Przetwarzanie języka naturalnego/05

2023-03-29

Krzysztof Misztal misztal.edu.pl

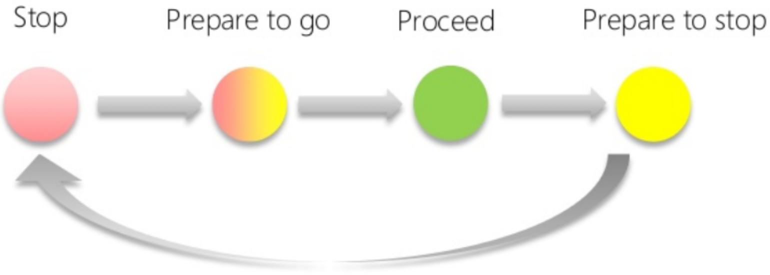
Spis treści

Spis treści

Spis treści

Ukryty Model Markowa

Rozważmy sygnalizację świetlną na skrzyżowaniu



Każdy stan jest zależny od poprzedniego. System jest deterministyczny. Możemy rozważać go jako proces Markowa.

Process Markowa

https://www.youtube.com/watch?v=EqUfuT3CC8s

Process Markowa

- Stan zależy wyłącznie od stanu poprzedniego
- Stan jest niezależny od czasu

Process Markowa

- **Proces Markowa** – ciąg zdarzeń, w którym prawdopodobieństwo każdego zdarzenia zależy jedynie od wyniku poprzedniego. W ujęciu matematycznym, procesy Markowa to takie procesy stochastyczne, które spełniają własność Markowa.
- **Łańcuchy Markowa** to procesy Markowa z czasem dyskretnym.
- Łańcuch Markowa jest ciągiem $X_1, X_2, X_3, \dots$ zmiennych losowych. Dziedzinę tych zmiennych nazywamy przestrzenią stanów, a realizacje X_n to stany w czasie n. Jeśli rozkład warunkowy X_{n+1} jest funkcją wyłącznie zmiennej X_n:

$$P(X_{n+1} \leqslant y | X_0, X_1, X_2, \dots, X_n) = P(X_{n+1} \leqslant y | X_n)$$

  to mówimy, że proces stochastyczny **posiada własność Markowa**.

Spis treści

Ukryty Model Markowa

Rozważmy system przewidywania pogody na postawie obserwacji wodorostów.

Ukryty Model Markowa

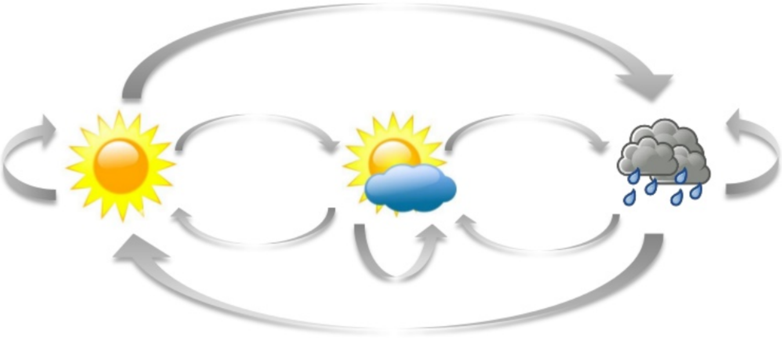
Rozważmy system przewidywania pogody na postawie obserwacji wodorostów.

- Folklore tells us that "soggy" seaweed means wet weather, while "dry" seaweed means sun.
- It the seaweed is in an ntermediate state "damp", then we cannot be sure.

