erkennen gefährlicher Bewegungen, wie Rennen oder Springen am Bahnsteig

- Moderne Benutzerschnittstelle
- Erkennung und Übersetzung der Gebärdensprache
 - ermöglicht Kommunikation zwischen Hörgeschädigten und Menschen, die der Gebärdensprache nicht mächtig sind
- große Bedeutung bei der Sportvideo-Analyse
 - konkrete Anfrage an ein System: „wie viele Pässe hat Person A gespielt“
 - individuelles Training z.B. beim Golf:
 - System kann ermitteln, ob der Abschlag optimal ausgeführt wurde
- großer Nachteil ist aber die aufwendige 3D-Analyse (auch finanziell)
- bei vielen Anwendungen ist es unmöglich, Markierungen an den Personen anzubringen
⇒ Einschränkung der möglichen Anwendungen

II Die Hauptfunktionen des Systems zur Erkennung menschlicher Aktionen

- bei diesem System werden menschliche Aktionen als eine stochastische vorhersagbare Folge von Teilbewegungen betrachtet
z.B.:

gesamte Bewegung: links-links-links-links-rechts-rechts-rechts-hoch-runter

Teilbewegungen: links, rechts, hoch, runter

- eine Folge von Teilbewegungen kann Störungen und Beobachtungsfehler enthalten, dann wäre diese Folge nur schwer durch eine einfache Grammatik auszudrücken
- stochastische Grammatiken sind hier wesentlich effektiver, weil diese für jede Produktion eine Wahrscheinlichkeit angeben
- es kann stets dieselbe Grammatik verwendet werden
- Bewegungen unterscheiden sich nur durch die Wahrscheinlichkeiten ihrer Produktionen
- so kann das System in der Lernphase eine bestimmte Menge von Aktionen lernen und in der Arbeitsphase erkennen

- Überblick:

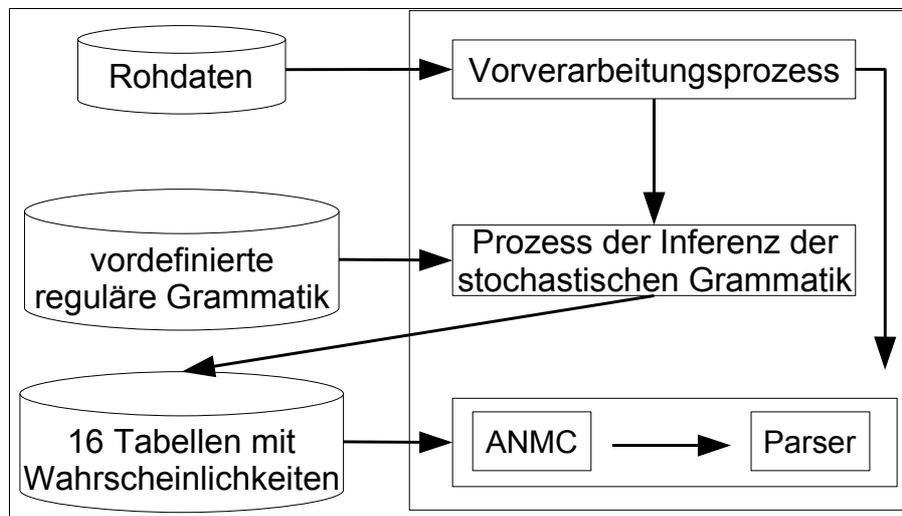


Abbildung 1: das System zur Erkennung von Aktionen

II.1 Der Vorverarbeitungsprozess

Gewinnung der Daten:

- an Person werden acht bunte Markierungen angebracht (an den Schultern, an den Ellenbogen, an den Händen und zwei am Kopf)
 - von den Bewegungen werden Videoaufzeichnungen in 3D erzeugt
 - mit der Software „STABIL++“ werden daraus Folgen von x, y, z – Koordinaten ermittelt
- Quantisierung der gewonnenen Daten, die Intervallbreite beträgt 3cm
⇒ geringfügige Bewegungen werden nicht weiter beachtet

- Abbilden der Daten auf die xy-Ebene und yz-Ebene, es ergeben sich zwei Datensätze
- Konvertierung beider Datensätze in den 8-Nachbarschaftscode

