

Wissensrepräsentation

Vorlesung

Sommersemester 2008

8. Sitzung

Dozent

Nino Simunic M.A.

Computerlinguistik, Campus DU

- (Fortsetzung LC-/Chart-Parsing)
- **Statistische Verfahren in der KI**
 - Impliziert Maschinelles Lernen

Bisherige Annahmen (insb. Bei den bisherigen Regelbasierten Systemen)

- **Clear-cut world:** Jede Hypothese ist entweder falsch, wahr, oder unbekannt.
 - ***Closedworld assumption:*** Jede Hypothese, die unbekannt ist, wird als falsch angenommen.
- Binäres System: 0/1, wahr/falsch, ja/nein
- In vielen Fällen okay, in ebenso vielen Fällen jedoch suboptimal.

Statistische vs. **symbolische** Verfahren

■ Symbolische Verfahren

- Basis: Regeln, Logik, wahr/falsch
 - »Was sind gültige bzw. ungültige Strukturen?«
- Wissen (Regeln) werden intellektuell (Experten) entwickelt
- Beispiel: Syntaxanalyse (Grammatik)

Statistische vs. symbolische Verfahren

- Basis: Statistische Modelle
 - Mathematik: Stochastik, Zufall
- »Intellektuelles« Wissen (im Idealfall) auf ein Minimum reduziert
 - Mensch/Experte stellt stattdessen Daten/Beispiele als Trainingsmaterial zur Verfügung
- Paradigma jetzt: Was sind wahrscheinliche / unwahrscheinliche Strukturen?
- Beispiel: Wortvervollständigung
 - Eclipse, SMS schreiben (T9 Wörterbuch), ...
 - »Gegeben ein Folge von ambigen Eingaben und Zeichen: Was ist die wahrscheinlichste Zeichenkette?«

T9 Wörterbuch (*Text on 9 keys*)

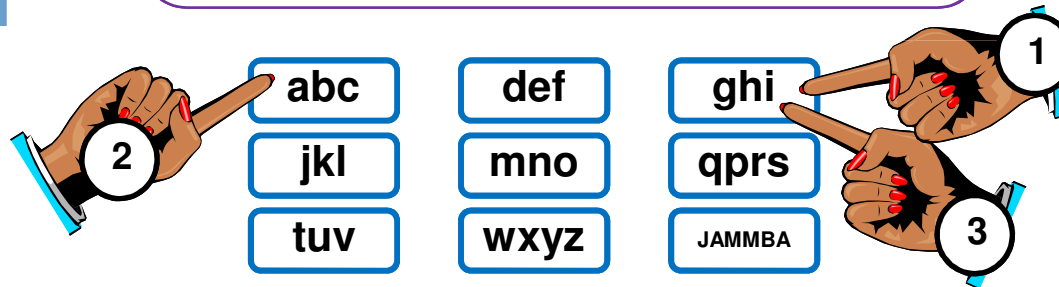


SMS schreiben an: Petra

Hallo Petra,

gestern war [?][?][?]

① ② ③
[i] [ha] [ich]



Wie kommt es, dass aus **ghi** → **abc** geschlossen wird, dass **ha**?

System **vermutet**, dass ha am **wahrscheinlichsten** ist, **gegeben** Kontext (ghi) und aktuelle Eingabe (abc).

Wie kommt es, dass aus **ghi** → **abc** → **ghi** **ich** geschlossen wird, und als nächste Vorschläge **hai**, gag, hah, hag, häh angeboten/vermutet werden?

System schlägt sortiert nach Wahrscheinlichkeit vor: **ich am wahrscheinlichsten**, **hai am zweitwahrscheinlichsten**, ..., gegeben Kontext und aktuelle Eingabe.