Słownictwo:
soggy - damp - dryish -dry
rozmoczony - wilgotny - suchawy - suchy

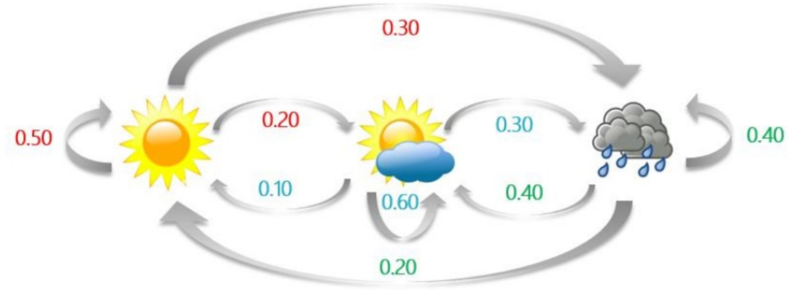
Ukryty Model Markowa

Rozważmy model pogodowy

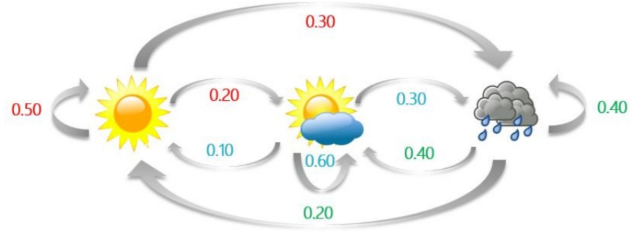


Ukryty Model Markowa

Rozważmy model pogodowy



Ukryty Model Markowa

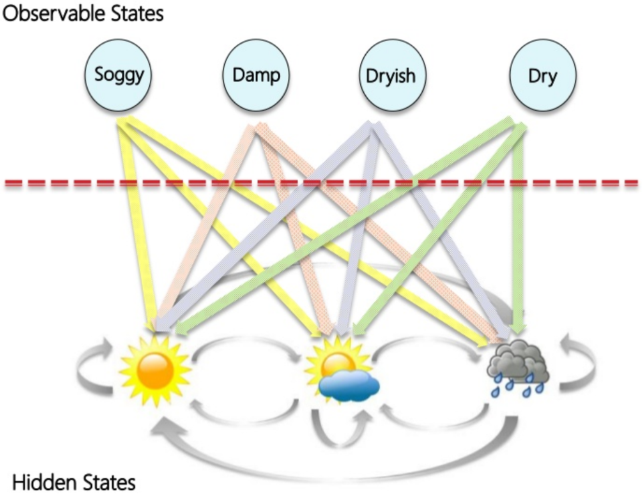


		Weather today		
		Sunny	Cloudy	Rainy
	Sunny	0,5	0,2	0,3
Weather yesterday	Cloudy	0,1	0,6	0,3
	Rainy	0,2	0,4	0,4

Macierz przejść (transition matrix/state transition matrix)

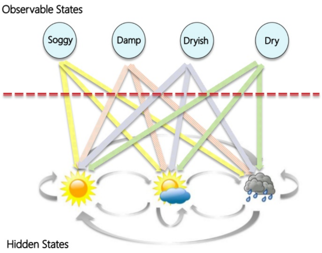
Ukryty Model Markowa

Wprowadzenie procesu Markowa



Ukryty Model Markowa

Wprowadzenie procesu Markowa



		Seaweed State Today			
		Dry	Dryish	Damp	Soggy
	Sunny	0,6	0,2	0,15	0,05
Weather yesterday	Cloudy	0,25	0,25	0,25	0,25
	Rainy	0,05	0,1	0,35	0,5

Macierz emisji (emission matrix/confusion matrix)

Ukryty Model Markowa

- Rozkład początkowy $\pi = [0.63, .017, 0.20]$
- Macierz przejść (transition matrix/state transition matrix)

$$T = \begin{array}{cccc} & \textit{Sunny} & \textit{Cloudy} & \textit{Rainy} \\ \textit{Sunny} & 0,5 & 0,2 & 0,3 \\ \textit{Cloudy} & 0,1 & 0,6 & 0,3 \\ \textit{Rainy} & 0,2 & 0,4 & 0,4 \end{array}$$

- Macierz emisji (emission matrix/confusion matrix)

$$E = \begin{array}{ccccc} & \textit{Dry} & \textit{Dryish} & \textit{Damp} & \textit{Soggy} \\ \textit{Sunny} & 0,6 & 0,2 & 0,15 & 0,05 \\ \textit{Cloudy} & 0,25 & 0,25 & 0,25 & 0,25 \\ \textit{Rainy} & 0,05 & 0,1 & 0,35 & 0,5 \end{array}$$

Spis treści

Ukryty Model Markowa

Ukryty Model Markowa

Trójke (π, T, E) nazywamy **Ukrytym Modelem Markowa** (Hidden Markov Model, HMM), gdy π jest wektorem rozkładu prawdopodobieństwa rozmiaru n, T jest macierzą $n \times n$, której każdy rząd jest wektorem rozkładu prawdopodobieństwa a E jest macierzą $n \times m$ o tej samej własności.

