Modele Markov Ascunse

De la Teorie la Aplicații

Alexandru Sorici, Tudor Berariu

Asociația Română pentru Inteligență Artificială
în colaborare cu
Laboratorul AI-MAS

7 noiembrie 2012

Outline

1. Aplicații în Învățarea Automată pentru MMA
   - Învățarea Automată
   - MMA în Învățarea Automată

Outline

1. Aplicații în Învățarea Automată pentru MMA
   - Învățarea Automată
   - MMA în Învățarea Automată
2. Teoria MMA
   - Cele Trei Probleme ale MMA
   - Fundamente Matematice

Outline

1. Aplicații în Învățarea Automată pentru MMA
   - Învățarea Automată
   - MMA în Învățarea Automată
2. Teoria MMA
   - Cele Trei Probleme ale MMA
   - Fundamente Matematice
3. Implementarea MMA
   - Problema Evaluării: Algoritmul Forward-Backward
   - Problema Interpretării: Algoritmul Viterbi
   - Problema Estimării: Algoritmul Baum-Welch

Outline

1. Aplicații în Învățarea Automată pentru MMA
   - Învățarea Automată
   - MMA în Învățarea Automată
2. Teoria MMA
   - Cele Trei Probleme ale MMA
   - Fundamente Matematice
3. Implementarea MMA
   - Problema Evaluării: Algoritmul Forward-Backward
   - Problema Interpretării: Algoritmul Viterbi
   - Problema Estimării: Algoritmul Baum-Welch
4. Demo: Recunoașterea Simbolurilor

Outline

1. Aplicații în Învățarea Automată pentru MMA
   - Învățarea Automată
   - MMA în Învățarea Automată
2. Teoria MMA
   - Cele Trei Probleme ale MMA
   - Fundamente Matematice
3. Implementarea MMA
   - Problema Evaluării: Algoritmul Forward-Backward
   - Problema Interpretării: Algoritmul Viterbi
   - Problema Estimării: Algoritmul Baum-Welch
4. Demo: Recunoașterea Simbolurilor
5. Tipuri de MMA

Outline

1. Aplicații în Învățarea Automată pentru MMA
   - Învățarea Automată
   - MMA în Învățarea Automată
2. Teoria MMA
   - Cele Trei Probleme ale MMA
   - Fundamente Matematice
3. Implementarea MMA
   - Problema Evaluării: Algoritmul Forward-Backward
   - Problema Interpretării: Algoritmul Viterbi
   - Problema Estimării: Algoritmul Baum-Welch
4. Demo: Recunoașterea Simbolurilor
5. Tipuri de MMA
6. Discuții și Concluzii

Outline

1. Aplicații în Învățarea Automată pentru MMA
   - Învățarea Automată
   - MMA în Învățarea Automată
2. Teoria MMA
   - Cele Trei Probleme ale MMA
   - Fundamente Matematice
3. Implementarea MMA
   - Problema Evaluării: Algoritmul Forward-Backward
   - Problema Interpretării: Algoritmul Viterbi
   - Problema Estimării: Algoritmul Baum-Welch
4. Demo: Recunoașterea Simbolurilor
5. Tipuri de MMA
6. Discuții și Concluzii

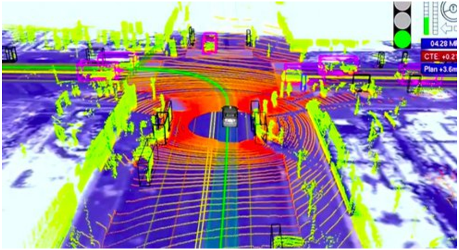
Ce este Învățarea Automată?

Învățarea automată

Un program se spune că învață dintr-o mulțime de experiențe E relativ la o clasă de task-uri T și o măsură a performaneței P dacă performața acestuia la execuția task-urilor T, măsurată de P, se îmbunătățește în urma experiențelor E. [Mit97]

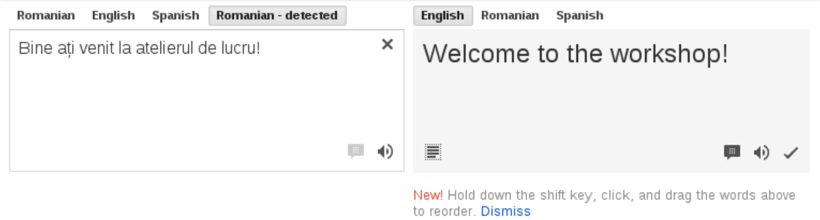
Aplicații ale Învățării Automate

- Self-Driving Car: Google Car
- Traducere Automată: Google Translate
- Sisteme de Recomandare
    - Filme: ImDB, NetFlix
    - Publicitate Inteligentă: Google Ads, Facebook Ads



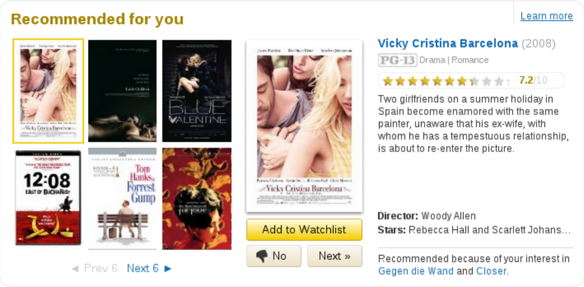
Aplicații ale Învățării Automate

- Self-Driving Car: Google Car
- Traducere Automată: Google Translate
- Sisteme de Recomandare
  - Filme: ImDB, NetFlix
  - Publicitate Inteligentă: Google Ads, Facebook Ads



Aplicații ale Învățării Automate

- Self-Driving Car: Google Car
- Traducere Automată: Google Translate
- Sisteme de Recomandare
  - Filme: ImDB, NetFlix
  - Publicitate Inteligentă: Google Ads, Facebook Ads



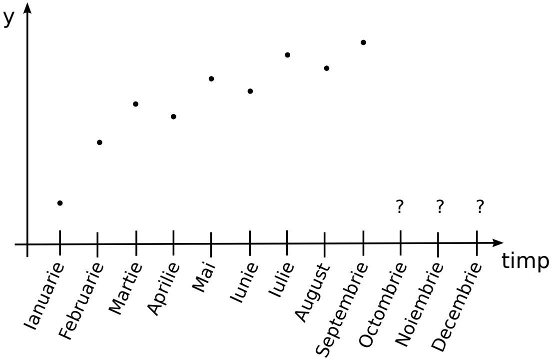
Aplicații ale Învățării Automate

- Self-Driving Car: Google Car
- Traducere Automată: Google Translate
- Sisteme de Recomandare
  - Filme: ImDB, NetFlix
  - Publicitate Inteligentă: Google Ads, Facebook Ads

Învățarea Automată: Tipuri de probleme

Tipuri de probleme

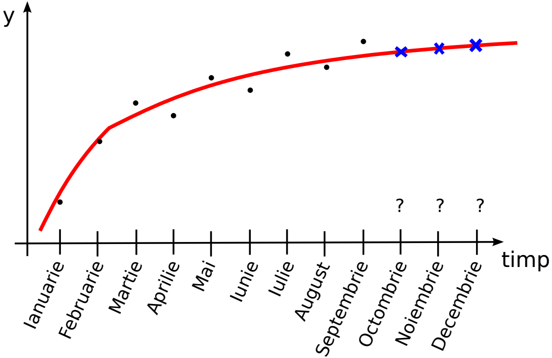
- Regresie
  - predicția evoluției prețului unui bun



Învățarea Automată: Tipuri de probleme

Tipuri de probleme

- Regresie
  - predicția evoluției prețului unui bun



Învățarea Automată: Tipuri de probleme

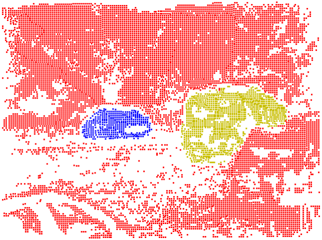
Tipuri de probleme

- Regresie
  - predicția evoluției prețului unui bun
- Clasificare
  - clasificarea obiectelor dintr-o imagine (imagine de la Albert-Ludwigs-Universität, Lehrstuhl für Mustererkennung und Bildverarbeitung)

Învățarea Automată: Tipuri de probleme

Tipuri de probleme

- Regresie
  - predicția evoluției prețului unui bun
- Clasificare
  - clasificarea obiectelor dintr-o imagine
    (imagine de la Albert-Ludwigs-Universität, Lehrstuhl für Mustererkennung und Bildverarbeitung)



Învățarea Automată: Tipuri de probleme

Tipuri de probleme

- Regresie
  - predicția evoluției prețului unui bun
- Clasificare
  - clasificarea obiectelor dintr-o imagine (imagine de la Albert-Ludwigs-Universität, Lehrstuhl für Mustererkennung und Bildverarbeitung)
- Învățare prin Recompensă
  - jucător inteligent pentru backgammon

Tipuri de învățare

Învățare Supervizată

Învățarea supervizată se face pe baza unor date de antrenare etichetate.

- construirea unui model climatic pe baza datelor din utimii 30 de ani (regresie)
- construirea unui detector de mesaje spam (clasificare)

Tipuri de învățare

Învățare Supervizată

Învățarea supervizată se face pe baza unor date de antrenare etichetate.

- construirea unui model climatic pe baza datelor din utimii 30 de ani (regresie)
- construirea unui detector de mesaje spam (clasificare)

Învățare Nesupervizată

În învățarea neupervizată nu există date etichetate.