Notwendige Statistiken für

ghi

abc

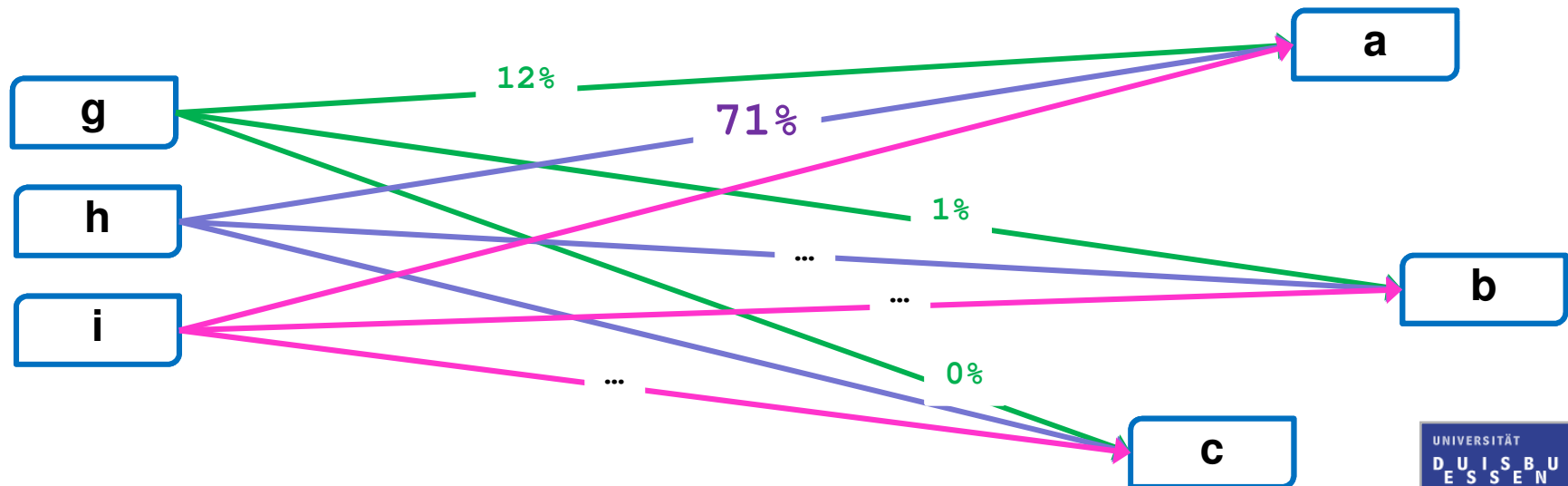
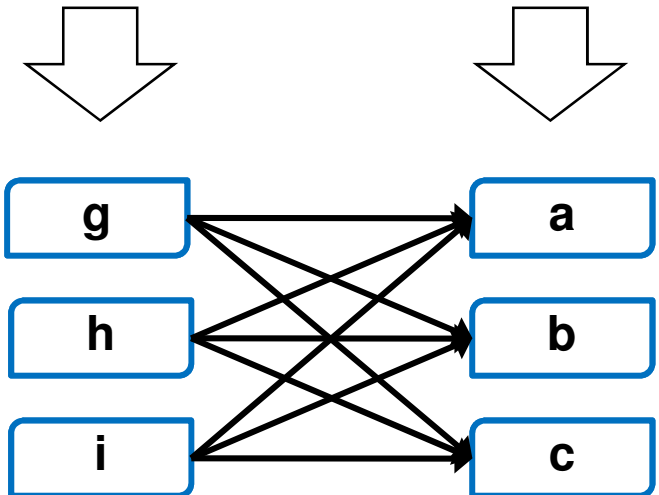
Tasten(-eingabe) zunächst nicht eindeutig:

Software muss also alle Möglichkeiten berücksichtigen.

Wie wahrscheinlich ist es,

- dass a auf g folgt,
- dass b auf g folgt,
- dass c auf g folgt,
- dass a auf h folgt, dass b auf h folgt,

ghi → abc



Woher kommen die Wahrscheinlichkeiten?