Spis treści

Podstawowe zadania HMM

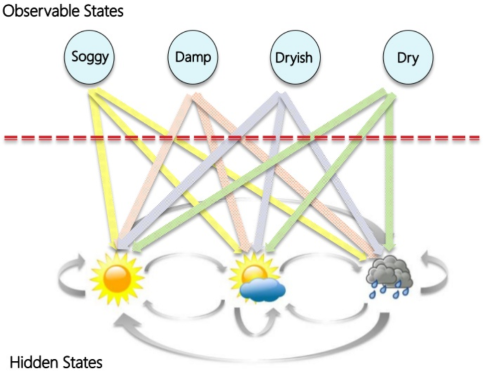
1. Dla zadanego HMM i zadanego ciągu obserwacji $(O_1, \ldots, O_k)$ oceń jak jest to prawdopodobne zdarzenie
2. Dla zadanego HMM i zadanego ciągu obserwacji $(O_1, \ldots, O_k)$ znajdź najbardziej prawdopodobny ciąg stanów ukrytych $(H_1, \ldots, H_k)$
3. Dla zbioru obserwacji $\{(O_1, \ldots, O_{k_i})\}_{i=1}^{N}$ $=1$ znajdź najbardziej prawdopodobny HMM o określonej strukturze (n, m)
4. Dla zadanego zbioru HMM i zadanego ciagu obserwacji $(O_1, \ldots, O_k)$ wskaz najbardziej prawdopodobny HMM

Spis treści

Podstawowe zadania HMM

- Dla zadanego HMM i zadanego ciągu obserwacji $(O_1, \dots, O_k)$ oceń **jak jest to prawdopodobne zdarzenie**

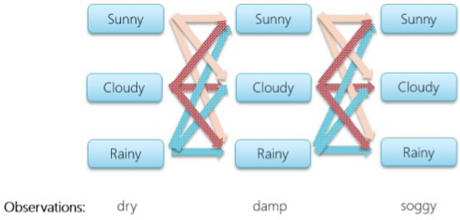
Podstawowe zadania HMM



$$O = (O_1, O_2, O_3) = (dry, damp, soggy)$$

$$P(O|HMM) = ?$$

Podstawowe zadania HMM



$$O = (O_1, O_2, O_3) = (dry, damp, soggy)$$

$$P(O|HMM) = ?$$

Podstawowe zadania HMM

$$O = (O_1, O_2, O_3) = (dry, damp, soggy)$$

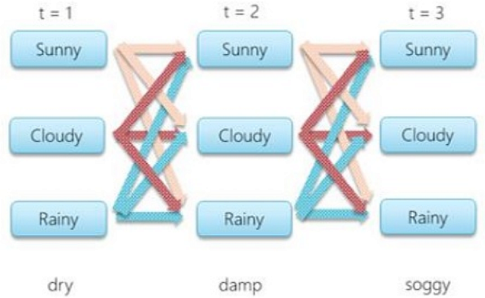
$$\begin{aligned} P(O|HMM) &= P(O|(sunny, sunny, sunny)) + \\ &\quad +P(O|(sunny, sunny, cloudy)) + \\ &\quad + \ldots + \\ &\quad +P(O|(rain, rain, rain)) \\ &= \sum_{(H_1,H_2,H_3)\in\{sunny,cloudy,rain\}^3} P(O|(H_1, H_2, H_3)) \end{aligned}$$