- detectarea profilurilor de comportament ale cumpărătorilor

Outline

1. Aplicații în Învățarea Automată pentru MMA
   - Învățarea Automată
   - MMA în Învățarea Automată
2. Teoria MMA
   - Cele Trei Probleme ale MMA
   - Fundamente Matematice
3. Implementarea MMA
   - Problema Evaluării: Algoritmul Forward-Backward
   - Problema Interpretării: Algoritmul Viterbi
   - Problema Estimării: Algoritmul Baum-Welch
4. Demo: Recunoașterea Simbolurilor
5. Tipuri de MMA
6. Discuții și Concluzii

Probleme cu Secvențe Temporale (I)

URMĂRIREA OBIECTELOR

Urmărirea Vorbitorului

GPS

Robotică

Rachete Sol-Aer

Sistem de Navigare pentru Nave

RECUNOAȘTEREA VORBIRII

Interfețe utilizator vocale

Procesare speech-to-text

Comanda Vocală Directă - Avioane

RECUNOAȘTEREA GESTURILOR

Recunoașterea Personalizată a Semnăturilor

Recunoaștera Limbajului Semenlor

Probleme cu Secvențe Temporale (I)

URMĂRIREA OBIECTELOR

Urmărirea Vorbitorului

GPS

Robotică

Rachete Sol-Aer

Sistem de Navigare pentru Nave

RECUNOAȘTEREA VORBIRII

Interfețe utilizator vocale

Procesare speech-to-text

Comanda Vocală Directă - Avioane

RECUNOAȘTEREA GESTURILOR

Recunoașterea Personalizată a Semnăturilor

Recunoaștera Limbajului Semenlor

Probleme cu Secvențe Temporale (I)

URMĂRIREA OBIECTELOR

Urmărirea Vorbitorului

GPS

Robotică

Rachete Sol-Aer

Sistem de Navigare pentru Nave

RECUNOAȘTEREA VORBIRII

Interfețe utilizator vocale

Procesare speech-to-text

Comanda Vocală Directă - Avioane

RECUNOAȘTEREA GESTURILOR

Recunoașterea Personalizată a Semnăturilor

Recunoaștera Limbajului Semenlor

Probleme cu Secvențe Temporale (I)

URMĂRIREA OBIECTELOR

Urmărirea Vorbitorului

GPS

Robotică

Rachete Sol-Aer

Sistem de Navigare pentru Nave

RECUNOAȘTEREA VORBIRII

Interfețe utilizator vocale

Procesare speech-to-text

Comanda Vocală Directă - Avioane

RECUNOAȘTEREA GESTURILOR

Recunoașterea Personalizată a Semnăturilor

Recunoaștera Limbajului Semenlor

Probleme cu Secvențe Temporale (I)

URMĂRIREA OBIECTELOR

Urmărirea Vorbitorului

GPS

Robotică

Rachete Sol-Aer

Sistem de Navigare pentru Nave

RECUNOAȘTEREA VORBIRII

Interfețe utilizator vocale

Procesare speech-to-text

Comanda Vocală Directă - Avioane

RECUNOAȘTEREA GESTURILOR

Recunoașterea Personalizată a Semnăturilor

Recunoaștera Limbajului Semenlor

Probleme cu Secvențe Temporale (I)

URMĂRIREA OBIECTELOR

Urmărirea Vorbitorului

GPS

Robotică

Rachete Sol-Aer

Sistem de Navigare pentru Nave

RECUNOAȘTEREA VORBIRII

Interfețe utilizator vocale

Procesare speech-to-text

Comanda Vocală Directă - Avioane

RECUNOAȘTEREA GESTURILOR

Recunoașterea Personalizată a Semnăturilor

Recunoaștera Limbajului Semenlor

Probleme cu Secvențe Temporale (II)

BIOINFORMATICĂ

Secvențierea Proteinelor

Modelarea unei Rețele Regulatoare Genetice

ECONOMIE

Predicția Valorilor Bursiere

Econometrie

- estimarea în timp a indicatorilor economici ai unei țari -

Probleme cu Secvențe Temporale (II)

BIOINFORMATICĂ

Secvențierea Proteinelor

Modelarea unei Rețele
Regulatoare Genetice

ECONOMIE

Predicția Valorilor Bursiere

Econometrie
- estimarea în timp a indicatorilor economici ai unei țari -

Raționament Probabilistic Temporal - Modele

Să ne gândim la unele din problemele anterioare ...

Raționament Probabilistic Temporal - Modele

Să ne gândim la unele din problemele anterioare ...

Cum modelăm astfel de situații dinamice?

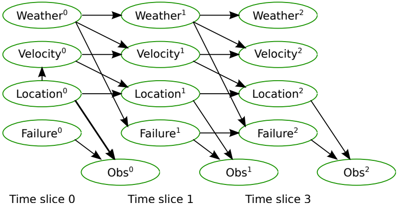
Raționament Probabilistic Temporal - Modele

Să ne gândim la unele din problemele anterioare ...

Cum modelăm astfel de situații dinamice?

Stări și Observații

- Procesul de schimbare este văzut ca o serie de snapshot-uri
- Fiecare snapshot conține un set de variabile aleatoare
  - $\mathbf{O}_t$ - setul tuturor variabilelor de măsurare (*observabile*) la momentul t
  - $\mathbf{Q}_t$ - setul tuturor variabilelor de stare (*neobservabile / ascunse*) la momentul t



Exemplu: problemă de localizare a unui vehicul [KF09]

Raționament Probabilistic Temporal - Presupuneri

Să ne gândim la unele din problemele anterioare ...

Raționament Probabilistic Temporal - Presupuneri

Să ne gândim la unele din problemele anterioare ...

Ce presupuneri facem?

Raționament Probabilistic Temporal - Presupuneri

Să ne gândim la unele din problemele anterioare ...

Ce presupuneri facem?

Proces staționar

Procesul de schimbare este guvernat de legi care nu se schimba in timp.
Urmare: trebuie sa specificăm relațiile între variabile doar pentru un snapshot *reprezentativ*.

Raționament Probabilistic Temporal - Presupuneri

Să ne gândim la unele din problemele anterioare ...

Ce presupuneri facem?

Proces staționar

Procesul de schimbare este guvernat de legi care nu se schimba in timp.
Urmare: trebuie sa specificăm relațiile între variabile doar pentru un snapshot *reprezentativ*.

Presupunerea Markov

Starea curentă a unui proces de schimbare depinde doar de o istorie finită de stări anterioare.
Urmare: avem un număr limitat de "parinți" pentru variabilele din fiecare snapshot.

Raționament Probabilistic Temporal - Inferență

Care sunt principalele inferențe ce se doresc făcute?

Raționament Probabilistic Temporal - Inferență

Care sunt principalele inferențe ce se doresc făcute?

Filtrare (monitorizare)

Sarcina de a calcula starea de fapt - distribuția posterioară de probabilitate a stării curente, date fiind toate observațiile de până acum.

Raționament Probabilistic Temporal - Inferență

Care sunt principalele inferențe ce se doresc făcute?

Filtrare (monitorizare)

Sarcina de a calcula starea de fapt - distribuția posterioară de probabilitate a stării curente, date fiind toate observațiile de până acum.

Evaluare

Sarcina de a calcula probabilitatea (likelihood) a observațiilor făcute până în prezent.

Raționament Probabilistic Temporal - Inferență

Care sunt principalele inferențe ce se doresc făcute?

Predicție

Sarcina de a calcula distribuția posterioară de probabilitate peste o stare viitoare, date fiind toate observațiile de până acum.

Raționament Probabilistic Temporal - Inferență

Care sunt principalele inferențe ce se doresc făcute?

Predicție

Sarcina de a calcula distribuția posterioară de probabilitate peste o stare viitoare, date fiind toate observațiile de până acum.

Netezire (hindsight)

Sarcina de a calcula distribuția posterioară de probabilitate peste o stare anterioară, date fiind toate observațiile de până acum.
Furnizează o estimare mai buna asupra stării respective, decât a fost posibil la momentul respectiv.

Raționament Probabilistic Temporal - Inferență

Care sunt principalele inferențe ce se doresc făcute?

Cea mai probabilă explicație

Dându-se o *secvență de observații*, se cere găsirea celei mai probabile secvenței de stări care a generat acele observații.

Raționament Probabilistic Temporal - Inferență

Care sunt principalele inferențe ce se doresc făcute?

Cea mai probabilă explicație

Dându-se o *secvență de observații*, se cere găsirea celei mai probabile secvenței de stări care a generat acele observații.

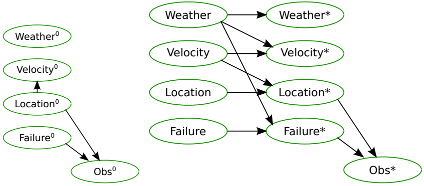
Învățare

Dându-se *un set de secvențe de observații*, găsește o metodă de a învăța modelele de tranziție și senzoriale / de măsurare pe baza acelor observații.

Raționament Probabilistic Temporal - Metode

Rețele Bayesiene Dinamice (RBD)

O RBD este o rețea Bayesiană ce reprezintă un model temporal de probabilitate.



(a) Rețeaua la momentul 0

(b) Rețeaua Bayesiană în două snapshot-uri

(c) RBD desfășurată pe 3 momente de timp

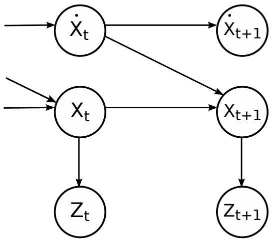
RBD simplificată pentru monitorizarea unui vehicul [KF09]

Aplicată în probleme precum: urmărirea obiectelor, recunoașterea activității umane, secvențierea proteinelor, etc.

Raționament Probabilistic Temporal - Metode

Filtre Kalman (Sistem Dinamice Lineare)

Un model temporal având una sau mai multe variabile care evoluează linear în timp, la care se adaugă zgomot Gaussian.



Structura unei RB pentru un sistem linear dinamic cu variabile de poziție X_t, viteză $\dot{X}_t$, și măsurare a poziției Z_t

- Poate fi văzut ca o RBD în care toate variabilele sunt continue, iar dependențele sunt linear gaussiane.
- Aplicații multiple în **urmărirea obiectelor**

Raționament Probabilistic Temporal - Metode

Modele Markov Ascunse (MMA)

Un MMA (HMM) este un model probabilistic temporal în care *starea* procesului de schimbare este descrisă de o singură variabilă aleatoare discretă. Valorile posibile ale variabilei reprezintă stările posibile ale lumii modelate.