Denkbar: Von Experten ermittelt

Fundierte Wissen über die Sprache (T9 in 90 Sprachen) und den allgemeinen (~wahrscheinlichsten) »Sprachgebrauch«

Immenser Wortschatz

Zeit

Geld

»Es ist wahrscheinlicher, dass aus {ghi}{abc}{ghi} *ich* entsteht, als (Platz 2) hai, (3) gag, ...«

»Einem *e* folgt öfter *r* als ein *f*. ... Einem *o* folgt häufiger *n* als ein */* ...«

**Statistische Verfahren,
Maschinelles Lernen**

viel zu zeit- und kostenintensiv, fehleranfällig, nicht repräsentativ für den allgemeinen Sprachgebrauch, ...

Training, Algorithmus, Lernen, n-Gramme

■ Trainingsdaten

- Text, der dem allgemeinen Sprachgebrauch entspricht, z.B. Zeitungstexte

■ Lern-Algorithmus

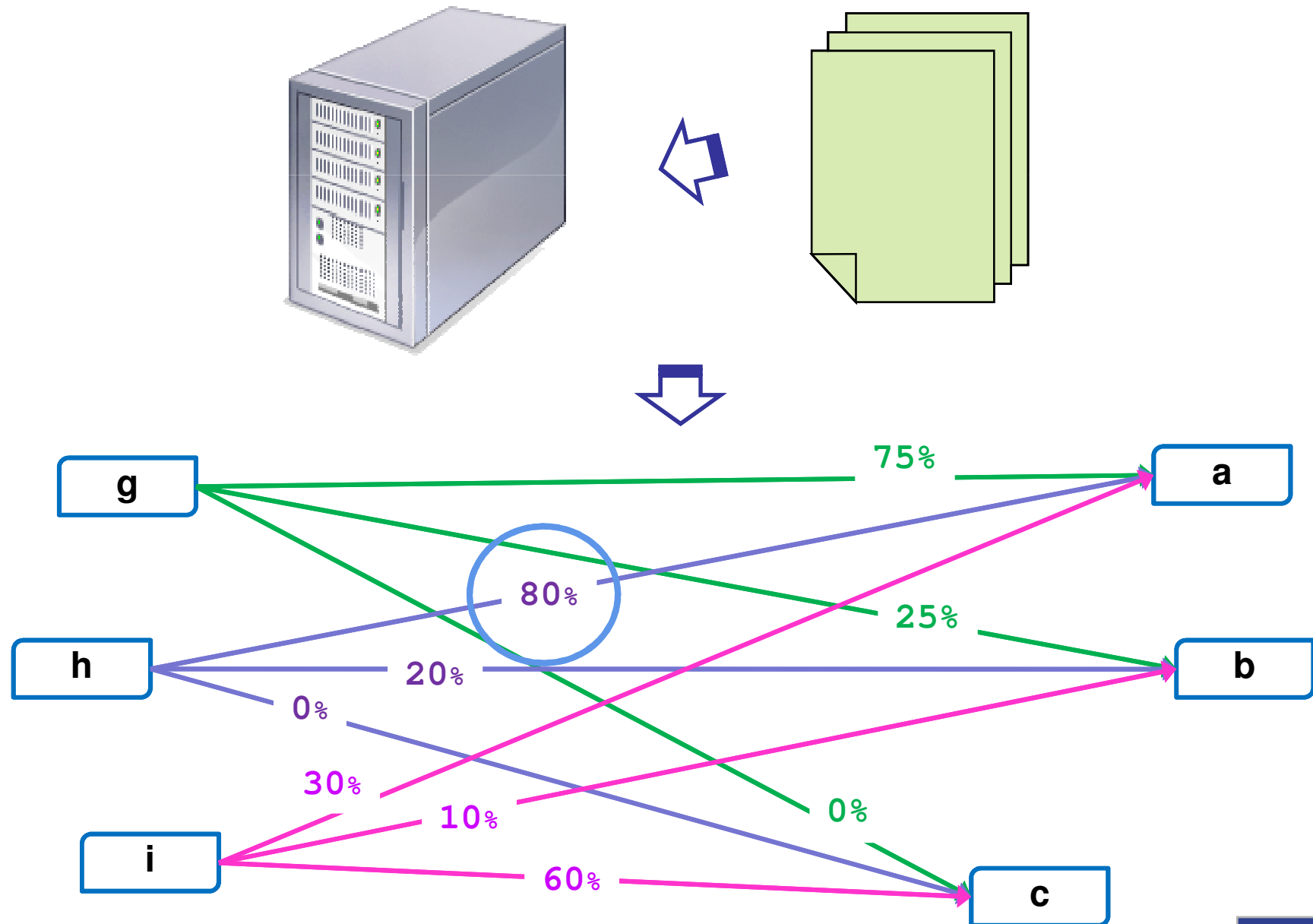
- Erstelle Statistiken zu Zeichensequenzen des Texts (der Wörter)