Spis treści

Podejscie naiwne

- Zwraca poprawny wynik
- Wymaga policzenia $n^{|O|}$ prawdopodobieństw, gdzie n to liczba stanów ukrytych a O to ciąg obserwacji, dla naszego przypadku $3^3 = 27$
- Nie korzysta z własności Markowa

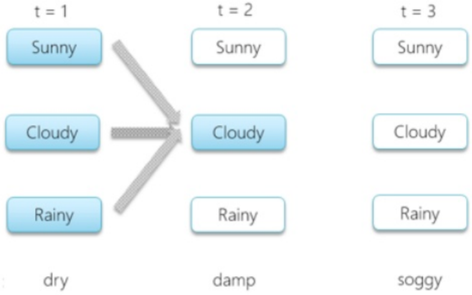
Forward algorithm



- Rozważmy obliczanie prawdopodobieństw wystąpienia sekwencji stanów zadanych przez HMM

Spis treści

Forward algorithm



$$\alpha_t(x) = P(O_t|H_t = x)P(\text{all paths to } x \text{ before } t)$$

Forward algorithm

$$\alpha_1(x) = P(O_1|H_1 = x)\pi_x = E_{x,O_1}\pi_x$$

$$\alpha_{t+1}(x) = P(O_{t+1}|H_{t+1} = x)\sum_{i=1}^{n}\alpha_t(i)T_{i,x} = E_{x,O_{t+1}}\sum_{i=1}^{n}\alpha_t(i)T_{i,x}$$

$$P(O|HMM) = \sum_{i=1}^{n}\alpha_{|O|}(i)$$

Forward algorithm

1. Policz $\alpha_1(x)$ dla każdego stanu ukrytego x, dla pierwszej obserwacji O_1
2. Dla każdej obserwacji O_t $(t > 1)$ i każdego stanu ukrytego x policz $\alpha_t(x)$ używając $\alpha_{t-1}(y)$
3. Zwróć sumę $\alpha_{|O|}(x)$ po wszystkich stanach ukrytych x

Forward algorithm

- Zwraca poprawny wynik
- Ma złożoność $|O|n$ (zamiast $n^{|O|}$) (przy ewaluacji od lewej do prawej)
- Korzysta z własności Markowa

Spis treści

Podstawowe zadania HMM

- Dla zadanego zbioru HMM i zadanego ciagu obserwacji $(O_1, \ldots, O_k)$ wskaz najbardziej prawdopodobny HMM

Przykład - Pogoda

Załóżmy, ze mamy wymodelowane poszczególne pory roku jako Ukryte Modele Markowa.
Mając dany ciąg obserwacji naszej rośliny z kilku dni - w jaki sposób odpowiedzieć na pytanie **"Jaka mamy porę roku?"**

Przykład - NLP

Załóżmy, ze mamy wymodelowane wypowiedzi różnych osób jako Ukryte Modele Markowa.
Mając dany ciąg słów (zdanie) - w jaki sposób odpowiedzieć na pytanie **"Kto jest autorem tych słów?"**

Problem

- Dane:

$$HMMMs = \{HMM_1, \ldots, HMM_l\}$$

$$O$$

- Szukamy:

$$\arg\max_{HMM \in HMMs} P(HMM)P(O|HMM)$$

bo

$$P(HMM|O) = P(O|HMM)P(HMM)\frac{1}{P(O)}$$

HMM a NB

W tym konkretnym zastosowaniu - bardzo podobne do Naive Bayesa, tylko poziom abstrakcji wyżej

$$\mathrm{argmax}_{HMM \in HMMs} P(HMM) P(O|HMM)$$

wygląda analogicznie jak

$$\underset{c \in C}{\mathrm{argmax}}\, P(c) P(d|c)$$

Spis treści

Podstawowe zadania HMM

- Dla zadanego HMM i zadanego ciągu obserwacji $(O_1, \ldots, O_k)$ znajdź najbardziej prawdopodobny ciąg stanów ukrytych $(H_1, \ldots, H_k)$

Spis treści

Tagging

Tagowanie tekstu

Mając dany tekst w formie ciągu słów $d = word_1, \ldots, word_k$ oraz ciągu tagów $t_d = tag_1, \ldots, tag_k$ celem tagowania jest zbudowanie funkcji $f : words \rightarrow tags$, która dobrze odwzorowuje to tagowanie, tj. minimalizuje jakąś funkcje błędu $E(f(d), t_d)$.