Utilizat cu succes in aplicații precum:

- Recunoașterea Scrisului
- Recunoașterea Gesturilor
- Recunoașterea Vorbirii
- Determinarea Parților de Vorbire (Part-of-Speech Tagging)
- Secvențiere ADN

Outline

1. Aplicații în Învățarea Automată pentru MMA
   - Învățarea Automată
   - MMA în Învățarea Automată
2. Teoria MMA
   - Cele Trei Probleme ale MMA
   - Fundamente Matematice
3. Implementarea MMA
   - Problema Evaluării: Algoritmul Forward-Backward
   - Problema Interpretării: Algoritmul Viterbi
   - Problema Estimării: Algoritmul Baum-Welch
4. Demo: Recunoașterea Simbolurilor
5. Tipuri de MMA
6. Discuții și Concluzii

Cele 3 Probleme ale MMA

3 probleme fundamentale [Rab89]

- Particularizarea inferenței în probleme cu secvențe temporale pe cazul MMA
- Structura restricționată a MMA permite implementări elegante ale tuturor algoritmilor de bază

Cele 3 Probleme ale MMA

3 probleme fundamentale [Rab89]

- Particularizarea inferenței în probleme cu secvențe temporale pe cazul MMA
- Structura restricționată a MMA permite implementări elegante ale tuturor algoritmilor de bază

Problema Evaluării

Dându-se un model și o secvență de observații, cum calculăm probabilitatea ca secvența observată sa fi fost produsă de model?

Cele 3 Probleme ale MMA

3 probleme fundamentale [Rab89]

- Particularizarea inferenței în probleme cu secvențe temporale pe cazul MMA
- Structura restricționată a MMA permite implementări elegante ale tuturor algoritmilor de bază

Problema Evaluării

Dându-se un model și o secvență de observații, cum calculăm probabilitatea ca secvența observată sa fi fost produsă de model?

Problema Interpretării (cea mai bună explicație a observațiilor)

Dându-se un model și o secvență de observații, cum alegem o secvență corespunzătoare de stări care *dau sens* observațiilor?

Cele 3 Probleme ale MMA

3 probleme fundamentale [Rab89]

- Particularizarea inferenței în probleme cu secvențe temporale pe cazul MMA
- Structura restricționată a MMA permite implementări elegante ale tuturor algoritmilor de bază

Problema Evaluării

Dându-se un model și o secvență de observații, cum calculăm probabilitatea ca secvența observată sa fi fost produsă de model?

Problema Interpretării (cea mai bună explicație a observațiilor)

Dându-se un model și o secvență de observații, cum alegem o secvență corespunzătoare de stări care *dau sens* observațiilor?

Problema Estimării (Antrenării) Modelului

Dându-se mai multe secvențe de observații, cum putem ajusta parametrii modelului MMA care explică cel mai bine observațiile făcute?

Outline

1. Aplicații în Învățarea Automată pentru MMA
   - Învățarea Automată
   - MMA în Învățarea Automată
2. Teoria MMA
   - Cele Trei Probleme ale MMA
   - Fundamente Matematice
3. Implementarea MMA
   - Problema Evaluării: Algoritmul Forward-Backward
   - Problema Interpretării: Algoritmul Viterbi
   - Problema Estimării: Algoritmul Baum-Welch
4. Demo: Recunoașterea Simbolurilor
5. Tipuri de MMA
6. Discuții și Concluzii

O perspectivă teoretică asupra MMA

- Vom răspunde la următoarele întrebări:

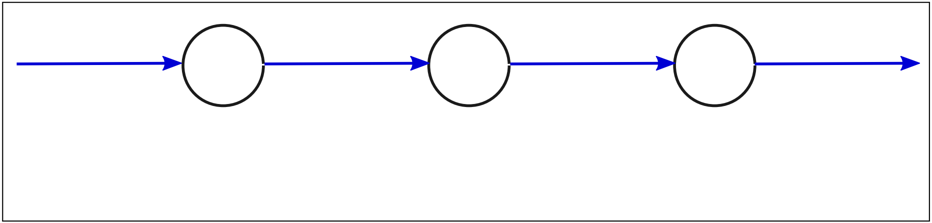
1. Cum definim un Model Markov Ascuns?
2. Cum exprimăm cele trei întrebări fundamentale cu ajutorul probabilităților?

Ce este un MMA?

Definiție

Un Model Markov Ascuns este un dublu proces stocastic cu două componente:

- un proces Markov

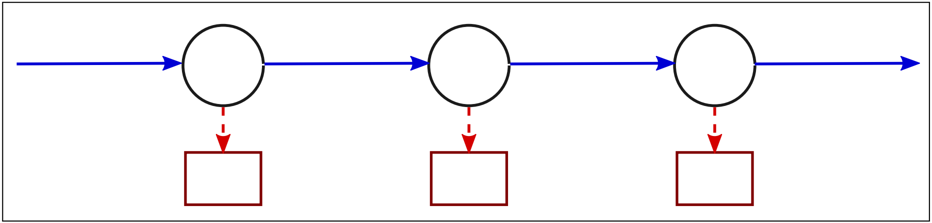


Ce este un MMA?

Definiție

Un Model Markov Ascuns este un dublu proces stocastic cu două componente:

- un proces Markov *neobservabil* (*ascuns*),
- un set de procese stocastice care produc partea *observabilă*.



Exemplu: Urmărirea stărilor emoționale

Să considerăm următorul exmeplu:

- un robot ce urmărește evoluția stărilor emoționale ale unui om

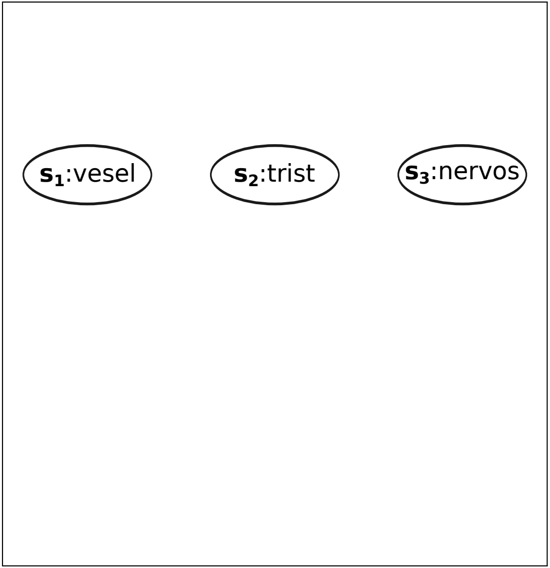
Senzor:

- cameră video

Exemplu adaptat după :

R. Zubek. Introduction to hidden markov models. *AI Game Programming Wisdom*, 3:633–646, 2006

Exemplu: Urmărirea stărilor emoționale



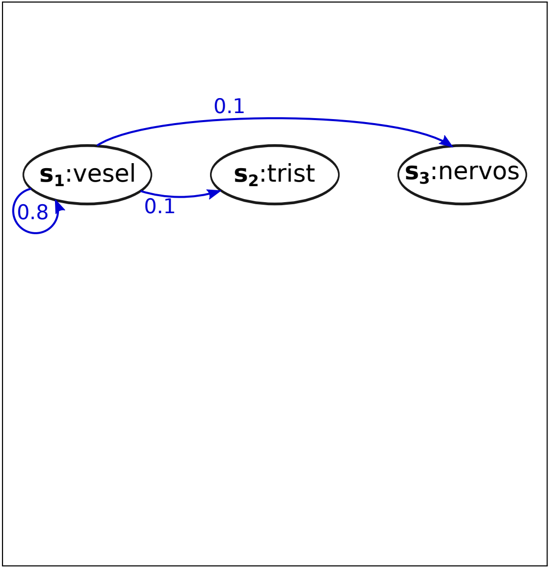
N - numărul de stări ascunse
S - mulțimea stărilor

$\mathbf{N} = 3$

Stări:

- $\mathbf{s_1}$: vesel
- $\mathbf{s_2}$: trist
- $\mathbf{s_3}$: nervos

Exemplu: Urmărirea stărilor emoționale



A - matricea distribuțiilor de probabilitate ale tranzițiilor între stări

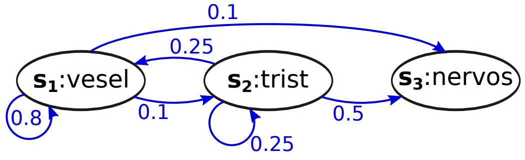
$\mathbf{A} = \{a_{i,j}\}$, $1 \leq i, j \leq N$

$a_{i,j} = P(q_{t+1} = s_j | q_t = s_i)$

$$\sum_{j=1}^{N} a_{i,j} = 1, \quad 1 \leq i \leq N$$

$$\mathbf{A} = \begin{array}{c c} & \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ & & \\ & & \end{pmatrix} \end{array}$$

Exemplu: Urmărirea stărilor emoționale



A - matricea distribuțiilor de probabilitate ale tranzițiilor între stări

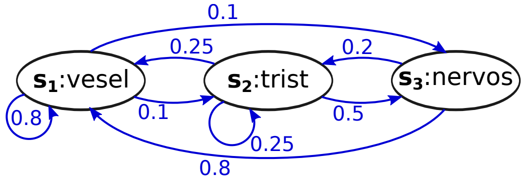
$\mathbf{A} = \{a_{i,j}\},\ 1 \leq i,j \leq N$

$a_{i,j} = P(q_{t+1} = s_j | q_t = s_i)$

$$\sum_{j=1}^{N} a_{i,j} = 1, \quad 1 \leq i \leq N$$

$$\mathbf{A} = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.25 & 0.25 & 0.5 \\ & & \end{pmatrix} \end{array}$$

Exemplu: Urmărirea stărilor emoționale



A - matricea distribuțiilor de probabilitate ale tranzițiilor între stări

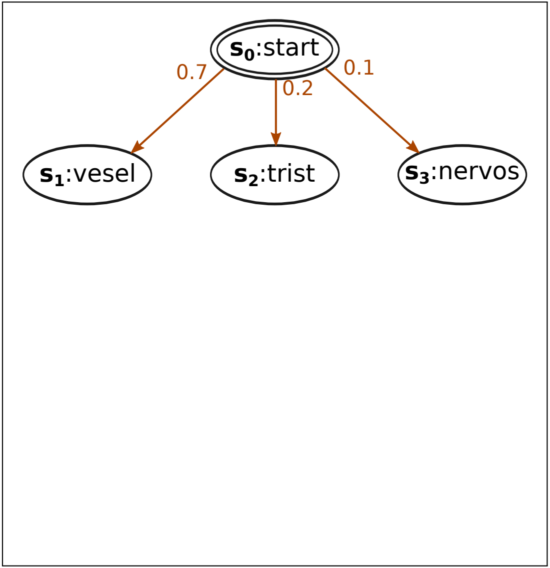
$\mathbf{A} = \{a_{i,j}\},\ 1 \leq i,j \leq N$

