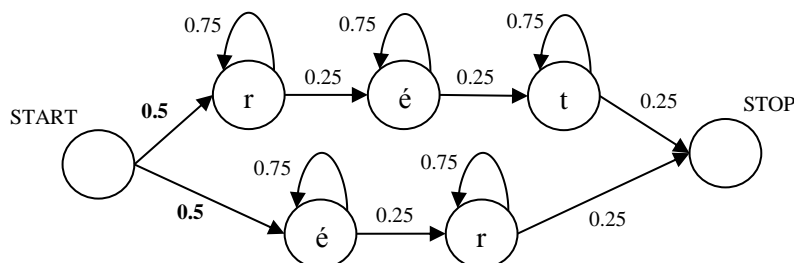


# BESZÉDINFORMÁCIÓS RENDSZEREK

## Dekódolás a beszédfelismerésben - Rejtett Markov-modellek

Tarján Balázs – tarjanb@tmit.bme.hu  
2017. november 14.

### 1. Gyakorló példa



a) Az ábrán látható rejtett Markov-modell hálózat milyen szavak felismerésére alkalmas, ha az egyes állapotok eloszlásfüggvényeit az azonosító betűjüknek megfelelő hangok jellemzővektoraival tanítottuk?

b) A fenti hálózatot izolált szavas beszédfelismerésre akarjuk használni. A beszédhangokat egy dimenziós, egység szórású Gauss-függvényekkel modellezzük, melyek várható értékei a következők:  $m_{r'} = 1.0$ ;  $m_{é'} = 2.0$ ;  $m_{t'} = 3.0$   
Összesen 3 db jellemzővektor érkezett a lényegkiemelőből:  $O_1=2.0$ ;  $O_2=2.0$ ;  $O_3=1.0$   
Melyik szó a felismerés eredménye?

### Megoldás

a) A „rét” és az „ér” szavakat tudja felismerni a modell.

### b) Normál megoldás

$\mu$  – megfigyelés (most valós szám)

$\delta$  – szórás

$m$  – várható érték

$$b(\mu) = \frac{1}{\sqrt{2\pi}\delta} e^{-\frac{(\mu-m)^2}{2\delta^2}} =_{(\delta=1)} \frac{1}{\sqrt{2\pi}} e^{-\frac{(\mu-m)^2}{2}}$$

$$b(\mu) = e^{-\frac{(\mu-m)^2}{2}} \left( \frac{1}{\sqrt{2\pi}} \text{ minden behelyettesítésben szereplő konstans} \right)$$

Ez alapján:

$$b(|O-m|=0) = 1.0$$

$$b(|O-m|=1) = 0.606$$

$$b(|O-m|=2) = 0.135$$

3 módon járható be a hálózat a megadott 3 jellemzővektorral

I. (r-é-t)

$$0.5 * b_r(O_1) * 0.25 * b_é(O_2) * 0.25 * b_t(O_3) * 0.25 = 0.5 * 0.606 * 0.25 * 1 * 0.25 * 0.135 * 0.25$$

II. (é-r-r)

$$0.5 * b_é(O_1) * 0.25 * b_r(O_2) * 0.75 * b_r(O_3) * 0.25 = 0.5 * 1 * 0.25 * 0.606 * 0.75 * 1 * 0.25$$

III. (é-é-r)

$$0.5 * b_é(O_1) * 0.75 * b_é(O_2) * 0.25 * b_r(O_3) * 0.25 = 0.5 * 1 * 0.75 * 1 * 0.25 * 1 * 0.25$$

Egyszerűsítés után:

I.  $0.606 \cdot 0.25 \cdot 0.135$

II.  $0.75 \cdot 0.606$

III.  $0.75 \Rightarrow$  Ez a legvalószínűbb útvonal, tehát a felismert szó: ér (é-é-r)

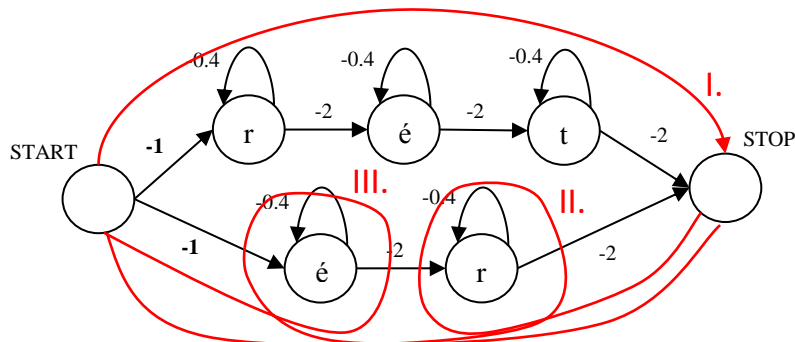
b) Megoldás logaritmussal

$$\log_2 b(o) = \log_2 e^{-\frac{(o-m)^2}{2}} = -\frac{(o-m)^2}{2} \log_2 e \sim -0.72(o-m)^2$$

$|o-m| = 0 ; \log_2 b(o) = 0$

$|o-m| = 1 ; \log_2 b(o) = -0.72$

$|o-m| = 2 ; \log_2 b(o) = -2.88$

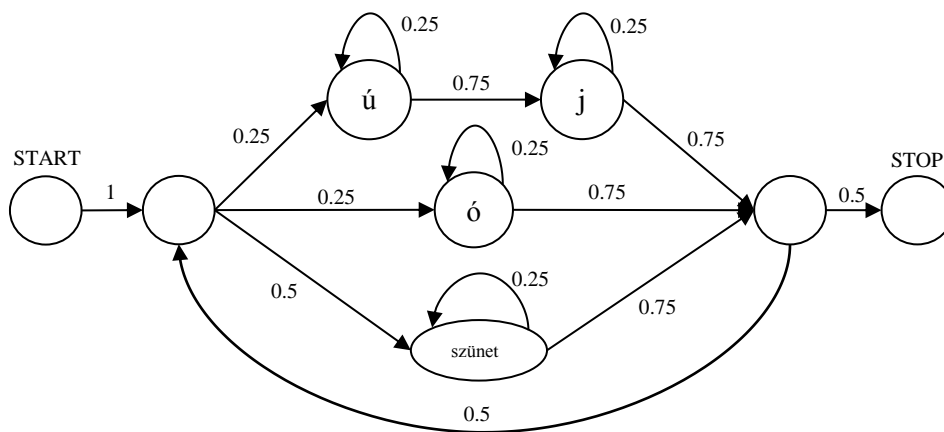


I.  $-1 + \log_2 b_r(o_1) - 2 + \log_2 b_\varepsilon(o_2) - 2 + \log_2 b_t(o_3) - 2 = -1 - 0.72 - 2 - 0 - 2 - 2.88 - 2 = -10.6$

II.  $-1 + \log_2 b_\varepsilon(o_1) - 2 + \log_2 b_r(o_2) - 0.4 + \log_2 b_r(o_3) - 2 = -1 - 0 - 2 - 0.72 - 0.4 - 0 - 2 = -6.12$

III.  $-1 + \log_2 b_\varepsilon(o_1) - 0.4 + \log_2 b_\varepsilon(o_2) - 2 + \log_2 b_r(o_3) - 2 = -1 - 0 - 0.4 - 0 - 2 - 0 - 2 = -5.4 \rightarrow$  Nyertes útvonal!