→ Maschine lernt die Wahrscheinlichkeiten

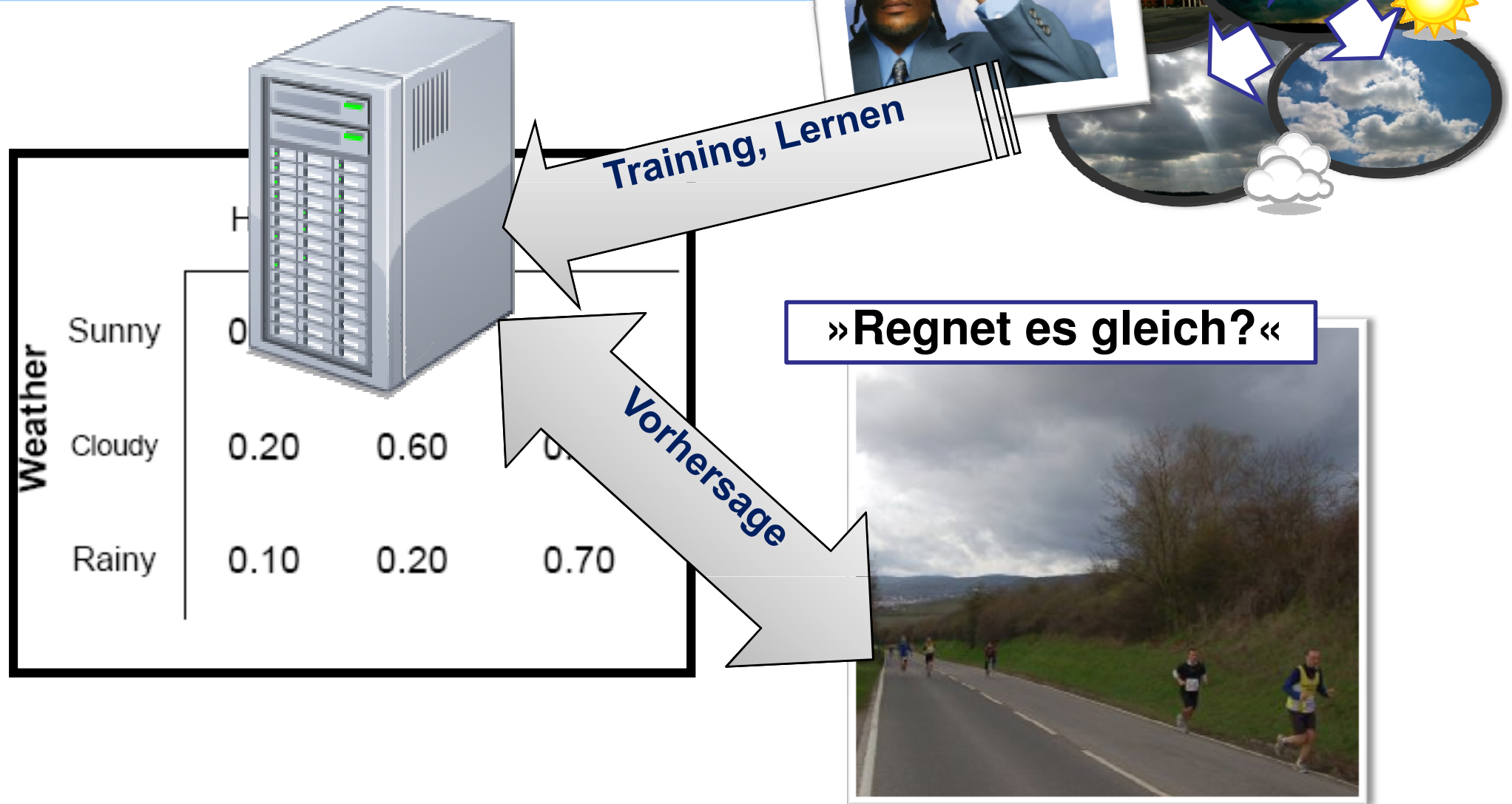
- Minimaler intellektueller Aufwand
- Wahl des richtigen Korpus (repräsentativ, balanciert)
- Vorarbeit: Wörter in Zeichen-**n-Gramme** zerlegen:

Unigramm (n=1):	{_B,a,n,a,n,e_}
Bigramm (n=2):	{_Ba, an, na, an, ne_}
Trigramm (n=3):	{_Ban, ana, nan, ane_}
Tetragramm (n=4):	{_Bana, anan, nane_}
Pentagramm (n=5):	{_Banan, anane_}

Maschine lernt aus Beobachtungen: Statistiken aus adäquaten, intellektuell ausgesuchten Daten



Aus Beobachtungen lernen



Aus Beobachtungen lernen: Wettervorhersage über Zeitraum



$t \backslash t+1$	Regen	Sonne	Wolken
Regen	1	2	0
Sonne	0	1	2
Wolken	0	1	1

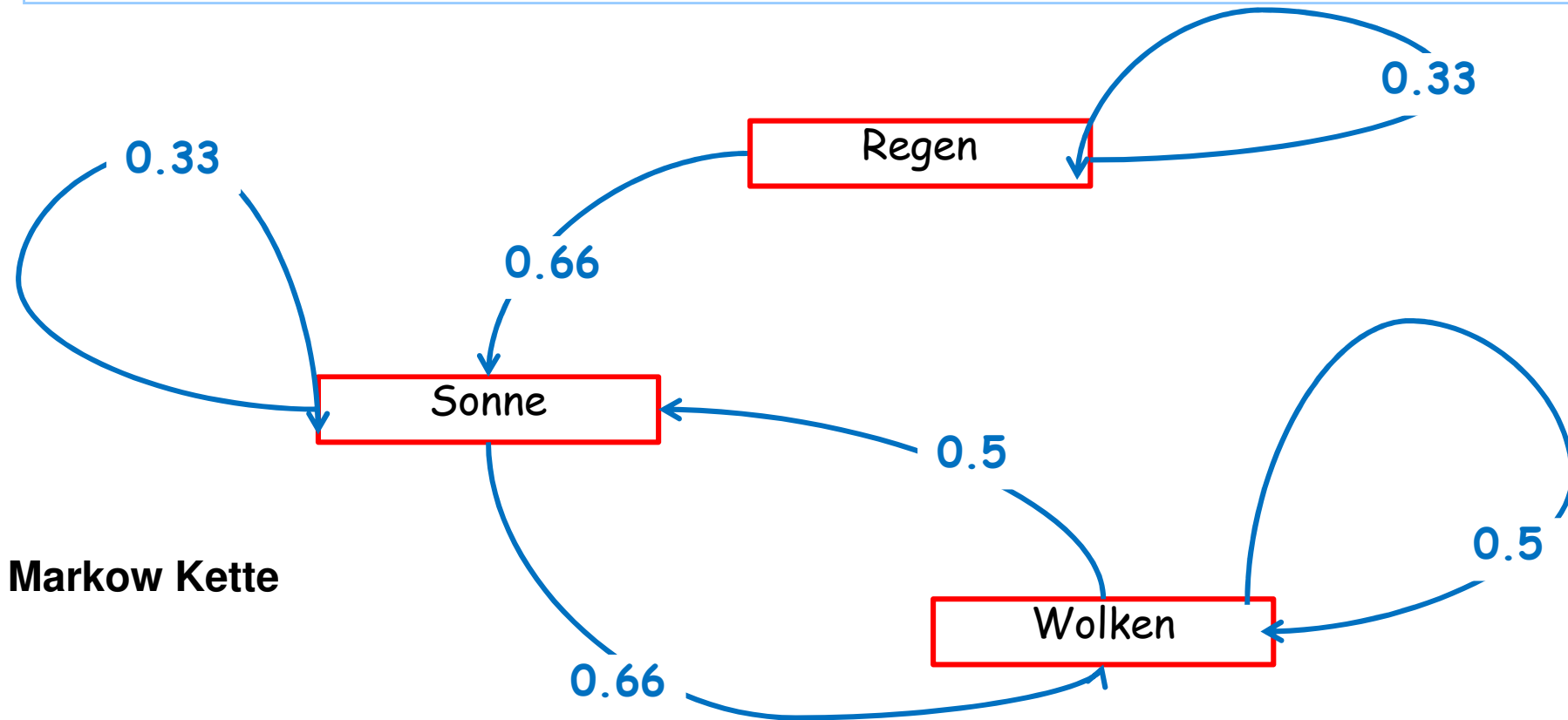
Zeilen müssen in der Summe jeweils 1 ergeben. (Hier nur wg. Rundung nicht möglich.)

