Piotr Szczuko

UKRYTE MODELE MARKOWA

HMM – HIDDEN MARKOV MODELS

Ukryte modele Markowa

- Ukryte modele Markowa są statystyczną metodą klasyfikacji **sekwencji zdarzeń**.
- Do celów rozpoznawania mowy po raz pierwszy zostały użyte przez Bakera, Jelinka i Levinsona (firma IBM) (1975)
- Sygnał – ciąg „zdarzeń” które są **parametryzowane**
  - Segmenty 10-30ms, opisywane wektorami parametrów
  - Zakłada się, że w segmencie sygnał jest stacjonarny
  - Słowo (klasa) to ciąg **O**bserwacji/zdarzeń/segmentów

$$O=o_1, o_2, \dots o_T$$

Gdzie T=liczba segmentów, na które podzielono słowo

Podsumowanie (1)

- **Klasa** = pojedyncze słowo
- **Obiekt** = ciąg obserwacji, sposób wypowiedzenia słowa
- **Klasyfikacja** = porównanie ciągu obserwacji do zapamiętanego wzorca

Automat skończony

- Każde słowo wzorcowe (klasa) jest osobnym automatem skończonym o N-stanach:

$$Q = \{q_1, q_2, \dots, q_N\}$$

- Model markowa definiujemy jako $\lambda = <\pi, A, B>$, gdzie:

$$\pi = [\pi_1, \pi_2, \dots, \pi_N]$$

rozkład prawdopodobieństw znalezienia się w stanie q_i w chwili $t = 0$,

$$A = [a_{ij}] \ (i,j = 1..\ N)$$

Macierz prawdopodobieństw przejść między stanami,

$B = [b_i(o_j)]$ (i = 1.. N, j = 1.. M ; M – liczba możliwych zdarzeń generowanych przez dany stan)

macierz prawdopodobieństw pojawienia się j-tej obserwacji w stanie q_i:

$$b_i(o_j) = P(o_j | q_i)$$

Przykład

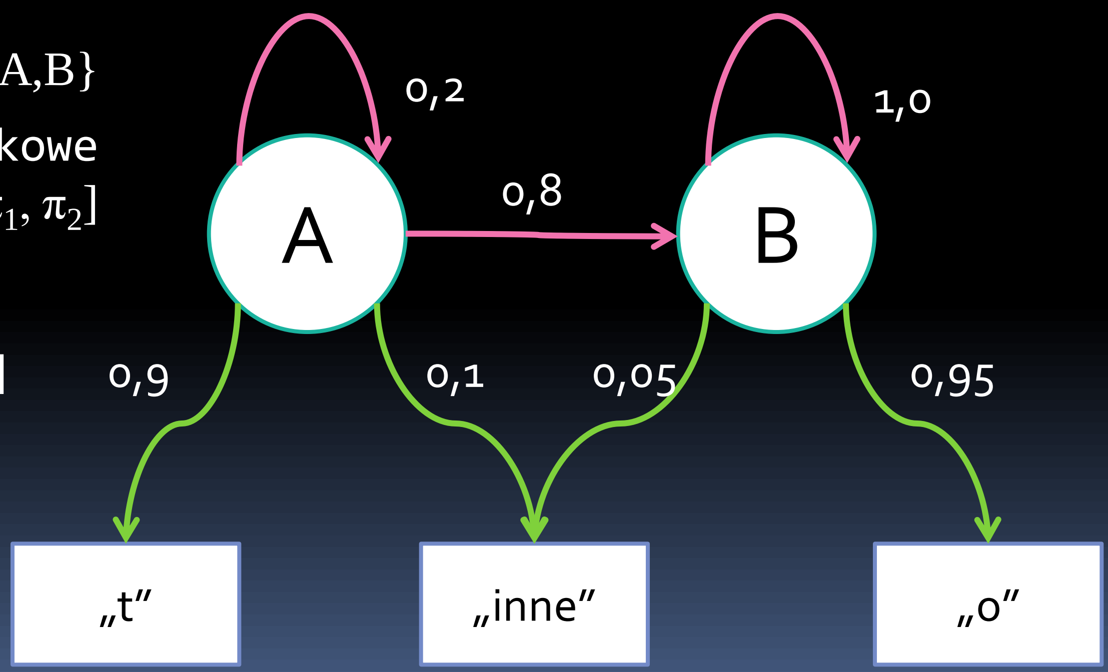
- Model słowa „to”