## 2. Gyakorló példa



a) Az ábrán látható rejtett Markov-modell hálózat milyen szavak felismerésére alkalmas, ha az egyes állapotok eloszlásfüggvényeit az azonosító betűjüknek megfelelő hangok jellemzővektoraival tanítottuk?

b) A fenti hálózatot beszédfelismerésre akarjuk használni. A beszédhangokat és a szünetmodell egyaránt egy dimenziós, egység szórású Gauss-függvényekkel modellezzük, melyek várható értékei a következők:

$m_{\text{ú}} = 3.0$ ;  $m_{\text{j}} = 2.0$ ;  $m_{\text{ó}} = 4.0$ ;  $m_{\text{szünet}} = 5.0$

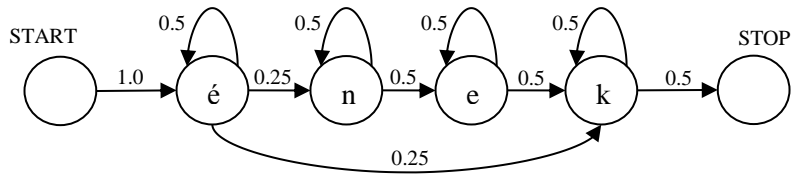
Összesen 2 db jellemzővektor érkezett a lényegkiemelőből:  $O_1 = 3.0$ ;  $O_2 = 4.0$

Mi a felismerés eredménye?

### Megoldás

- a) ó, új
- b) ó (ó-ó)

### 3. Gyakorló példa

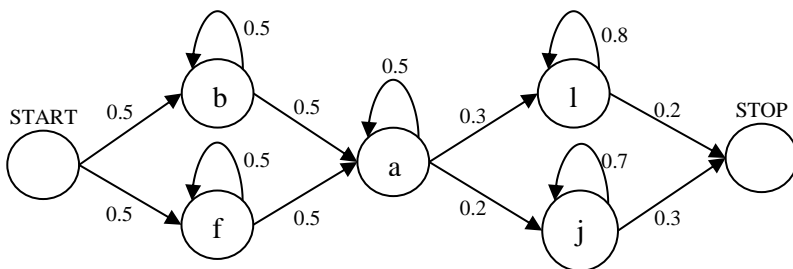


- a) Az ábrán látható rejtett Markov-modell hálózat milyen szavak felismerésére alkalmas, ha az egyes állapotok eloszlásfüggvényeit az azonosító betűjüknek megfelelő hangok jellemzővektoraival tanítottuk?
- b) A fenti hálózatot izolált szavas beszéd felismerésre akarjuk használni. A beszédhangokat egy dimenziós, egység szórású Gauss-függvényekkel modellezzük, melyek várható értékei a következők:  $m_{\text{é}} = 2.0$ ;  $m_{\text{n}} = 4.0$ ;  $m_{\text{e}} = 1.0$ ;  $m_{\text{k}} = 3.0$ . Összesen 4 db jellemzővektor érkezett a lényegkiemelőből:  $O_1=1.0$ ;  $O_2=2.0$ ;  $O_3=2.0$ ;  $O_4=4.0$ . Melyik szó a felismerés eredménye?

### Megoldás

- a) ének, ék
- b) ék (é-é-é-k)

### 4. Gyakorló példa



- a) Az ábrán látható rejtett Markov-modell hálózat milyen értelmes, magyar szavak felismerésére alkalmas, ha az egyes állapotok eloszlásfüggvényeit az azonosító betűjüknek megfelelő hangok jellemzővektoraival tanítottuk?
- b) A fenti hálózatot izolált szavas beszéd felismerésre akarjuk használni. A beszédhangokat egy dimenziós, egység szórású Gauss-függvényekkel modellezzük, melyek várható értékei a következők:  $m_{\text{b}} = 0.0$ ;  $m_{\text{f}} = 2.0$ ;  $m_{\text{a}} = 3.0$ ;  $m_{\text{l}} = 1.0$ ;  $m_{\text{j}} = 4.0$ . Összesen 3 db jellemzővektor érkezett a lényegkiemelőből, sorrendben:  $O_1 = 2.0$ ;  $O_2 = 4.0$ ;  $O_3 = 3.0$ . Melyik szó a felismerés eredménye?

### Megoldás

- a) bal, baj, fal, faj
- b) faj (f-a-j)