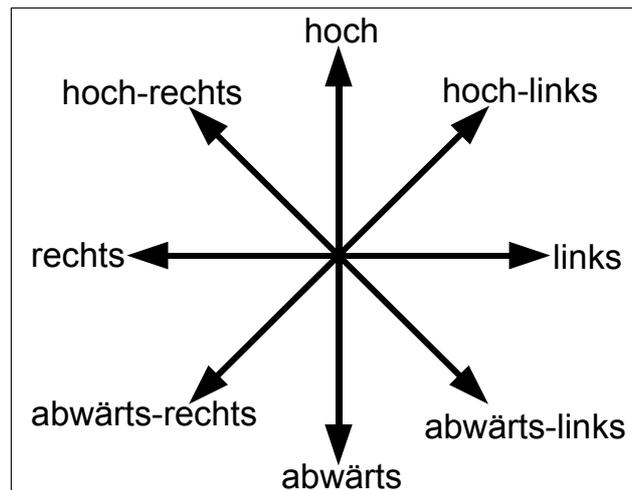


Abbildung 2: xy-Ebene: 8-Nachbarschaftscode
(Anmerkung: aus der Sicht der Versuchsperson)

- Konvertierung der Datensätze in den 4-Nachbarschaftscode, weil sonst:
 - zu viele Produktionen benötigt würden
 - zu viele Übergänge zwischen zwei Zuständen stattfinden würden
 - die Rate der Wiedererkennung sinken würde
- letztendlich kann man dasselbe ausdrücken
z.B.: „links-abwärts“ wird zu „links“ u. „abwärts“
- ein Folge von Daten im 4-Nachbarschaftscode hat folgenden Aufbau: Code Code_Häufigkeit
z.B.: links 12 hoch 11 rechts 4 ...

II.2 Die stochastische Grammatik

II.2.1 Spezifikation der Grammatik

- die Grammatik G_{sk} sei gegeben durch $G_{sk}=(N, \Sigma, P, D_k, S)$ für $k=1,2,\dots,M$ (wobei M die Anzahl der zu erkennenden Aktionen sei)
- N sei eine endliche Menge von Nichtterminaln,
 $N=\{S, R, L, H, A\}$
- Σ sei eine endliche Menge von Terminaln,
 $\Sigma=\{rechts, links, hoch, abwärts\}$
- D_k sei die Menge der Wahrscheinlichkeiten P_{kij} der zugehörigen Produktionen $A_i \rightarrow \beta_j$
- S sei das Startsymbol
- P sei eine endliche Menge von Produktionen / Regeln

$$P = \{ \begin{array}{lllll} S \rightarrow rechts R & R \rightarrow rechts R & L \rightarrow rechts R & H \rightarrow rechts R & A \rightarrow rechts R \\ S \rightarrow links L & R \rightarrow links L & L \rightarrow links L & H \rightarrow links L & A \rightarrow links L \\ S \rightarrow hoch H & R \rightarrow hoch H & L \rightarrow hoch H & H \rightarrow hoch H & A \rightarrow hoch H \\ S \rightarrow abwärts A & R \rightarrow abwärts A & L \rightarrow abwärts A & H \rightarrow abwärts A & A \rightarrow abwärts A \\ & R \rightarrow rechts & L \rightarrow rechts & H \rightarrow rechts & A \rightarrow rechts \\ & R \rightarrow links & L \rightarrow links & H \rightarrow links & A \rightarrow links \\ & R \rightarrow hoch & L \rightarrow hoch & H \rightarrow hoch & A \rightarrow hoch \\ & R \rightarrow abwärts & L \rightarrow abwärts & H \rightarrow abwärts & A \rightarrow abwärts \end{array} \}$$

- die Wahrscheinlichkeiten P_{kij} der Menge D_k werden durch eine Menge X mit Mustern ein und derselben Aktion durch folgende Gleichung bestimmt:

$$P_{kij} = \frac{n_{kij}}{\sum_r n_{kir}}$$

II Die Hauptfunktionen des Systems zur Erkennung menschlicher Aktionen

- $\sum_r n_{kir}$ wird über alle Produktionen in G_{sk} berechnet, die dasselbe A_i haben
- n_{kij} steht für die Häufigkeit, mit der die Produktion $A_i \rightarrow \beta_j$ in G_{sk} benutzt wird, und ergibt sich wie folgt:

$$n_{kij} = \sum_{x_h \text{ in } X} n(x_h) \cdot p(G_{sk}/x_h) \cdot N_{kij}(x_h)$$

- $n(x_h)$ ist die Häufigkeit, mit der alle Muster in X auftauchen
- $N_{kij}(x_h)$ ist die Häufigkeit, mit der $A_i \rightarrow \beta_j$ benutzt wird, wenn ein Muster x_h analysiert wird
- $p(G_{sk}/x_h)$ steht für die Wahrscheinlichkeit, mit der ein Muster x_h von einer Grammatik G_{sk} erzeugt wird
- in der Lernphase wird D_k für $k=1 \dots M$ bestimmt
- diese Lernphase muss für die beiden Ebenen und alle acht Markierungen ausgeführt werden