Przykłady

- Wykrywanie czy kropka jest końcem zdania czy użyto jej w innej formie
- Określanie czy dane słowo jest nazwa własna
- **Określanie części mowy**

Określanie części mowy - Part of speech

- **Noun**: a part of speech inflected for case, signifying a concrete or abstract entity
- **Verb**: a part of speech without case inflection, but inflected for tense, person and number, signifying an activity or process performed or undergone
- **Participle**: a part of speech sharing the features of the verb and the noun
- **Interjection**: a part of speech expressing emotion alone
- **Pronoun**: a part of speech substitutable for a noun and marked for a person
- **Preposition**: a part of speech placed before other words in composition and in syntax
- **Adverb**: a part of speech without inflection, in modification of or in addition to a verb, adjective, clause, sentence, or other adverb

Spis treści

Naiwne podejście

Dla danego słowa w

1. Jeśli w nie ma w zbiorze uczącym to zwróć tag, który występuje najczęściej w tym zbiorze
2. W przeciwnym razie zwróć tag, który występuje w zbiorze uczącym najczęściej dla słowa w

Naiwne podejście

Jak skuteczne jest tego typu podejście?

Naiwne podejście

Jak skuteczne jest tego typu podejście?

ok. 90% dla języka angielskiego (dla podstawowej wersji V/Adj/N/Det)

Spis treści

POS jako HMM

Założenia:

- Tag (cześć mowy) słowa w_i zależy wyłącznie od tagu (części mowy) poprzedniego słowa w_{i-1}
- Tag jest niezależny od czasu
- Słowo jest "generowane" przez swój tag (cześć mowy), czyli jest zależne tylko od niego

Od strony matematycznej:

- Założenie Markowa
- Stacjonarność procesu

POS jako HMM

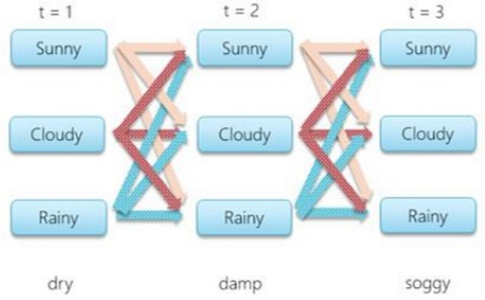
- Rozwiązujemy problem POS używając HMM:
  - Stany ukryte to możliwe części mowy
  - Stany widzialne (obserwacje) to słowa z języka

Spis treści

Podstawowe zadania HMM – wracamy

- Dla zadanego HMM i zadanego ciągu obserwacji $(O_1, \ldots, O_k)$ znajdź najbardziej prawdopodobny ciąg stanów ukrytych $(H_1, \ldots, H_k)$

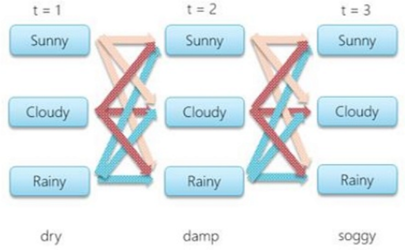
Naiwne podejście



$$O = (dry, damp, soggy)$$

$$\arg \max_{(H_1, H_2, H_2) \in H} P(O | H_1, H_2, H_3)$$

Naiwne podejście



$$O = (dry, damp, soggy)$$

$$\arg \max_{(H_1, H_2, H_2) \in H} P(O|H_1, H_2, H_3)$$

$$= \text{argmax}\{P(O|(sunny, sunny, sunny)), P(O|(sunny, sunny, cloudy)),$$

$$\ldots, P(O|(cloudy, cloudy, clody))\}$$

Naiwne podejście

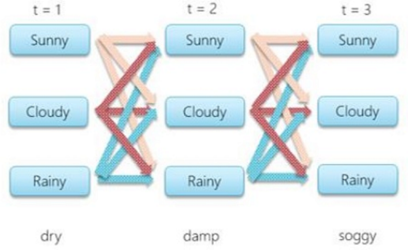
- Bardzo podobna sytuacja do poprzedniej
- Złożoność wykładnicza ze względu na długość obserwacji
- ... ale zwraca dobry wynik