$a_{i,j} = P(q_{t+1} = s_j | q_t = s_i)$

$$\sum_{j=1}^{N} a_{i,j} = 1, \quad 1 \leq i \leq N$$

$$\mathbf{A} = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.25 & 0.25 & 0.5 \\ 0.8 & 0.2 & 0 \end{pmatrix} \end{array}$$

Exemplu: Urmărirea stărilor emoționale



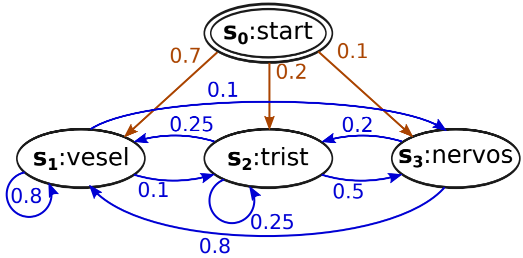
$\mathbf{\Pi}$ - distribuția stării inițiale

$$\mathbf{\Pi} = \{\pi_i\}, \quad 1 \le i \le N$$

$$\pi_i = P(q_1 = s_i)$$

$$\mathbf{\Pi} = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.7 & 0.2 & 0.1 \end{pmatrix} \end{array}$$

Exemplu: Urmărirea stărilor emoționale

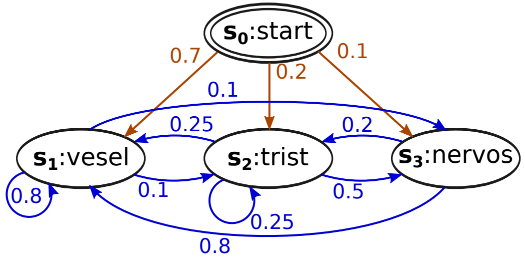


Deocamdată am descris un *proces Markov*.

$$A = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.25 & 0.25 & 0.5 \\ 0.8 & 0.2 & 0 \end{pmatrix} \end{array}$$

$$\Pi = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.7 & 0.2 & 0.1 \end{pmatrix} \end{array}$$

Exemplu: Urmărirea stărilor emoționale



vesel → vesel → nervos → trist

$\mathbf{Q} = [\mathbf{q_1}{:}s_1 \ \mathbf{q_2}{:}s_1 \ \mathbf{q_3}{:}s_3 \ \mathbf{q_4}{:}s_2]$

Deocamdată am descris un *proces Markov*.

$$A = \begin{array}{c|ccc} & s_1 & s_2 & s_3 \\ \hline s_1 & 0.8 & 0.1 & 0.1 \\ s_2 & 0.25 & 0.25 & 0.5 \\ s_3 & 0.8 & 0.2 & 0 \end{array}$$

$$\Pi = \begin{array}{ccc} s_1 & s_2 & s_3 \\ \hline 0.7 & 0.2 & 0.1 \end{array}$$

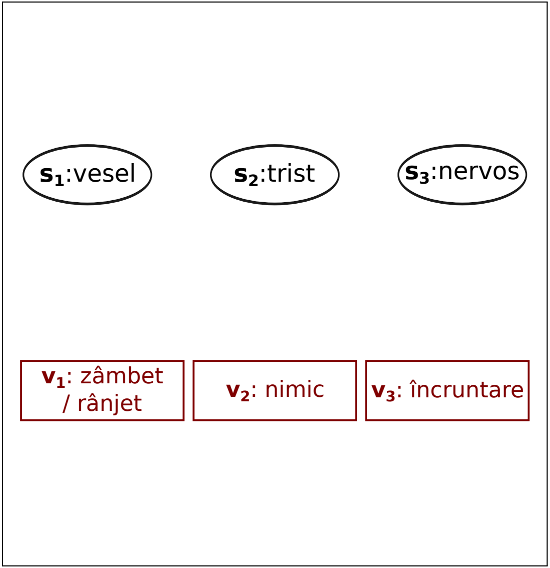
Notație: $\mathbf{Q} = [q_1 q_2 \cdots q_T]$

$P(Q|A,\Pi) = \pi_{q_1} a_{q_1,q_2} \cdots a_{q_{T-1},q_T}$

$P(s_1,s_1,s_3,s_2|A,\Pi) = \pi_1 \cdot a_{1,1} \cdot a_{1,3} \cdot a_{3,2} =$

$= 0.8 \cdot 0.8 \cdot 0.1 \cdot 0.2 = 0.0128$

Exemplu: Urmărirea stărilor emoționale



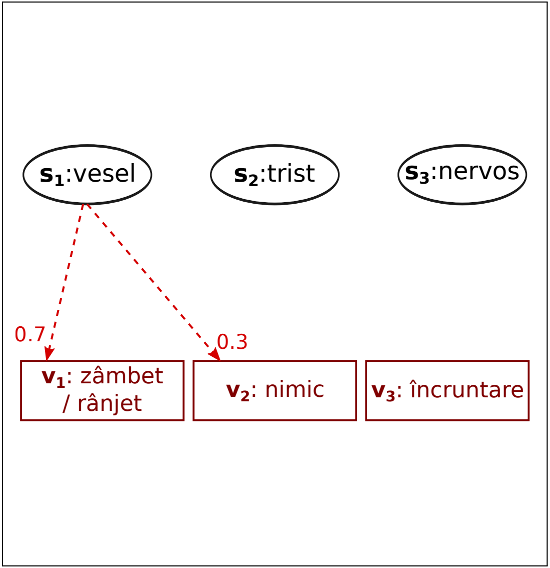
M - numărul de valori observabile distincte

$\mathbf{M} = 3$

valori observabile:

- $\mathbf{v_1}$: zâmbet / rânjet
- $\mathbf{v_2}$: nimic
- $\mathbf{v_3}$: încruntare

Exemplu: Urmărirea stărilor emoționale



B - matricea distribuțiilor de probabilitate ale valorilor observabile

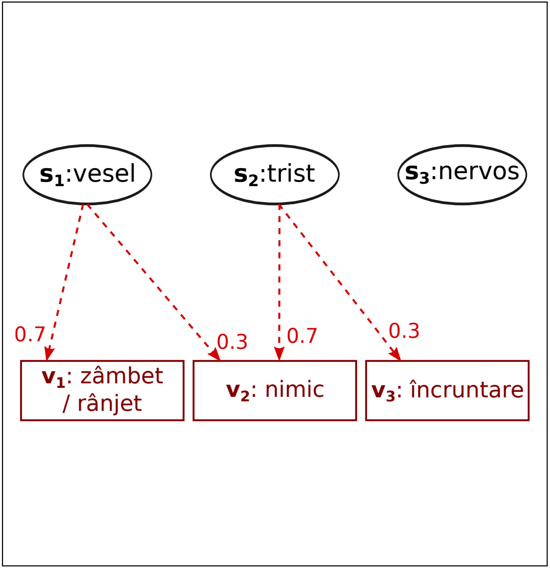
$$\mathbf{B} = \{b_{j,k}\}_{1 \le j \le N, 1 \le k, \le M}$$

$$\begin{aligned} b_{j,k} &= b_j(v_k) \\ &= P(o_t = v_k | q_t = s_j) \end{aligned}$$

$$\sum_{k=1}^{M} b_{j,k} = 1, \quad 1 \le j \le N$$

$$\mathbf{B} = \begin{array}{c c} & \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} 0.7 & 0.3 & 0 \\ & & \\ & & \end{pmatrix} \end{array}$$

Exemplu: Urmărirea stărilor emoționale



B - matricea distribuțiilor de probabilitate ale valorilor observabile

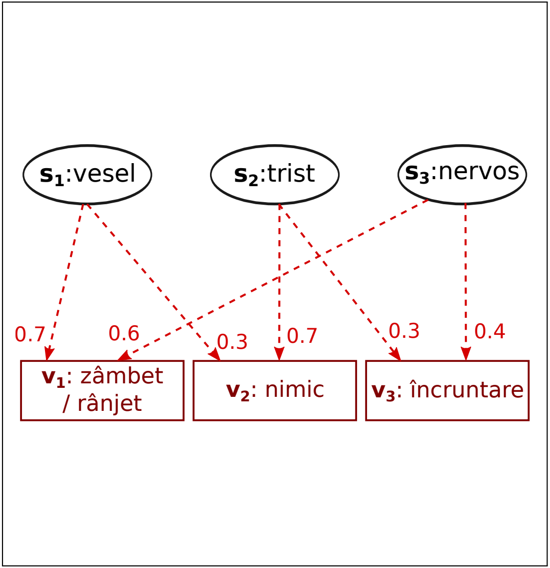
$$\mathbf{B} = \{b_{j,k}\}_{1 \le j \le N, 1 \le k, \le M}$$

$$\begin{aligned} b_{j,k} &= b_j(v_k) \\ &= P(o_t = v_k | q_t = s_j) \end{aligned}$$

$$\boxed{\sum_{k=1}^{M} b_{j,k} = 1, \quad 1 \le j \le N}$$

$$\mathbf{B} = \begin{array}{c c} & \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} 0.7 & 0.3 & 0 \\ 0 & 0.7 & 0.3 \\ & & \end{pmatrix} \end{array}$$