0.33	0.66	0
0	0,33	0,66
0	0.5	0,5

Berechnung der Wahrscheinlichkeit:
Häufigkeit / Anzahl Beobachtungen

Unbeobachtete Wetter-Wechsel: Häufigkeit und Wahrscheinlichkeit = 0;

Endlicher (nicht-deterministischer) Automat stellt Zustände mit Übergängen und ihren Wahrscheinlichkeiten dar



t \ t+1	Regen	Sonne	Wolken
Regen	0.33	0.66	0
Sonne	0	0,33	0,66
Wolken	0	0.5	0,5

Zustände

Zustandsübergänge

Wie wird das Wetter **morgen**, wenn es **heute** geregnet hat?

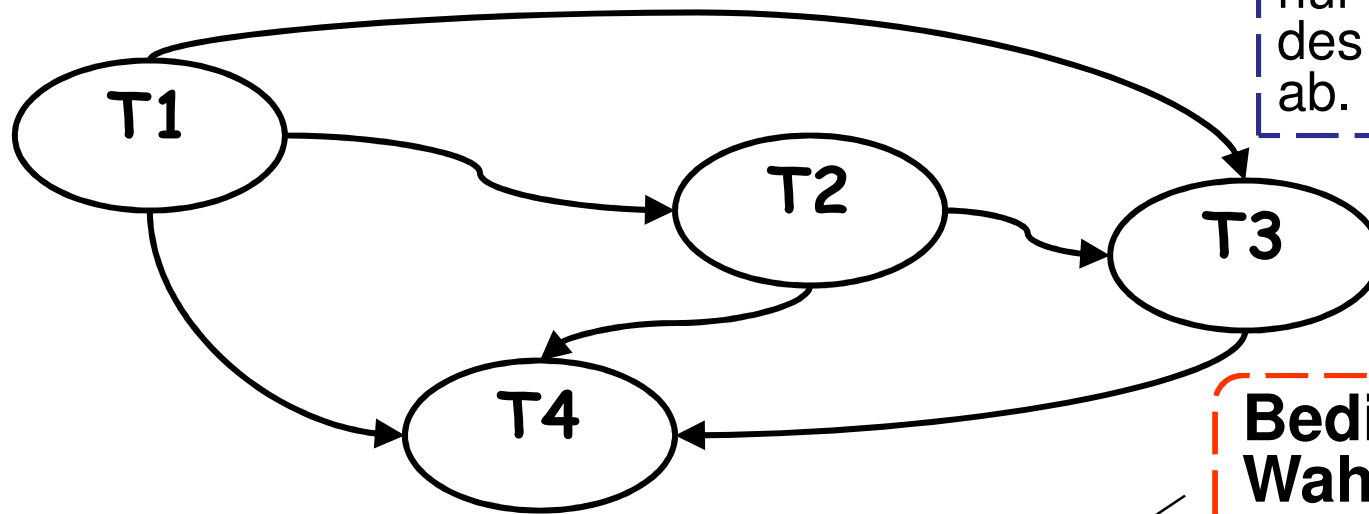
t \ t+1			
	Regen	Sonne	Wolken
Regen	0.33	0.66	0
Sonne	0	0,33	0,66
Wolken	0	0.5	0,5