Przejścia $A = [a_{ij}]$

Stany $Q = \{q_1, q_2\} = \{A, B\}$

P-stwa początkowe $\pi = [\pi_1, \pi_2]$

$B = [b_i(o_j)]$



Obserwacje:

Budowa klasyfikatora

- Przygotowanie danych:
  - Rejestracja sygnałów, segmentacja, parametryzacja,
  - Słownik - transkrypcja słów na obserwacje (np. „tszy”, „ośem”)
  - Gramatyka – jakie słowa występują po sobie (np. „skasuj+to”)
- Trenowanie ukrytych modeli Markowa na podstawie każdej z reprezentacji danego słowa
  - Np. wszystkich powtórzeń słowa „to”
  - Określenie złożoności modelu (liczby stanów modelu),
  - Prowadzi do wyznaczenia $\lambda = <\pi, A, B>$ dla każdego słowa
- Opracowanie procedury klasyfikacji (o tym dalej)

Trening

- Algorytmy:
  - Viterbiego,
  - Baum-Welcha,
  - „embedded training”.
- Problemy:
  - Segmentacja sygnału mowy
    - Stałe odcinki czasowe?
    - Segmentacja podyktowana maksymalizacją p-stwa generowania danej obserwacji
  - Konieczna kwantyzacja obserwacji (wyznaczanie p-stw występowania dowolnej obserwacji w dowolnym stanie jest praktycznie nierealizowalne).

Klasyfikacja

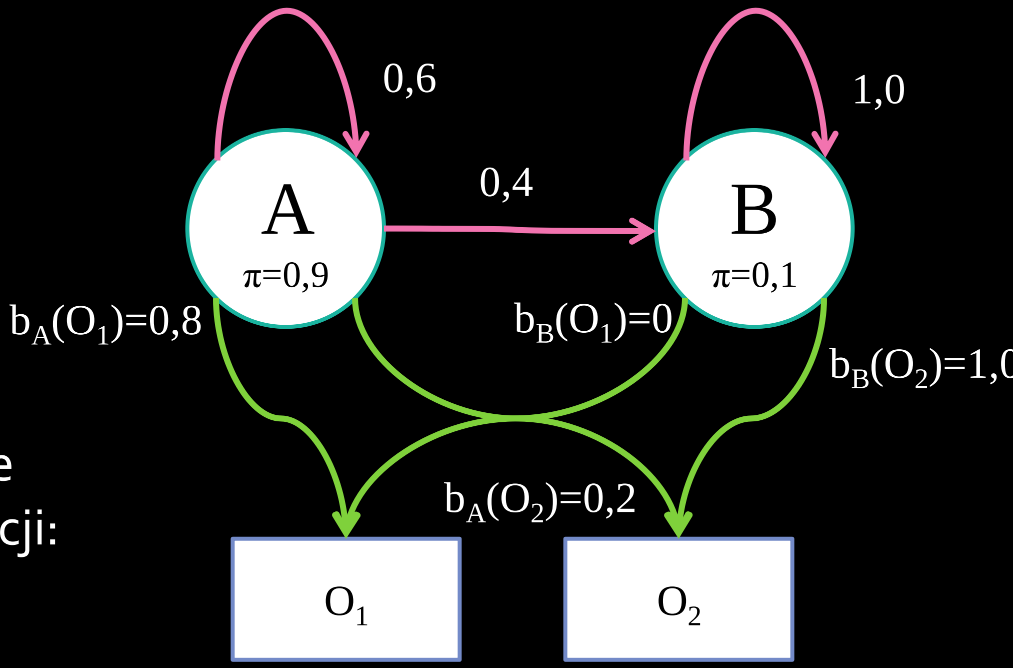
- Stwierdzenie, który model najlepiej pasuje do danego ciągu wejściowego.
  - tj. ma największe prawdopodobieństwo wygenerowania tego konkretnego ciągu zdarzeń.
  - obliczanie p-stwa (iteracyjne):

$$\alpha_1(i) = \pi_1 \cdot b_i(O_1)$$

$$\alpha_{t+1}(j) = [\ \textstyle\sum_{i=1}^{N} \alpha_t(i)a_{ij}] \cdot b_j(O_{t+1})$$