Spis treści

Viterbi

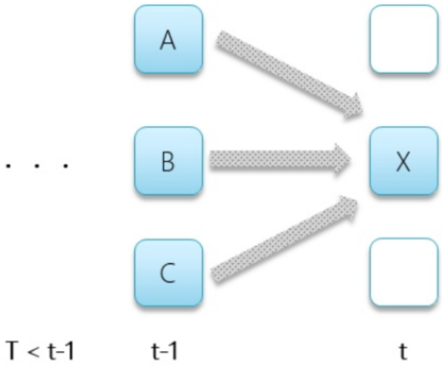
`https://www.youtube.com/watch?v=6JVqutwtzmo&t=1s`

Viterbi - podejście trellis



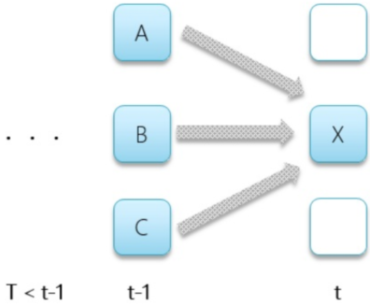
Viterbi - najbardziej prawdopodobne ścieżki

Viterbi - najbardziej prawdopodobne ścieżki



$$P_{best,t}(X) = ?$$

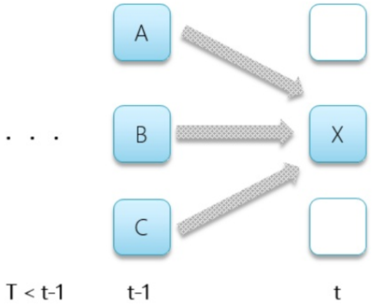
Viterbi - najbardziej prawdopodobne ścieżki



Najlepsza ścieżka do x musi mieć postać:

- ... A X
- ... B X
- ... C X

Viterbi - najbardziej prawdopodobne ścieżki



Najlepsza ścieżka do x musi mieć postać:

- ... A X $\quad P_{best,t}(X) = P_{best,t-1}(A)P(A \rightarrow X)P(O_t|X)$
- ... B X $\quad P_{best,t}(X) = P_{best,t-1}(B)P(B \rightarrow X)P(O_t|X)$
- ... C X $\quad P_{best,t}(X) = P_{best,t-1}(C)P(C \rightarrow X)P(O_t|X)$

$$P_{best,t}(X) = \max_i P_{best,t-1}(i)P(i \rightarrow X)P(O_t|X)$$

Viterbi - najbardziej prawdopodobne ścieżki

- $P_{best,t}(X) = \pi_X P(O_1|X)$
- $P_{best,t}(X) = \max_i P_{best,t-1}(i)P(i \to X)P(O_t|X)$
- $\delta_t(X) = \max_i \delta_{t-1}(i) T_{i,X} E_{X,O_t}$

Viterbi - najbardziej prawdopodobne ścieżki

Mamy prawdopodobieństwo ścieżki, trzeba ja tylko odzyskać

Viterbi - najbardziej prawdopodobne ścieżki

- $$\delta_t(X) = \max_i \delta_{t-1}(i) T_{i,X} E_{X,O_t}$$

- $$\phi_t(x) = \operatorname*{argmax}_i \delta_{t-1}(j) P(i \to x) = \operatorname*{argmax}_i \delta_{t-1}(j) T_{i,x}$$

Viterbi - najbardziej prawdopodobne ścieżki

- Mając dane O, HMM
- Obliczamy δ, ϕ
- Odpowiadamy $H = (H_{a_1}, \ldots, H_{a_{|O|}})$, gdzie :
  - $a_{|O|+1} := \operatorname{argmax} \delta_{|O|}(i)$
  - $a_t = \phi_{t+1}(a_{t+1})$