Antwort: **Sonnig**

» Wird es wohl in **den nächsten zwei Tagen** sonnig oder verregnet? Heute ist es verregnet.«

$$P(\cancel{T_2} \cancel{T_3} T_4 \mid T_1) = P(\cancel{T_2} \mid T_1) * P(\cancel{T_3} \mid \cancel{T_2}) * P(\cancel{T_4} \mid \cancel{T_3})$$

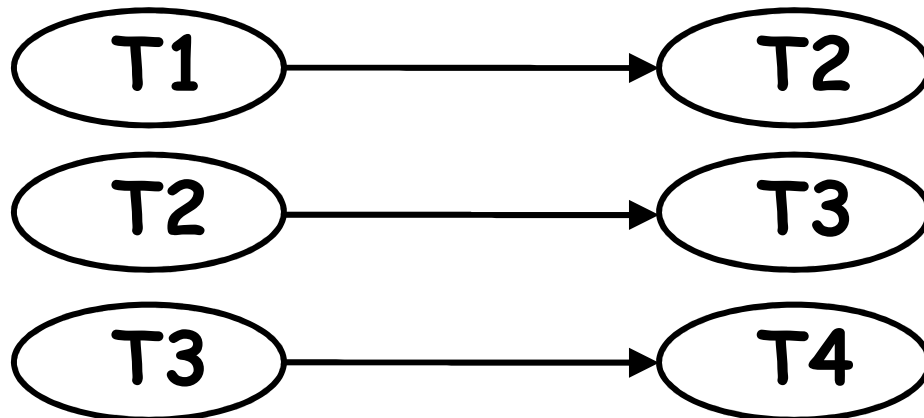
$$P(T2 \ T3 \ T4 \mid T1)$$



Eigenschaft einer Markow-Kette:
Wahrscheinlichkeit des Zustands zum Zeitpunkt t hängt nur von der Wahrscheinlichkeit des Zustands zum Zeitpunkt $t-1$ ab.

Bedingte Wahrscheinlichkeiten.

$$P(T2 \ T3 \ T4 \mid T1) =$$



$$\begin{aligned}
 &P(T2 \mid T1) \\
 &* P(T3 \mid T2) \\
 &* P(T4 \mid T3)
 \end{aligned}$$

Berechnung von der Wahrscheinlichkeit von Sequenzen, z.B. **Regen Regen | Wolken** und **Sonne Sonne | Wolken**

»Wird es wohl **in den nächsten zwei Tagen sonnig oder verregnet?**
Heute ist es **bewölkt.**« Vergleich von Pfaden nötig.

t \ t+1	Regen	Sonne	Wolken
Regen	0.33	0.66	0
Sonne	0	0.33	0.66
Wolken	0	0.5	0.5

Problematische Aussage:
 Beruht auf fehlenden
 Beobachtungen, und nicht,
 weil es z.B. Regen → Regen de
 facto nicht gibt. (Schön wär's).