Przykład

- Jakie dla tego modelu jest p-stwo zajścia ciągu obserwacji $O=\{O_2, O_2, O_2\}$
- Uwzględnić wszystkie stany, które mogą prowadzić do tych obserwacji: AAA, AAB, ~~ABA,~~ ABB, BBB



$P(AAA) = \pi_A \cdot \alpha_{AA} \cdot \alpha_{AA} = 0,9 \cdot 0,6 \cdot 0,6 = 0,324$

P(…)=…. – p-stwo wystąpienia sekwencji stanów (dla wszystkich)

$P(O/AAA) = b_A(O_2) \cdot b_A(O_2) \cdot b_A(O_2) = 0,2 \cdot 0,2 \cdot 0,2 = 0,008$

P(O/…) = … – p-stwo obserwacji $O=\{O_{2,}\ O_{2,}\ O_2\}$ dla sekwencji (dla wszystkich)

$P(O/AAA) \cdot P(AAA) = 0,008 \cdot 0,324 = 0,002592$

P(O/…)·P(…) = … - p-stwo całkowite

Łącznie N^T obliczeń

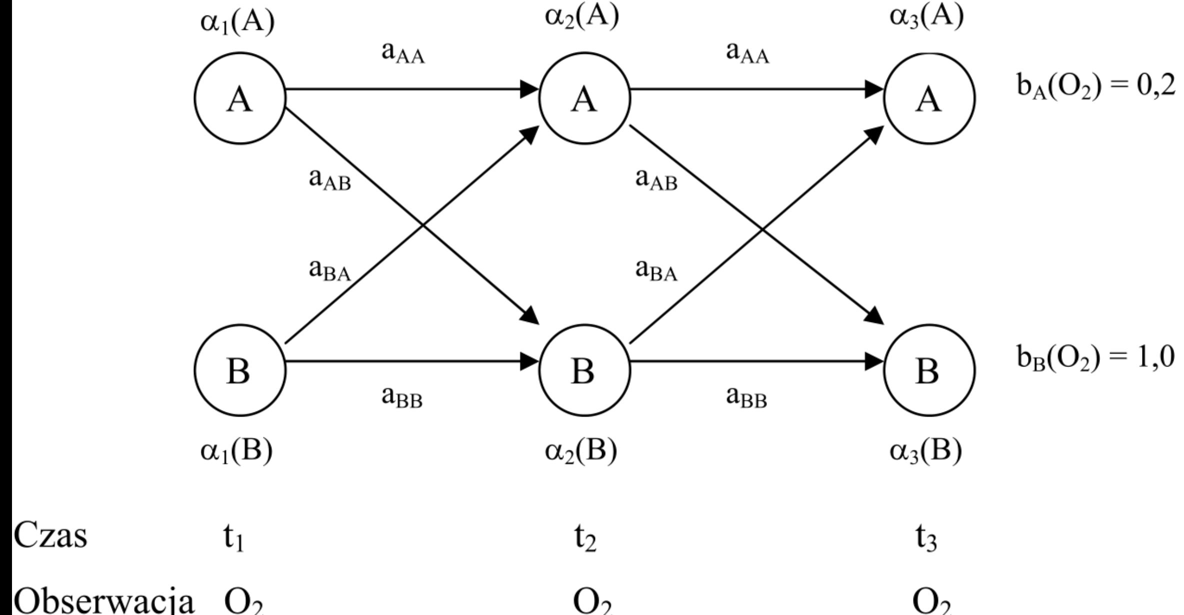
N-stanów

T-długość obserwacji

$P(O/M) = P(O/AAA) \cdot P(AAA) + P(O/AAB) \cdot P(AAB) + P(O/ABB) \cdot P(ABB) + P(O/BBB) \cdot P(BBB)$

P(O/M) – p-stwo zajścia obserwacji O dla tego modelu M

Przykład



$\alpha_1(i) = \pi_1 \cdot b_i(O_1)$

$\alpha_{t+1}(j) = [\ \sum_{i=1}^{N} \alpha_t(i)a_{ij}] \cdot b_j(O_{t+1})$

$\alpha_1(A) = \pi_A \cdot b_A(O_2) = 0{,}9 \cdot 0{,}2 = 0{,}18$

$\alpha_1(B) = \pi_B \cdot b_B(O_2) = 0{,}1 \cdot 1{,}0 = 0{,}10$

$\alpha_2(A) = (\alpha_1(A) \cdot a_{AA} + \alpha_1(B) \cdot a_{BA}) \cdot b_A(O_2) = (0{,}18 \cdot 0{,}6 + 0{,}1 \cdot 0{,}0) \cdot 0{,}2 = 0{,}0216$

