

GWV – Grundlagen der Wissensverarbeitung

Tutorial 9 : Belief Networks

*Enrico Milutzki (6671784), Love Kumar (7195374),
Nikolai Poets (6911432), Jan Willruth (6768273)*

Exercise 9.2 : (Language Modelling)



Code in markov.py, die einzulesende Datei muss sich im selben Ordner befinden.
Wir haben nur die ersten 100.000 Wörter eingelesen, da die Berechnung ansonsten deutlich länger dauern würde; funktioniert natürlich trotzdem mit "etwas" mehr Geduld.

Hier einmal beispielhaft generierte Sätze der Länge 20 mit dem Startwort "hallo".

```
Bigram:  
hallo dies ist wenn er zu mal wieder alles in der titel der kann ich jetzt ist ja schon die  
  
Trigram:  
hallo ach komm auch die inszenierung gruendlich armklagen als die engine neue zelda bekannt war es wird verkrieche ich glaube  
  
Quadrogram:  
hallo sagen dir mal spontan ein kind total aus jux zu zwecken einer fehlinterpretation bzw. fehlapplikation auf meinem next-gen-sabbern tm
```

Im Vergleich zu realen Sätzen lässt sich erkennen, dass jeweils 2 aufeinander folgende Wörter natürlich Sinn machen müssen, gesetzt den Fall, dass der eingelesene Text ein realer Text ist, da ein folgendes Wort ja auf dem Vorgänger basiert. Auch Artikel sind aus diesem Grund richtig verwendet. Allerdings sind Wortfolgen von 3, 4 oder mehr Wörtern dadurch selten. Darüber hinaus fällt im Vergleich zwischen den mit Bi-, Tri- und Quadrogrammen generierten Sätzen auf, dass die Wörter (von Bi-, zu Tri-, zu Quadrogrammen) durchschnittlich länger und die Sätze sinnvoller werden, da spezifische Wortkombinationen wahrscheinlicher werden.

Exercise 9.3 : (Diagnosis (cont.))



$$P(\text{Battery}) = 1,0 - 0,1 = 0,9 = 90 \%$$

$$P(\text{Starter}) = P(\text{Battery}) * P(\text{Ignition Key}) * P(\text{Starter}) = 0,9^3 = 0,729 = 72,9 \%$$

$$P(\text{Engine}) = P(\text{Battery}) * P(\text{Ignition Key}) * P(\text{Starter}) * P(\text{Electronic Fuel Regulation}) *$$

$$P(\text{Fuel Tank}) * P(\text{Fuel Pump}) * P(\text{Filter}) * P(\text{Engine}) = 0,9^8 = 0,4305 = 43,05 \%$$

$$P(\text{Engine} | \text{Pump is working}) = P(\text{Starter}) * P(\text{Filter}) * P(\text{Engine}) = 0,9^3 = 0,729 = 72,9 \%$$

Exercise 9.4 : (Bayesian Probabilities)



1.

$f_0(\text{Smuggler})$:

Smuggler	Value
True	0,01
False	0,99

$f_1(\text{Fever})$:

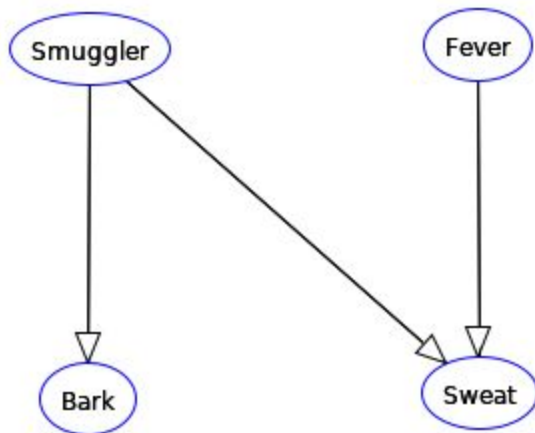
Fever	Value
True	0,013
False	0,987

$f_2(\text{Smuggler}, \text{Bark})$:

Smuggler	Bark	Value
True	True	0,8
True	False	0,2
False	True	0,05
False	False	0,95

f3(Smuggler, Fever, Sweat):

Smuggler	Fever	Sweat	Value
True	True	True	0,8
True	True	False	0,2
True	False	True	0,4
True	False	False	0,6
False	True	True	0,6
False	True	False	0,4
False	False	True	0
False	False	False	1



Probability network in the .zip

2.