Exemplu: Urmărirea stărilor emoționale



B - matricea distribuțiilor de probabilitate ale valorilor observabile

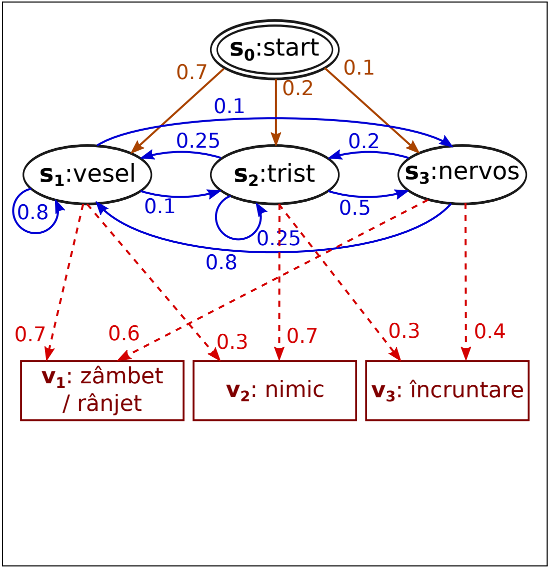
$$\mathbf{B} = \{b_{j,k}\}_{1 \le j \le N, 1 \le k, \le M}$$

$$\begin{aligned} b_{j,k} &= b_j(v_k) \\ &= P(o_t = v_k | q_t = s_j) \end{aligned}$$

$$\sum_{k=1}^{M} b_{j,k} = 1, \quad 1 \le j \le N$$

$$\mathbf{B} = \begin{array}{c c} & \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} 0.7 & 0.3 & 0 \\ 0 & 0.7 & 0.3 \\ 0.6 & 0 & 0.4 \end{pmatrix} \end{array}$$

Exemplu: Urmărirea stărilor emoționale



λ - parametrii Modelului Markov Ascuns

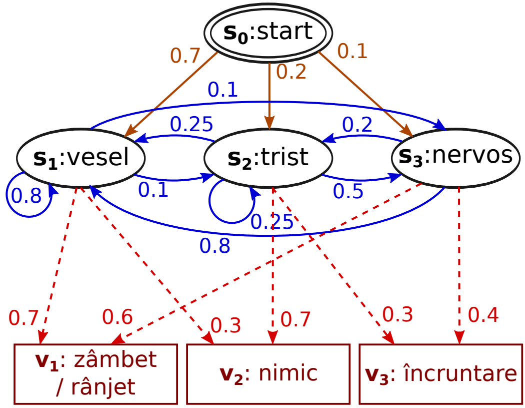
$\lambda = (A, B, \Pi)$

A - matricea distribuțiilor de probabilitate ale tranzițiilor între stări

B - matricea distribuțiilor de probabilitate ale valorilor observabile

Π - distribuția stării inițiale

Exemplu: Urmărirea stărilor emoționale



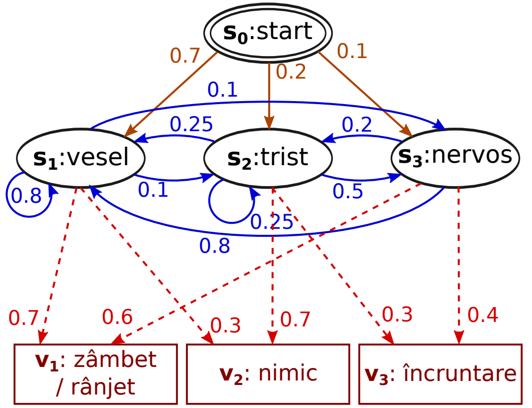
O = [**o$_1$**:zâmbet **o$_2$**:nimic **o$_3$**:nimic **o$_4$**:încruntare **o$_5$**:zâmbet]

O - secvența de observații

T - lungimea secvenței de observații

$$O = [o_1 o_2 \cdots o_T]$$

Exemplu: Urmărirea stărilor emoționale



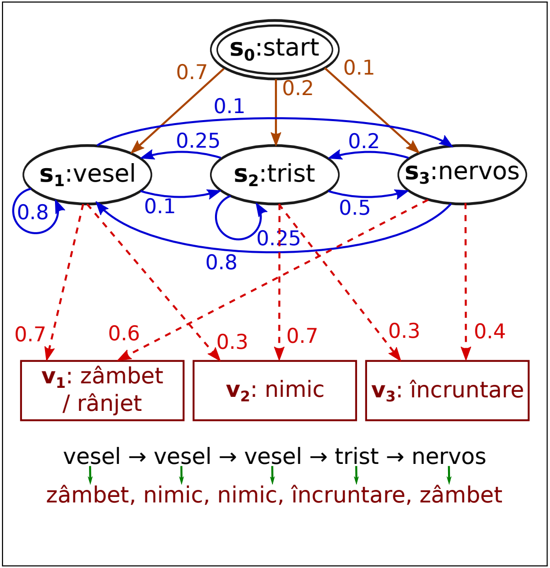
zâmbet, nimic, nimic, încruntare, zâmbet

O - secvența de observații

T - lungimea secvenței de observații

$$O = [o_1 o_2 \cdots o_T]$$

Exemplu: Urmărirea stărilor emoționale

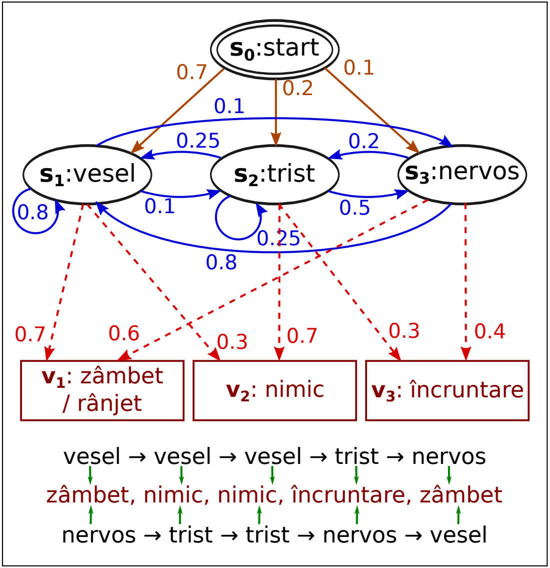


O - secvența de observații

T - lungimea secvenței de observații

$$O = [o_1 o_2 \cdots o_T]$$

Exemplu: Urmărirea stărilor emoționale



O - secvența de observații

T - lungimea secvenței de observații

$$O = [o_1 o_2 \cdots o_T]$$

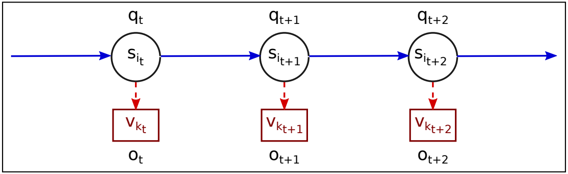
Modele Markov Ascunse

Definiție

Un Model Markov Ascuns este un tuplu $\langle S, V, A, B, P \rangle$:

- S - mulțimea stărilor
- V - mulțimea valorilor observabile
- A - matricea de tranziție
- B - matricea de emisie
- Π - matricea distribuției stării inițiale

- Notație: $\lambda = (A, B, \Pi)$ - parametrii modelului



Reformularea problemelor fundamentale ale MMA

Problema evaluării

Date fiind un model și o secvență de observații , cum calculăm probabilitatea ca secvența de observații să fi fost generată de acel model?

Reformularea problemelor fundamentale ale MMA

Problema evaluării

Date fiind un model $\lambda = (A, B, \Pi)$ și o secvență de observații $O = [o_1 o_2 \cdots o_T]$, cum calculăm probabilitatea $P(O|\lambda)$ ca secvența de observații să fi fost generată de acel model?

Reformularea problemelor fundamentale ale MMA

Problema evaluării

Date fiind un model $\lambda = (A, B, \Pi)$ și o secvență de observații $O = [o_1 o_2 \cdots o_T]$, cum calculăm probabilitatea $P(O|\lambda)$ ca secvența de observații să fi fost generată de acel model?

- Prin enumerarea tuturor secvențelor posibile de stări:

$$P(O|\lambda) = \sum_{\text{all } Q} P(O|Q,\lambda) \cdot P(Q|\lambda) \tag{1}$$

(legea probabilității totale)

Reformularea problemelor fundamentale ale MMA

$$P(O|\lambda) = \sum_{\text{all } Q} P(O|Q, \lambda) \cdot P(Q|\lambda) \tag{1}$$

Reformularea problemelor fundamentale ale MMA

$$P(O|\lambda) = \sum_{\text{all } Q} P(O|Q,\lambda) \cdot P(Q|\lambda) \tag{1}$$

Primul factor (independență condițională):

$$P(O|Q,\lambda) = \prod_{t=1}^{T} P(o_t|q_t,\lambda) = \prod_{t=1}^{T} b_{q_t}(o_t) = b_{q_1}(o_1) \cdot \ldots \cdot b_{q_T}(o_T) \tag{2}$$

Reformularea problemelor fundamentale ale MMA

$$P(O|\lambda) = \sum_{\text{all } Q} P(O|Q,\lambda) \cdot P(Q|\lambda) \quad (1)$$

Primul factor (independență condițională):

$$P(O|Q,\lambda) = \prod_{t=1}^{T} P(o_t|q_t,\lambda) = \prod_{t=1}^{T} b_{q_t}(o_t) = b_{q_1}(o_1) \cdot \ldots \cdot b_{q_T}(o_T) \quad (2)$$

Al doilea factor (presupunerea Markov):

$$P(Q|\lambda) = \pi_{q_1} \prod_{t=2}^{T} a_{q_{t-1},q_t} = \pi_{q_1} \cdot a_{q_1,q_2} \cdot a_{q_2,q_3} \cdot \ldots \cdot a_{q_{T-1},q_T} \quad (3)$$

Reformularea problemelor fundamentale ale MMA

$$P(O|\lambda) = \sum_{\text{all } Q} P(O|Q,\lambda) \cdot P(Q|\lambda) \quad (1)$$

Primul factor (independență condițională):

$$P(O|Q,\lambda) = \prod_{t=1}^{T} P(o_t|q_t,\lambda) = \prod_{t=1}^{T} b_{q_t}(o_t) = b_{q_1}(o_1) \cdot \ldots \cdot b_{q_T}(o_T) \quad (2)$$

Al doilea factor (presupunerea Markov):

$$P(Q|\lambda) = \pi_{q_1} \prod_{t=2}^{T} a_{q_{t-1},q_t} = \pi_{q_1} \cdot a_{q_1,q_2} \cdot a_{q_2,q_3} \cdot \ldots \cdot a_{q_{T-1},q_T} \quad (3)$$

$$P(O|\lambda) = \sum_{\text{all } Q} \left(\pi_{q_1} \cdot b_{q_1}(o_1) \cdot \prod_{t=2}^{T} b_{q_t}(o_t) a_{q_{t-1},q_t} \right) \quad (1)$$

Reformularea problemelor fundamentale ale MMA

Problema interpretării unei secvențe de observații

Date fiind un model și o secvență de observații , cum alegem o secvență corespunzătoare de stări care *să dea un înțeles* observațiilor? Cum *descoperim* partea ascunsă a modelului?