$\alpha_2(B) = (\alpha_1(A) \cdot a_{AB} + \alpha_1(B) \cdot a_{BB}) \cdot b_B(O_2) = (0{,}18 \cdot 0{,}4 + 0{,}1 \cdot 1{,}0) \cdot 1{,}0 = 0{,}172$

$\alpha_3(A) = (\alpha_2(A) \cdot a_{AA} + \alpha_2(B) \cdot a_{BA}) \cdot b_A(O_2) = (0{,}0216 \cdot 0{,}6 + 0{,}172 \cdot 0{,}0) \cdot 0{,}2 = 0{,}002592$

$\alpha_3(B) = (\alpha_2(A) \cdot a_{AB} + \alpha_2(B) \cdot a_{BB}) \cdot b_B(O_2) = (0{,}0216 \cdot 0{,}4 + 0{,}172 \cdot 1{,}0) \cdot 1{,}0 = 0{,}18064$

Łącznie N·T obliczeń

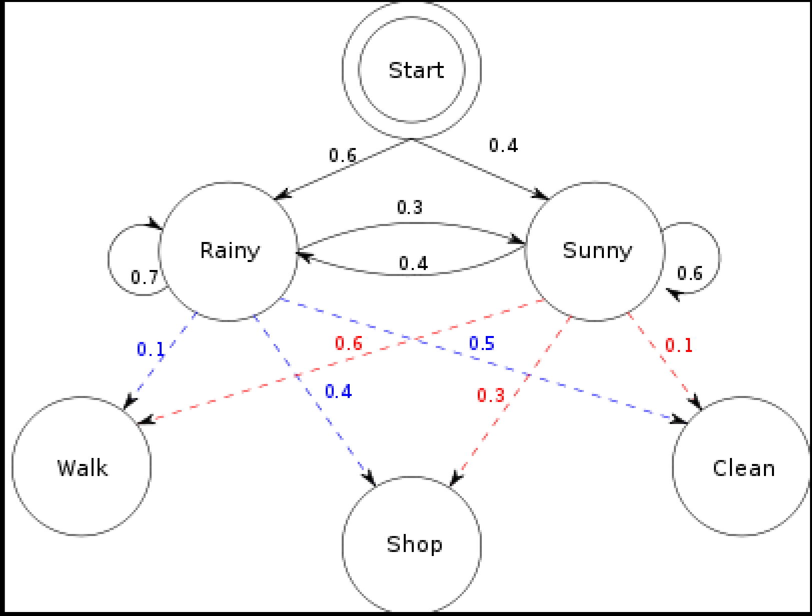
N-stanów

T-długość obserwacji

otrzymujemy:

$P(O/M) = \alpha_3(A) + \alpha_3(B) = 0{,}183232$

Przykład – rozmowa telefoniczna

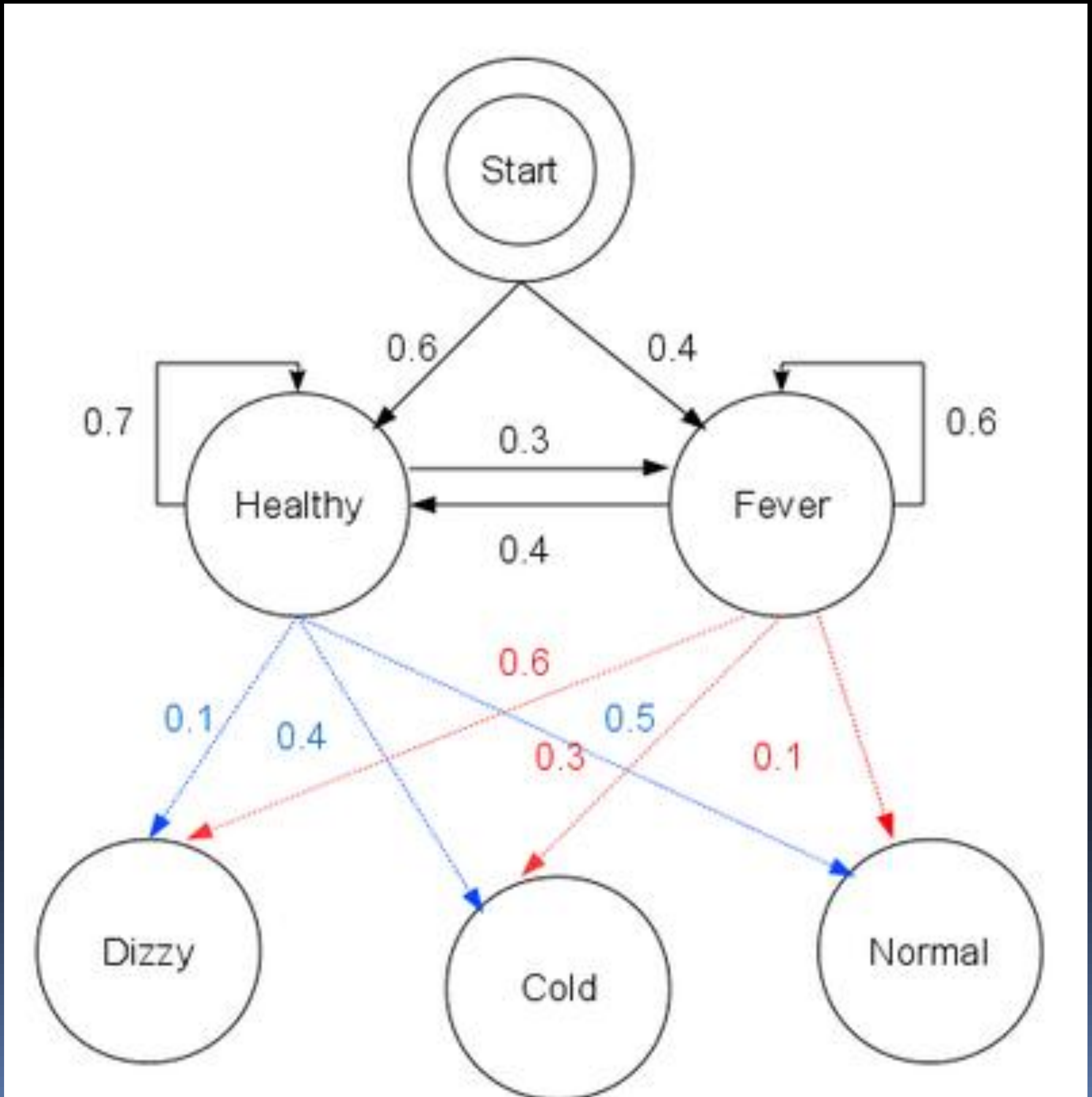


Bob do Alicji – „dziś sprzątam”

Alicja do Boba – „a więc u Ciebie pada”

- Stany = (‘Deszcz', ‘Słońce')
- Obserwacje = (‘Spacer', ‘Zakupy', ‘Sprzątanie')
- Pstwa_początkowe= {Deszcz': 0.6, ‘Słońce': 0.4}
- Pstwa_przejść=
  - { ‘Deszcz' : {'Deszcz': 0.7, 'Słońce': 0.3},
  - 'Słońce' : {'Deszcz': 0.4, 'Słońce': 0.6}, }
- Pstwa_obserwacji =
  - { 'Deszcz' : {‘Spacer': 0.1, ‘Zakupy': 0.4, ‘Sprzątanie': 0.5},
  - 'Słońce' : {' Spacer': 0.6, ‘Zakupy‘: 0.3, ‘Sprzątanie ': 0.1}, }

Przykład – symptomy choroby



Zastosowania

- Celem jest „odzyskanie” **sekwencji stanów**, które można obserwować tylko **pośrednio**, jako manifestację w postaci **„obserwacji”**
  - Kryptoanaliza
  - Rozpoznawanie mowy
  - Synteza mowy
  - Tłumaczenie maszynowe
  - Sekwencjonowanie genów
  - Rozpoznawanie czynności
  - Detekcja wirusów komputerowych
  - …

DZIĘKUJĘ ZA UWAGĘ