Spis treści

Części zdania

- **NN** – singular or mass noun
- **VB** – verb, base form
- **TO** – infinitive marker to
- **PPSS** – other nominative personal pronoun (I, we, they, you)

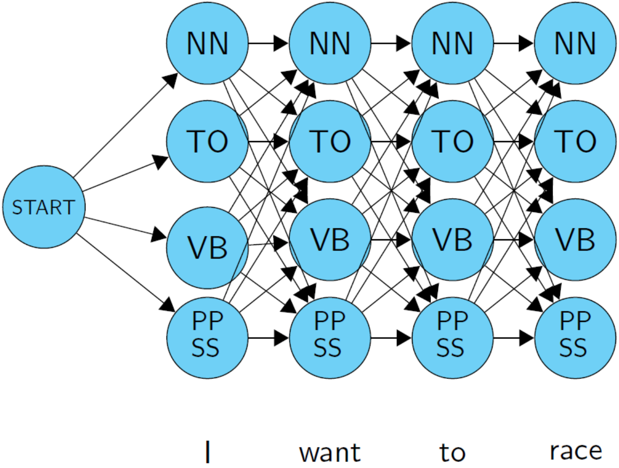
Części zdania

- **NN** – singular or mass noun
- **VB** – verb, base form
- **TO** – infinitive marker to
- **PPSS** – other nominative personal pronoun (I, we, they, you)

Jakie jest tagowanie dla zdania?

I want to race

Viterbi


START
NN
TO
VB
PP SS
NN
TO
VB
PP SS
NN
TO
VB
PP SS
NN
TO
VB
PP SS
I
want
to
race

Viterbi – przykład

- Macierz przejść (transition matrix/state transition matrix)

	VB	TO	NN	PPPS
$< s >$	.019	.0043	.041	.67
VB	.0038	.035	.047	.0070
TO	.83	0	.00047	0
NN	.0040	.016	.087	.0045
PPPS	.23	.00079	.001	.00014

- Macierz emisji (emission matrix/confusion matrix)

	I	want	to	race
VB	0	.0093	0	.00012
TO	0	0	.99	0
NN	0	.000054	0	.00057
PPSS	.37	0	0	0

Viterbi

q_4	NN	0				
q_3	TO	0				
q_2	VB	0				
q_1	PPSS	0				
q_0	start	**1.0**				
		$\langle s \rangle$	I	want	to	race
			w_1	w_2	w_3	w_4

- Utwórz macierz prawdopodobieństw z kolumną dla każdej obserwacji (np. word token), i jednym wierszem dla każdego stanu (np. POS tag)
- Wypełniamy komórki kolumna po kolumnie
- Wpisz w i-tą kolumnę oraz j-ty wiersz prawdopodobieństwo najbardziej prawdopodobnej ścieżki do stanu q_j, który wyemitował $w_1, \ldots, w_i$

Viterbi

q_4	NN	0	$1.0 \times .041 \times 0$			
q_3	TO	0	$1.0 \times .0043 \times 0$			
q_2	VB	0	$1.0 \times .19 \times 0$			
q_1	PPSS	0	$1.0 \times .67 \times .37$			
q_0	start	**1.0**				
		$\langle s \rangle$	I	want	to	race
			w_1	w_2	w_3	w_4

- Dla każdego stanu q_j w czasie i, wyznacz

$$v_i(j) = \max_{k=1}^{n} v_{i-1}(k) \cdot a_{kj} \cdot b_j(w_i)$$

- $v_{i-1}(k)$ to poprzednie prawdopodobieństwo ścieżki, a_{kj} – prawdopodobieństwo przejścia, $b_j(w_i)$ – prawdopodobieństwo emisji

Viterbi

q_4	NN	0	0	$.025 \times .0012 \times .000054$		
q_3	TO	0	0	$.025 \times .00079 \times 0$		
q_2	VB	0	0	$.025 \times .23 \times .0093$		
q_1	PPSS	0	**.025**	$.025 \times .00014 \times 0$		
q_0	start	**1.0**				
		$\langle s \rangle$	I	want	to	race
			w_1	w_2	w_3	w_4