Reformularea problemelor fundamentale ale MMA

Problema interpretării unei secvențe de observații

Date fiind un model $\lambda = (A, B, \Pi)$ și o secvență de observații , cum alegem o secvență corespunzătoare de stări care *să dea un înțeles* observațiilor? Cum *descoperim* partea ascunsă a modelului?

Reformularea problemelor fundamentale ale MMA

Problema interpretării unei secvențe de observații

Date fiind un model $\lambda = (A, B, \Pi)$ și o secvență de observații $O = [o_1 o_2 \cdots o_T]$, cum alegem o secvență corespunzătoare de stări care *să dea un înțeles* observațiilor? Cum *descoperim* partea ascunsă a modelului?

Reformularea problemelor fundamentale ale MMA

Problema interpretării unei secvențe de observații

Date fiind un model $\lambda = (A, B, \Pi)$ și o secvență de observații $O = [o_1 o_2 \cdots o_T]$, cum alegem o secvență corespunzătoare de stări $Q = [q_1 q_2 \cdots q_T]$ care *să dea un înțeles* observațiilor? Cum *descoperim* partea ascunsă a modelului?

Reformularea problemelor fundamentale ale MMA

Problema interpretării unei secvențe de observații

Date fiind un model $\lambda = (A, B, \Pi)$ și o secvență de observații $O = [o_1 o_2 \cdots o_T]$, cum alegem o secvență corespunzătoare de stări $Q = [q_1 q_2 \cdots q_T]$ care *să dea un înțeles* observațiilor? Cum *descoperim* partea ascunsă a modelului?

- Există mai multe criterii pentru *cea mai bună* sevență
  - Secvența celor mai probabile stări (luate individual):

$$Q_{\text{best}} = [\hat{q}_1 \ \hat{q}_2 \ \ldots \hat{q}_T], \quad \hat{q}_t = \underset{s_i}{argmax}\ P(q_t = s_i | O, \lambda) \tag{4}$$

Reformularea problemelor fundamentale ale MMA

Problema interpretării unei secvențe de observații

Date fiind un model $\lambda = (A, B, \Pi)$ și o secvență de observații $O = [o_1 o_2 \cdots o_T]$, cum alegem o secvență corespunzătoare de stări $Q = [q_1 q_2 \cdots q_T]$ care *să dea un înțeles* observațiilor? Cum *descoperim* partea ascunsă a modelului?

- Există mai multe criterii pentru *cea mai bună* seventă
  - Secvența celor mai probabile stări (luate individual):

$$Q_{\text{best}} = [\hat{q}_1 \ \hat{q}_2 \ \ldots \hat{q}_T], \quad \hat{q}_t = \underset{s_i}{argmax}\ P(q_t = s_i | O, \lambda) \tag{4}$$

  - Cea mai bună *cale* (de dimensiune T)

$$Q_{\text{best}} = \underset{Q}{\operatorname{argmax}}\ P(Q|O, \lambda) = \underset{Q}{\operatorname{argmax}}\ P(Q, O|\lambda) \tag{5}$$

Reformularea problemelor fundamentale ale MMA

Problema Estimării Modelului (Învățării)

Date fiind niște secvențe de observații , cum *ajustăm* parametrii ai unui MMA astfel încât să explice cel mai bine observațiile?

Reformularea problemelor fundamentale ale MMA

Problema Estimării Modelului (Învățării)

Date fiind niște secvențe de observații $\mathcal{O} = [O_1 O_2 \cdots O_L]$, cum *ajustăm* parametrii ai unui MMA astfel încât să explice cel mai bine observațiile?

Reformularea problemelor fundamentale ale MMA

Problema Estimării Modelului (Învățării)

Date fiind niște secvențe de observații $\mathcal{O} = [O_1 O_2 \cdots O_L]$, cum *ajustăm* parametrii $\lambda = (A, B, \Pi)$ ai unui MMA astfel încât să explice cel mai bine observațiile?

Reformularea problemelor fundamentale ale MMA

Problema Estimării Modelului (Învățării)

Date fiind niște secvențe de observații $\mathcal{O} = [O_1 O_2 \cdots O_L]$, cum *ajustăm* parametrii $\lambda = (A, B, \Pi)$ ai unui MMA astfel încât să explice cel mai bine observațiile?

- Întrebarea se poate reformula matematic:

$$\lambda_{\text{best}} = \underset{\lambda}{\text{argmax}}\, P(\mathcal{O}|\lambda) \tag{6}$$

De înlocuit

:)

Outline

1. Aplicații în Învățarea Automată pentru MMA
   - Învățarea Automată
   - MMA în Învățarea Automată
2. Teoria MMA
   - Cele Trei Probleme ale MMA
   - Fundamente Matematice
3. Implementarea MMA
   - Problema Evaluării: Algoritmul Forward-Backward
   - Problema Interpretării: Algoritmul Viterbi
   - Problema Estimării: Algoritmul Baum-Welch
4. Demo: Recunoașterea Simbolurilor
5. Tipuri de MMA
6. Discuții și Concluzii

Algoritmul Forward-Backward

Problema Evaluării

Date fiind un model $\lambda = (A, B, \Pi)$ și o secvență de observații $O = [o_1 o_2 \dots o_T]$, care este probabilitatea ca secvența de observații să fi fost produsă de acel model?

$$P(O|\lambda) =? \tag{7}$$

Algoritmul Forward-Backward

Problema Evaluării

Date fiind un model $\lambda = (A, B, \Pi)$ și o secvență de observații $O = [o_1 o_2 \dots o_T]$, care este probabilitatea ca secvența de observații să fi fost produsă de acel model?

$$P(O|\lambda) = ? \tag{7}$$

Rezolvare

$P(O|\lambda)$ se calculează *eficient* cu algoritmul Forward - Backward.

Calculul se face cu ajutorul unor variabile auxiliare α și β.

Variabilele α - Motivație

- Calcul conform legii probabilităților totale:

$$P(O|\lambda) = \sum_{all\ Q} P(O, Q|\lambda) = \sum_{all\ Q} P(O, |Q, \lambda) \cdot P(Q, \lambda)$$
$$= \sum_{\text{all } Q} \left(\pi_{q_1} \cdot b_{q_1}(o_1) \cdot \prod_{t=2}^{T} b_{q_t}(o_t) a_{q_{t-1}, q_t} \right) \tag{1}$$

- Pentru secvențele de stări care încep cu o secvență comună $[q_1 q_2 \ldots q_Z]$, următorii factori sunt comuni:

$$\pi_{q_1} \cdot b_{q_1}(o_1) \cdot \prod_{z=2}^{Z} b_{q_z}(o_z) a_{q_{z-1}, q_z}$$

  - Calculul de $(T - Z)^N$ ori ar fi redundant!

Variabilele α - Definiție

Definim variabilele α astfel:

$$\alpha_{t,i} = P(o_1, o_2, \ldots, o_t, q_t = s_i|\lambda)_{1 \le t \le T, 1 \le i \le N} \tag{8}$$

(probabiliatea de a fi observat primele t valori observate ajungând la momentul t în starea s_i, dați fiind parametrii λ)

Variabilele α - Definiție

Definim variabilele α astfel:

$$\alpha_{t,i} = P(o_1, o_2, \ldots, o_t, q_t = s_i|\lambda) \quad \underset{1\leq t\leq T, 1\leq i\leq N}{} \tag{8}$$

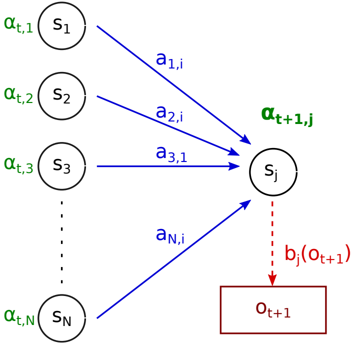
(probabiliatea de a fi observat primele t valori observate ajungând la momentul t în starea s_i, dați fiind parametrii λ)

Relația dintre $P(O|\lambda)$ și variabilele α este:

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_{T,i} \tag{9}$$

(conform legii probabilităților totale)

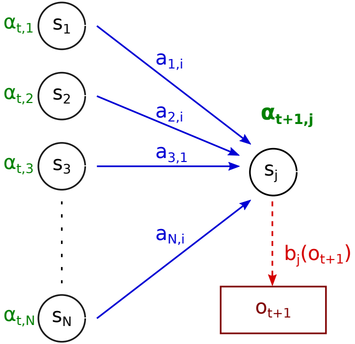
Calculul variabilelor α



- Inițializare ($t = 1$)

$$\alpha_{1,i} = \pi_i b_i(o_1), \quad 1 \leq i \leq N$$

Calculul variabilelor α



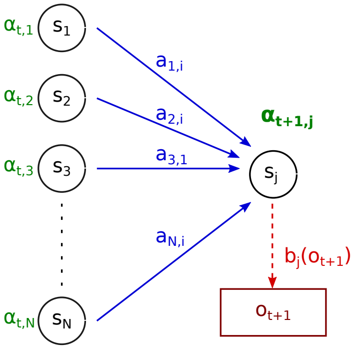
- Inițializare $(t = 1)$

$$\alpha_{1,i} = \pi_i b_i(o_1), \quad 1 \leq i \leq N$$

- Inducție $(t > 1)$

$$\alpha_{t+1,j} = \left[\sum_{i=1}^{N} \alpha_{t,j} a_{i,j} \right] b_j(o_{t+1}), \quad \substack{1 \leq t \leq T-1 \\ 1 \leq j \leq N}$$