% sonnig

$$P(\text{Sonne Sonne} \mid \text{Wolken}) = P(\text{Sonne} \mid \text{Wolken}) * P(\text{Sonne} \mid \text{Sonne})$$

$$0.5 * 0.33 = 0,165$$

% verregnet

$$P(\text{Regen Regen} \mid \text{Wolken}) = P(\text{Regen} \mid \text{Wolken}) * P(\text{Regen} \mid \text{Regen})$$

$$0 * 0.33 = 0$$

Sparse Data Problem: Abdeckung der Daten

- Nicht beobachtet wurden die Sequenzen:
 - Sonne → Regen
 - $P(\text{Regen}|\text{Sonne}) = 0$
 - Regen → Wolken (Wolken|Regen)
 - $P(\text{Wolken}|\text{Regen}) = 0$
 - Wolken → Regen wurde nicht beobachtet.
 - $P(\text{Regen}|\text{Wolken}) = 0$
- Also: Beobachten, bis man alle Zustandsübergänge beobachtet hat.
 - Suboptimal: Keine Garantie, dass die fehlenden Beobachtungen tatsächlich beobachtet werden.
 - Insbesondere in anderen Kontexten gravierend: Textsammlung mit allen Wörter der deutschen Sprache?

Problem-Beispiel

(»Lösungsbeispiel« hierzu auf S. 21)

Gesucht ist die wahrscheinlichere Sequenz von
 $P(\text{Sonne Regen} | \text{Wolken})$ und $P(\text{Regen Sonne} | \text{Wolken})$

t \ t+1	Regen	Sonne	Wolken
Regen	0.33	0.66	0
Sonne	0	0.33	0.66
Wolken	0	0.5	0.5

Keine Möglichkeit zu differenzieren, was wahrscheinlicher ist.

$$\begin{aligned} P(\text{Sonne Regen} | \text{Wolken}) &= \\ P(\text{Regen} | \text{Wolken}) * P(\text{Sonne} | \text{Regen}) &= \\ 0 * 0.33 &= 0 \end{aligned}$$

$$\begin{aligned} P(\text{Regen Sonne} | \text{Wolken}) &= \\ P(\text{Sonne} | \text{Wolken}) * P(\text{Regen} | \text{Sonne}) &= \\ 0.5 * 0 &= 0 \end{aligned}$$

Smoothing Verfahren (Glättungsverfahren)

- Allgemeines Problem im Kontext **supervised learning**: Sparse data.
 - Überwachte Lernverfahren
 - Lernen anhand von (intellektuell zusammengestellten / bearbeiteten **Beispielen**)
- Lösung: **Techniken**, die ungesehenen Beobachtungen eine Wahrscheinlichkeit > 0 zuweisen
→ ***Smoothing***

Smoothing der Wetterdaten: Add-one Smoothing Verfahren

$t \backslash t+1$	Regen	Sonne	Wolken
Regen	1  +1	2  +1	0  +1
Sonne	0  ...	1	2
Wolken	0	1	1

Zur absoluten Häufigkeit der Beobachtungen wird jeweils der Wert 1 addiert und die Wahrscheinlichkeiten dann berechnet.



$t \backslash t+1$	Regen	Sonne	Wolken
Regen	2	3	1
Sonne	1	2	3
Wolken	1	2	2

$t \backslash t+1$	Regen	Sonne	Wolken
Regen	0,33	0,5	0,16
Sonne	0,16	0,33	0,5
Wolken	0,2	0,4	0,4

Vorhersage des letzten (Problem-) Beispiels nun mit geglätteten Werten

t \ t+1	Regen	Sonne	Wolken
Regen	0,33	0,5	0,16
Sonne	0,16	0,33	0,5
Wolken	0,2	0,4	0,4

Gesucht ist die
wahrscheinlichere Sequenz
von $P(\text{Sonne Regen} | \text{Wolken})$
und $P(\text{Regen Sonne} | \text{Wolken})$

$$\begin{aligned}
 &P(\text{Sonne Regen} | \text{Wolken}) \\
 &P(\text{Regen} | \text{Wolken}) * P(\text{Sonne} | \text{Regen}) \\
 &\quad 0.2 \quad * \quad 0.5 = 0,1
 \end{aligned}$$



$$\begin{aligned}
 &P(\text{Regen Sonne} | \text{Wolken}) \\
 &P(\text{Sonne} | \text{Wolken}) * P(\text{Regen} | \text{Sonne}) \\
 &\quad 0.4 \quad * \quad 0.16 = 0,064
 \end{aligned}$$