



AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE

Przetwarzanie Języka Naturalnego

Lab 5 – Łańcuchy Markova

Zbigniew Kaleta
`zkaleta@agh.edu.pl`

Wydział IEiT
Katedra Informatyki

08.04.2015



Łańcuchem Markowa nazywamy proces stochastyczny w czasie dyskretnym* i o dyskretnym zbiorze stanów** bez pamięci.

Brak pamięci oznacza, że kolejny stan osiągniany przez system zależy wprost jedynie od stanu bezpośrednio go poprzedzającego (oraz czasu).

Ciąg zmiennych losowych $X_0, X_1 \dots$ o wartościach całkowitych**

Warunek Markowa:

$$\forall n \in \mathbb{N} \forall i_0, i_1 \dots i_n, i_{n+1} \in \mathbb{Z} P(X_{n+1} = i_{n+1} | X_0 = i_0 \wedge X_1 = i_1 \wedge \dots \wedge X_n = i_n) = \\ P(X_{n+1} = i_{n+1} | X_n = i_n) = p(i_n, i_{n+1}) = p_{i_n i_{n+1}}$$

Łańcuch (czasowo) jednorodny - prawdopodobieństwo nie zależy od n

$$P = \begin{bmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,m} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m,0} & p_{m,1} & \cdots & p_{m,m} \end{bmatrix}$$

$$x^{(n)} = x^{(n-1)} * P$$

$$x^{(n)} = x^{(0)} * P^n$$

Łańcuch Markowa rzędu m (z pamięcią m)

$$\begin{aligned} P(X_{n+1} = i_{n+1} | X_0 = i_0 \wedge X_1 = i_1 \wedge \dots \wedge X_n = i_n) = \\ = P(X_{n+1} = i_{n+1} | X_{n-m+1} = i_{n-m+1} \wedge \dots \wedge X_n = i_n) \quad m \in \mathbb{Z}_+ \end{aligned}$$

$$Y_n = (X_n, X_{n-1}, X_{n-2}, \dots, X_{n-m+1})$$

- ✦ modelowanie procesów fizycznych (termodynamika, mechanika)
- ✦ chemia
- ✦ rozpoznawanie mowy
- ✦ generacja tekstów, muzyki
- ✦ teoria gier, sztuczna inteligencja
- ✦ ekonomia

Na podstawie modelu:

- 1 Skonstruować łańcuch na podstawie modelu.
- 2 Warunek końca: odpowiednia wielkość wygenerowanego obiektu lub osiągnięcie stanu pochłaniającego.

"Ala ma kota i psa"

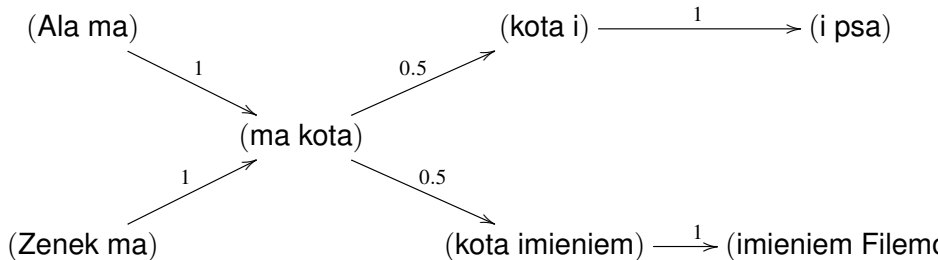
(Ala ma) \rightarrow (ma kota) \rightarrow (kota i) \rightarrow (i psa)

"Zenek ma kota imieniem Filemon"

(Zenek ma) \rightarrow (ma kota) \rightarrow (kota imieniem) \rightarrow (imieniem Filemon)

"Ala ma kota i psa"

"Zenek ma kota imieniem Filemon"



- 1 Napisać funkcję budującą łańcuch Markova z podanych n-gramów (1 pkt.)
- 2 Napisać funkcję generującą losowe notatki prasowe na podstawie korpusu PAP (1 pkt.)
- 3 Napisać funkcję generującą losowe słowa w języku polskim (1 pkt.)

Materiały:

<http://home.agh.edu.pl/~zkaleta/pjn/lab5.tar.gz>