Calculul variabilelor α



- Inițializare ($t = 1$)

$$\alpha_{1,i} = \pi_i b_i(o_1), \quad 1 \leq i \leq N$$

- Inducție ($t > 1$)

$$\alpha_{t+1,j} = \left[\sum_{i=1}^{N} \alpha_{t,j} a_{i,j}\right] b_j(o_{t+1}), \quad \substack{1 \leq t \leq T-1 \\ 1 \leq j \leq N}$$

- Probabilitatea secvenței observate:

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_{T,i}$$

Variabilele β - Definiție

Definim variabilele β astfel:

$$\beta_{t,i} = P(o_{t+1}o_{t+2}\cdots o_T | q_t = s_i, \lambda) \tag{10}$$

(probabiliatea de a fi observate valorile secvenței de la $t+1$ la T, condiționată de aflarea la momentul t în starea s_i și de parametrii λ)

Variabilele β - Definiție

Definim variabilele β astfel:

$$\beta_{t,i} = P(o_{t+1}o_{t+2}\cdots o_T|q_t = s_i, \lambda) \tag{10}$$

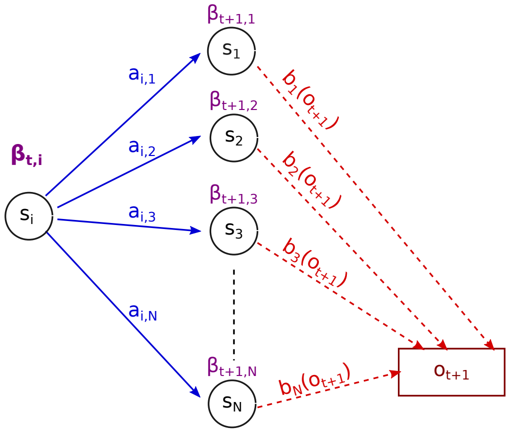
(probabiliatea de a fi observate valorile secvenței de la $t+1$ la T, condiționată de aflarea la momentul t în starea s_i și de parametrii λ)

Relația dintre $P(O|\lambda)$ și variabilele β este:

$$P(O|\lambda) = \sum_{i=1}^{N} \beta_{1,i} \tag{11}$$

(conform legii probabilităților totale)

Calculul Variabilelor β



- Inițializare ($t = T$)

$$\beta_{T,i} = 1, \quad 1 \leq i \leq N \tag{12}$$

- Pas de inducție ($t < T$)

$$\beta_{t,i} = \sum_{j=1}^{N} a_{i,j} b_j(o_{t+1}) \beta_{t+1,j}, \quad t = T-1, T-2, \ldots, 1, 1 \leq i \leq N$$

Probleme numerice

- să revedem definiția $P(O|\lambda)$:

$$P(O|\lambda) = \sum_{\text{all } Q} \left(\pi_{q_1} \cdot b_{q_1}(o_1) \cdot \prod_{t=2}^{T} b_{q_t}(o_t) a_{q_{t-1},q_t} \right)$$

Probleme numerice

- să revedem definiția $P(O|\lambda)$:

$$P(O|\lambda) = \sum_{\text{all } Q} \left(\pi_{q_1} \cdot b_{q_1}(o_1) \cdot \prod_{t=2}^{T} b_{q_t}(o_t) a_{q_{t-1},q_t} \right)$$

- toți cei $2T$ factori ai unui produs sunt subunitari
- pentru secvențe mari, produsele se apropie mult de zero
- se depășește precizia disponibilă pentru reprezentare
- trebuie introdus un mecanism de scalare

Algoritmul Forward-Backward cu scalare

- $\hat{\alpha}_{t,i}$ - variabilele α scalate
- $\hat{\beta}_{t,i}$ - variabilele β scalate

- C_t - coeficienții de scalare

- Variabilele α scalate
$$\hat{\alpha}_{t,i} = C_t \cdot \alpha_{t,i} \tag{13}$$

- Variabilele β scalate:
$$\hat{\beta}_{t,j} = C_t \cdot \beta_{t,j} \tag{14}$$

(notațiile sunt adoptate din literatură)

Calculul variabilelor α scalate

- Inițializare $(t = 1)$

  $1 \leq i \leq N$

$$\ddot{\alpha}_{1,i} = \alpha_{1,i} \tag{15}$$

$$c_1 = \frac{1}{\sum_{i=1}^{N} \ddot{\alpha}_{1,i}} \tag{16}$$

$$\hat{\alpha}_{1,i} = c_1 \cdot \ddot{\alpha}_{1,i} \tag{17}$$

Calculul variabilelor α scalate

- Inițializare ($t = 1$)

  $1 \leq i \leq N$

$$\ddot{\alpha}_{1,i} = \alpha_{1,i} \tag{15}$$

$$c_1 = \frac{1}{\sum_{i=1}^{N} \ddot{\alpha}_{1,i}} \tag{16}$$

$$\hat{\alpha}_{1,i} = c_1 \cdot \ddot{\alpha}_{1,i} \tag{17}$$

- Pas de inducție ($t > 1$)

  $1 \leq i \leq N$

$$\ddot{\alpha}_{t+1,i} = \left[\sum_{i=1}^{N} \hat{\alpha}_{t,i} a_{i,j} \right] b_j(o_{t+1}) \tag{18}$$

$$c_{t+1} = \frac{1}{\sum_{i=1}^{N} \ddot{\alpha}_{t+1,i}} \tag{19}$$

$$\hat{\alpha}_{t+1,i} = c_{t+1} \cdot \ddot{\alpha}_{t+1,i} \tag{20}$$

Coeficienții de scalare

Vom nota: $C_t = c_1 \cdot c_2 \cdot \ldots \cdot c_t$

Ipoteza: $\hat{\alpha}_{t,i} = C_t \alpha_{t,i}$

Coeficienții de scalare

Vom nota: $C_t = c_1 \cdot c_2 \cdot \ldots \cdot c_t$

Ipoteza: $\hat{\alpha}_{t,i} = C_t \alpha_{t,i}$

Demonstrație prin inducție:

- $(t = 1)$:

$$\hat{\alpha}_{1,i} = c_1 \cdot \ddot{\alpha}_{1,i} = c_1 \cdot \alpha_{1,i}$$

Coeficienții de scalare

Vom nota: $C_t = c_1 \cdot c_2 \cdot \ldots \cdot c_t$

Ipoteza: $\hat{\alpha}_{t,i} = C_t \alpha_{t,i}$

Demonstrație prin inducție:

- $(t = 1)$:

$$\hat{\alpha}_{1,i} = c_1 \cdot \ddot{\alpha}_{1,i} = c_1 \cdot \alpha_{1,i}$$

- $(t > 1)$: Presupunem adevărat: $\hat{\alpha}_{t,i} = C_t \alpha_{t,i}$

$$\ddot{\alpha}_{t+1,i} = \left[\sum_{i=1}^{N} \hat{\alpha}_{t,i} a_{i,j}\right] b_j(o_{t+1}) = \left[\sum_{i=1}^{N} C_t \alpha_{t,i} a_{i,j}\right] b_j(o_{t+1}) = C_t \alpha_{t+1,i}$$

$$\hat{\alpha}_{t+1,i} = c_{t+1} \cdot \ddot{\alpha}_{t+1,i} = c_{t+1} \cdot C_t \cdot \alpha_{t+1,i} = C_{t+1} \cdot \alpha_{t+1,i}$$

Coeficienții de scalare

Vom nota: $C_t = c_1 \cdot c_2 \cdot \ldots \cdot c_t$

Ipoteza: $\hat{\alpha}_{t,i} = C_t \alpha_{t,i}$

Atunci: $\mathbf{P}(\mathbf{O}|\lambda) = \sum_{i=1}^{N} \alpha_{T,i} = C_T^{-1} \cdot \sum_{i=1}^{N} \hat{\alpha}_{T,i} = C_T^{-1} = \prod_{\mathbf{t=1}}^{\mathbf{T}} \mathbf{c_t^{-1}}$

Coeficienții de scalare

Vom nota: $C_t = c_1 \cdot c_2 \cdot \ldots \cdot c_t$

Ipoteza: $\hat{\alpha}_{t,i} = C_t \alpha_{t,i}$

Atunci: $\mathbf{P}(\mathbf{O}|\lambda) = \sum_{i=1}^{N} \alpha_{T,i} = C_T^{-1} \cdot \sum_{i=1}^{N} \hat{\alpha}_{T,i} = C_T^{-1} = \prod_{\mathbf{t}=\mathbf{1}}^{\mathbf{T}} \mathbf{c}_\mathbf{t}^{-\mathbf{1}}$

Calculăm $P(O|\lambda)$?