Mögliche Ursachen für Schwitzen sind entweder, dass die Person ein Schmuggler ist, Fieber hat oder beides. Wenn aber nun beobachtet wird, dass der Hund bellt, so wird Fieber als Ursache für das Schwitzen “away explained”, da der Hund ja nicht bellt, nur weil eine Person Fieber hat.



3.

P(Smuggler | Dog is barking):

Eliminiere Faktoren f_1 (Fever) und f_3 (Smuggler, Fever, Sweat).

$(f_2 | \text{Dog is barking}) = f_4(\text{Smuggler})$: True = 0,8, False = 0,05

Multipliziere Faktoren:

$f_0 * f_4 = f_5(\text{Smuggler})$: True = $0,01 * 0,8 = 0,008$, False = $0,99 * 0,05 = 0,0495$

Normalisieren:

$f_5 / (0,008 + 0,0495) = f_6(\text{Smuggler})$: True = $0,008 / 0,0575 = 0,13913$,

False = $0,0495 / 0,0575 = 0,86087$

Die Wahrscheinlichkeit, dass eine Person ein Schmuggler ist unter der Bedingung, dass der Hund bellt ist gerundet 13,91 %.

P(Sweating):

Eliminiere Faktor f_2 (Smuggler, Bark).

Summiere Variablen nacheinander aus:

Eliminiere Smuggler von f_0 (Smuggler) und f_3 (Smuggler, Fever, Sweat):

$f_4(\text{fever, sweat})$:

True, True = $0,01 * 0,8 + 0,99 * 0,6 = 0,602$,

True, False = $0,01 * 0,2 + 0,99 * 0,4 = 0,398$,

False, True = $0,01 * 0,4 + 0,99 * 0 = 0,004$,

False, False = $0,01 * 0,6 + 0,99 * 1 = 0,996$

Eliminiere Fever von f_1 (Fever) und f_4 :

$f_5(\text{sweat})$: True = $0,013 * 0,602 + 0,987 * 0,004 = 0,01177$,

False = $0,013 * 0,398 + 0,987 * 0,996 = 0,98823$

Die Wahrscheinlichkeit, dass eine Person schwitzt ohne vorherige Beobachtungen ist gerundet 1,18 %.

P(Smuggler | Sweating and dog is barking):

Eliminiere Faktoren $f_2(\text{Smuggler}, \text{Bark})$ und $f_3(\text{Smuggler}, \text{Fever}, \text{Sweat})$. Es entstehen...

$f_4(\text{Smuggler})$: True = 0,8, False = 0,05

$f_5(\text{Smuggler}, \text{Fever})$: TT = 0,8, TF = 0,4, FT = 0,6, FF = 0

Summiere Fever aus:

$f_6(\text{Smuggler})$: True = $0,013 * 0,8 + 0,987 * 0,4 = 0,4052$,

False = $0,013 * 0,6 + 0,987 * 0 = 0,0078$

Multipliziere Faktoren:

$f_0 * f_4 * f_6 = f_7(\text{Smuggler})$: True = $0,01 * 0,8 * 0,4052 = 0,0032416$,

False = $0,99 * 0,05 * 0,0078 = 0,0003861$

Normalisieren:

$F_7 / (0,0032416 + 0,0003861) = f_8(\text{Smuggler})$:

True = $0,0042416 / 0,0036277 = 0,89357$,

False = $0,0003861 / 0,0036277 = 0,10643$

Die Wahrscheinlichkeit, dass eine Person ein Schmuggler ist unter den Beobachtungen, dass die Person schwitzt und der Hund bellt ist gerundet 89,36 %.