II.2.2 Erzeugung der Tabellen mit den Wahrscheinlichkeiten

- mit der Grammatik werden für alle acht Markierungen und beide Ebenen 16 Tabellen mit Wahrscheinlichkeiten erzeugt
- in diesen Tabellen werden die Wahrscheinlichkeiten der Produktionen P_{kij} angegeben
- dazu werden die Produktionen nach folgenden Regeln ersetzt:

$A1=S$	$\beta 1=rechts R$	$\beta 5=rechts$
$A2=R$	$\beta 2=links L$	$\beta 6=links$
$A3=L$	$\beta 3=hoch H$	$\beta 7=hoch$
$A4=H$	$\beta 4=abwärts A$	$\beta 8=abwärts$
$A5=A$		

	Produktionen	ersetzte Produktionen	D1	D2	...
1	S → rechts R	A1 → β1	P111	P211	
2	S → links L	A1 → β2	P112	P212	
3	S → hoch H	A1 → β3	P113	P213	
4	S → abwärts A	A1 → β4	P114	P214	
5	R → rechts R	A2 → β1	P121	P221	
...					
35	A → hoch	A5 → β7	P157	P257	
36	A → abwärts	A5 → β8	P158	P258	

Abbildung 3: eine Tabelle mit den Wahrscheinlichkeiten für die xy-Ebene

II.3 Klassifikation und Erkennung der Aktion

- einer Folge von Teilmustern eine gelernte Aktion zuordnen
- aus einem Aktionsmuster entstehen nach dem Vorverarbeitungsprozess 16 Folgen von Teilmustern
- ANMC (Approximate Node Movement Classifier) sortiert die Folgen der Teilmuster nach der Komplexität ihrer Bewegung
- so wird entschieden welche Klassen (gelernte Bewegungen) vergleichbar sind
- die resultierende Wahrscheinlichkeit für eine Aktion wird wie folgt berechnet:

$$P(Knoten_i \cap Knoten_j \cap \dots \cap Knoten_k) = P(Knoten_i) \times P(Knoten_j) \times \dots \times P(Knoten_k) \quad ^1$$

für $i, j, k \in \{1 \dots 16\}$ und $i \neq j \neq k$

¹ Diese Formel gilt wohl nur in erster Näherung. Denn sie geht davon aus, dass die Bewegungen einzelner Knoten statistisch unabhängig sind.

- $P(Knoten_i)$ steht für die Wahrscheinlichkeit eines Knotens
- bestimmt werden die Wahrscheinlichkeiten der Klassen, die vergleichbar sind
- gewählt wird die Klasse mit der höchsten Wahrscheinlichkeit

III Experimente und Ergebnisse

- drei verschiedene Experimente sollen die Effektivität beweisen
- es sollen die Daten, das Experiment an sich und dessen Ergebnisse vorgestellt werden

III.1 Die Daten der Experimente

- bestehen aus 900 Aktionen des menschlichen Oberkörpers, vollführt von drei Personen und aufgezeichnet mit „STABIL++“
- 490 Aktionen werden zum Anlernen des Systems verwendet und 410 Aktionen zum Testen des Systems
- es soll zwischen 60 Aktionen unterschieden werden
- diese Aktionen können sein: Bewegungen des Kopfes,

des Oberkörpers, der Arme und Bewegungen mehrerer Körperteile kombiniert

III.2 Ergebnisse der Experimente

Experiment	Beschreibung des Experiments	gelernte Daten	Rate der Erkennung
1	8 Aktionen des Kopfes ohne ANMC	40	93.8% (30/32)
2	60 Aktionen der Hände und des Oberkörpers ohne ANMC	490	64.6% (265/410)
3	60 Aktionen der Hände und des Oberkörpers mit ANMC	490	84.9% (348/410)

Abbildung 4: Ergebnisse der Experimente

- bei Experiment 1 ist ANMC unnötig, weil alle Bewegungen des Kopfes die selbe Komplexität haben
- Die Rate der Wiedererkennung wird bestimmt durch:

$$R = \frac{\text{Anzahl der korrekt erkannten Aktionen}}{\text{Anzahl aller Aktionen}}$$

- das System erreicht gute Werte bei der Erkennung von Aktionen mit geringer Komplexität (siehe Experiment 1)
- Experiment 2 und 3 haben eine höhere Komplexität und dafür ist die Rate der Erkennung sehr hoch
- ANMC führt zu einer deutlichen Steigerung der Rate (siehe Experiment 3)

IV Quellen

- Kyungeun Cho / Hyungje Cho / Kyhyun Um:
Human action recognition by inference of stochastic regular grammars. LNCS 3138 (2004), S. 388-396.
- Frank Pagel: Analyse menschlicher Bewegungen:
Eine Einführung (Teil 1). Vortrag:
(http://iaks-www.ira.uka.de/home/haimerl/seminar/SeminarBewWS01_02/pagel.ps.gz vom 20.01.07)