- Dla każdego stanu q_j w czasie i, wyznacz

$$v_i(j) = \max_{k=1}^{n} v_{i-1}(k) \cdot a_{kj} \cdot b_j(w_i)$$

- $v_{i-1}(k)$ to poprzednie prawdopodobieństwo ścieżki, a_{kj} – prawdopodobieństwo przejścia, $b_j(w_i)$ – prawdopodobieństwo emisji

Viterbi

q_4	NN	0	0	.000000002	$.000053 \times .047 \times 0$	
q_3	TO	0	0	0	$.000053 \times .035 \times .99$	
q_2	VB	0	0	**.00053**	$.000053 \times .0038 \times 0$	
q_1	PPSS	0	**.025**	0	$.000053 \times .0070 \times 0$	
q_0	start	**1.0**				
		$\langle s \rangle$	I	want	to	race
			w_1	w_2	w_3	w_4

- Dla każdego stanu q_j w czasie i, wyznacz

$$v_i(j) = \max_{k=1}^{n} v_{i-1}(k) \cdot a_{kj} \cdot b_j(w_i)$$

- $v_{i-1}(k)$ to poprzednie prawdopodobieństwo ścieżki, a_{kj} – prawdopodobieństwo przejścia, $b_j(w_i)$ – prawdopodobieństwo emisji

Viterbi

q_4	NN	0	0	.000000002	0	$.0000018 \times .00047 \times .00057$
q_3	TO	0	0	0	**.0000018**	$.0000018 \times 0 \times 0$
q_2	VB	0	0	**.00053**	0	$.0000018 \times .83 \times .00012$
q_1	PPSS	0	**.025**	0	0	$.0000018 \times 0 \times 0$
q_0	start	**1.0**				
		$\langle s \rangle$	I	want	to	race
			w_1	w_2	w_3	w_4

- Dla każdego stanu q_j w czasie i, wyznacz

$$v_i(j) = \max_{k=1}^{n} v_{i-1}(k) \cdot a_{kj} \cdot b_j(w_i)$$

- $v_{i-1}(k)$ to poprzednie prawdopodobieństwo ścieżki, a_{kj} – prawdopodobieństwo przejścia, $b_j(w_i)$ – prawdopodobieństwo emisji

Viterbi

q_4	NN	0	0	.000000002	0	4.8222e-13
q_3	TO	0	0	0	**.0000018**	0
q_2	VB	0	0	**.00053**	0	**1.7928e-10**
q_1	PPSS	0	**.025**	0	0	0
q_0	start	**1.0**				
		$\langle s \rangle$	I	want	to	race
			w_1	w_2	w_3	w_4

- Dla każdego stanu q_j w czasie i, wyznacz

$$v_i(j) = \max_{k=1}^{n} v_{i-1}(k) \cdot a_{kj} \cdot b_j(w_i)$$

- $v_{i-1}(k)$ to poprzednie prawdopodobieństwo ścieżki, a_{kj} – prawdopodobieństwo przejścia, $b_j(w_i)$ – prawdopodobieństwo emisji

Viterbi – przykład 2

Jakie jest tagowanie dla zdania?

deal talks fail

- Macierz przejść (transition matrix/state transition matrix)

	to N	to V
from start	.8	.2
from N	.4	.6
from V	.8	.2

- Macierz emisji (emission matrix/confusion matrix)

	deal	fail	talks
N	.2	.05	.2
V	.3	.3	.3

Viterbi

	deal	talks	fail
N	$.8 \times .2 = .16$	$\leftarrow .16 \times .4 \times .2 = .0128$ (bo $.16 \times .4 > .06 \times .8$)	$\swarrow .0288 \times .8 \times .05 = .001152$ (bo $.0128 \times .4 < .0288 \times .8$)
V	$.2 \times .3 = .06$	$\nwarrow .16 \times .6 \times .3 = .0288$ (bo $.16 \times .6 > .06 \times .2$)	$\nwarrow .0128 \times .6 \times .3 = .002304$ (bo $.0128 \times .6 > .0288 \times .2$)

Dziękuję za uwagę.