Coeficienții de scalare

Vom nota: $C_t = c_1 \cdot c_2 \cdot \ldots \cdot c_t$

Ipoteza: $\hat{\alpha}_{t,i} = C_t \alpha_{t,i}$

Atunci: $\mathbf{P}(\mathbf{O}|\lambda) = \sum_{i=1}^{N} \alpha_{T,i} = C_T^{-1} \cdot \sum_{i=1}^{N} \hat{\alpha}_{T,i} = C_T^{-1} = \prod_{\mathbf{t=1}}^{\mathbf{T}} \mathbf{c_t^{-1}}$

Calculăm $P(O|\lambda)$? **NU**

Coeficienții de scalare

Vom nota: $C_t = c_1 \cdot c_2 \cdot \ldots \cdot c_t$

Ipoteza: $\hat{\alpha}_{t,i} = C_t \alpha_{t,i}$

Atunci: $\mathbf{P}(\mathbf{O}|\lambda) = \sum_{i=1}^{N} \alpha_{T,i} = C_T^{-1} \cdot \sum_{i=1}^{N} \hat{\alpha}_{T,i} = C_T^{-1} = \prod_{\mathbf{t=1}}^{\mathbf{T}} \mathbf{c_t^{-1}}$

Calculăm $P(O|\lambda)$? **NU**

Calculăm:

$$log(P(O|\lambda)) = log(\prod_{t=1}^{T} c_t^{-1}) = \sum_{t=1}^{T} log(c_t^{-1}) = -\sum_{t=1}^{T} log(c_t) \qquad (21)$$

Calculul variabilelor β scalate

- Inițializare
  $(t = T)$

  $1 \leq i \leq N$

$$\ddot{\beta}_{T,i} = \beta_{T,i} = 1 \tag{22}$$

$$\hat{\beta}_{T,i} = c_T \cdot \ddot{\beta}_{T,i} \tag{23}$$

Calculul variabilelor β scalate

- Inițializare
  $(t = T)$
  $1 \leq i \leq N$

$$\ddot{\beta}_{T,i} = \beta_{T,i} = 1 \tag{22}$$

$$\hat{\beta}_{T,i} = c_T \cdot \ddot{\beta}_{T,i} \tag{23}$$

- Pas de inducție
  $(t < T)$
  $1 \leq i \leq N$

$$\ddot{\beta}_{t,i} = \sum_{j=1}^{N} a_{i,j} b_j(o_{t+1}) \hat{\beta}_{t+1,i} \tag{24}$$

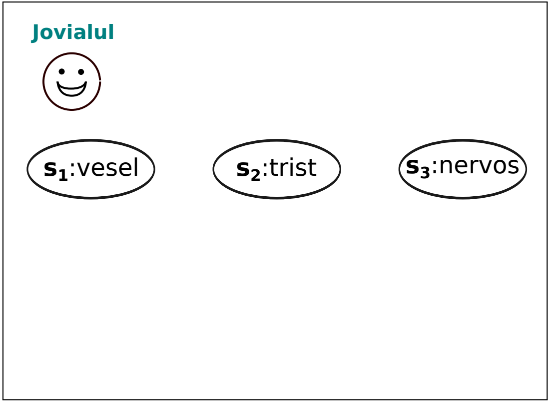
$$\hat{\beta}_{t,i} = c_t \cdot \ddot{\beta}_{t,i} \tag{25}$$

- Se demonstrează: $\hat{\beta}_{t,i} = c_t \cdot c_{t+1} \cdot \ldots \cdot c_T \cdot \beta_{t,i}$

Exemplu de aplicație

- Să reluăm exemplul de mai devreme cu robotul.
- Robotul acționează diferit în fața a 2 tipuri de personalitate:
  - *jovialul*
  - *morocănosul*.
- Pentru fiecare există un set de parametri:
  - $\lambda^1 = (A^1, B^1, \Pi^1)$
  - $\lambda^2 = (A^2, B^2, \Pi^2)$
- Robotul recunoaște următoarele gesturi:
  (zâmbet | rânjet) $\longrightarrow$ (zâmbet | rânjet) $\longrightarrow$ încruntare
- **Întrebare:** Cărei personalități este mai probabil să-i aparțină cel monitorizat?

Jovialul



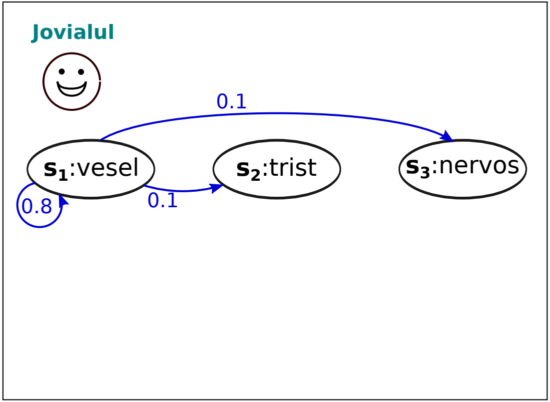
$$\lambda^1 = (A^1, B^1, \Pi^1)$$

$$A^1 = \begin{array}{c|ccc} & s_1 & s_2 & s_3 \\ \hline s_1 & & & \\ s_2 & & & \\ s_3 & & & \end{array}$$

$$\Pi^1 = \begin{array}{ccc} s_1 & s_2 & s_3 \\ (& &) \end{array}$$

$$B^1 = \begin{array}{c|ccc} & v_1 & v_2 & v_3 \\ \hline s_1 & & & \\ s_2 & & & \\ s_3 & & & \end{array}$$

Jovialul



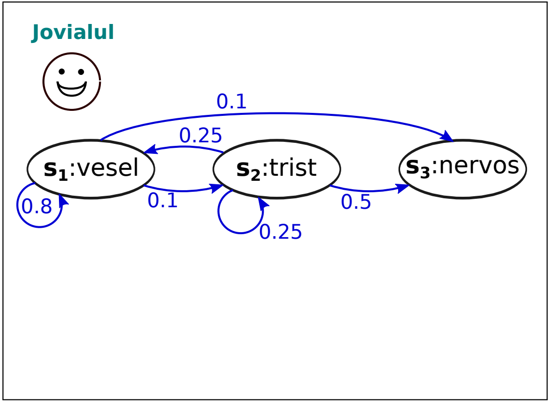
$$\lambda^1 = (A^1, B^1, \Pi^1)$$

$$A^1 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ & & \\ & & \end{pmatrix} \end{array}$$

$$\Pi^1 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} & & \end{pmatrix} \end{array}$$

$$B^1 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{pmatrix} & & \\ & & \\ & & \end{pmatrix} \end{array}$$

Jovialul



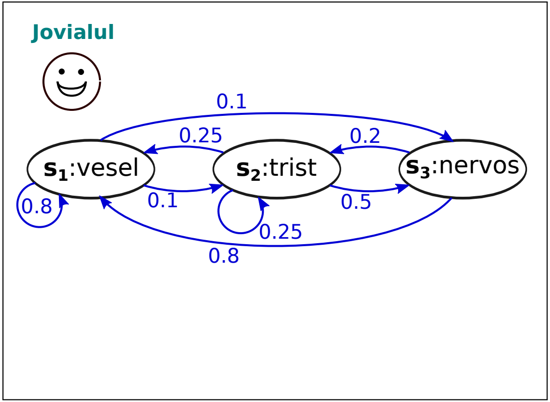
$$\lambda^1 = (A^1, B^1, \Pi^1)$$

$$A^1 = \begin{array}{c c} & \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.25 & 0.25 & 0.5 \\ & & \end{pmatrix} \end{array}$$

$$\Pi^1 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} & & \end{pmatrix} \end{array}$$

$$B^1 = \begin{array}{c c} & \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} & & \\ & & \\ & & \end{pmatrix} \end{array}$$

Jovialul



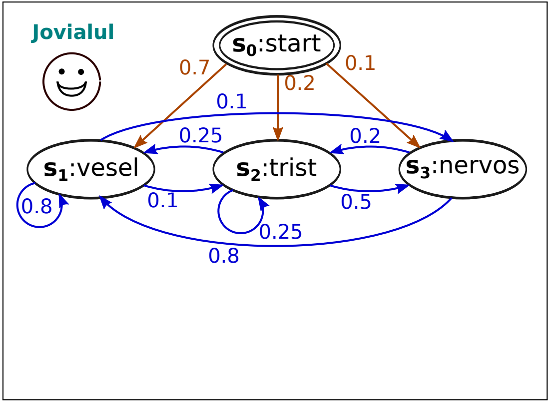
$$\lambda^1 = (A^1, B^1, \Pi^1)$$

$$A^1 = \begin{array}{c c} & \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.25 & 0.25 & 0.5 \\ 0.8 & 0.2 & 0 \end{pmatrix} \end{array}$$

$$\Pi^1 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} \quad & \quad & \quad \end{pmatrix} \end{array}$$

$$B^1 = \begin{array}{c c} & \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} \quad & \quad & \quad \\ & & \\ & & \end{pmatrix} \end{array}$$

Jovialul



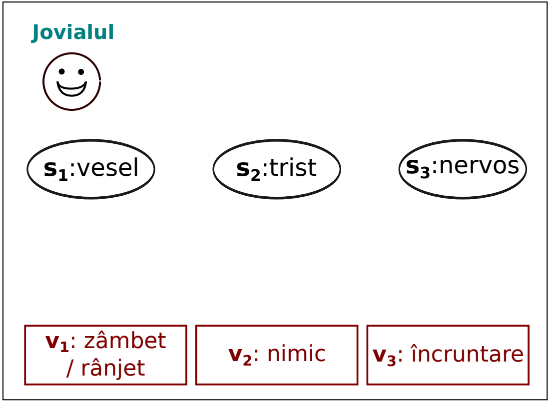
$$\lambda^1 = (A^1, B^1, \Pi^1)$$

$$A^1 = \begin{array}{c c} & \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.25 & 0.25 & 0.5 \\ 0.8 & 0.2 & 0 \end{pmatrix} \end{array}$$

$$\Pi^1 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.7 & 0.2 & 0.1 \end{pmatrix} \end{array}$$

$$B^1 = \begin{array}{c c} & \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} & & \\ & & \\ & & \end{pmatrix} \end{array}$$

Jovialul



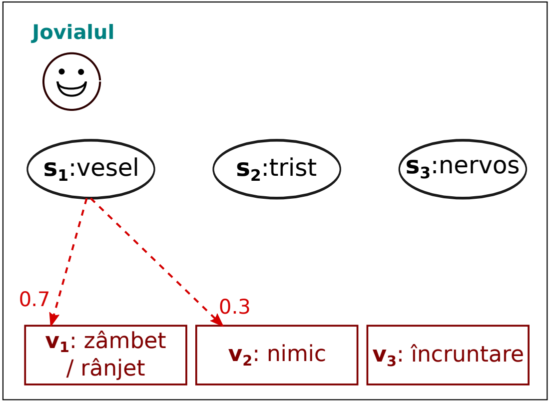
$$\lambda^1 = (A^1, B^1, \Pi^1)$$

$$A^1 = \begin{array}{c c} & \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.25 & 0.25 & 0.5 \\ 0.8 & 0.2 & 0 \end{pmatrix} \end{array}$$

$$\Pi^1 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.7 & 0.2 & 0.1 \end{pmatrix} \end{array}$$

$$B^1 = \begin{array}{c c} & \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} \quad & \quad & \quad \\ & & \\ & & \end{pmatrix} \end{array}$$

Jovialul



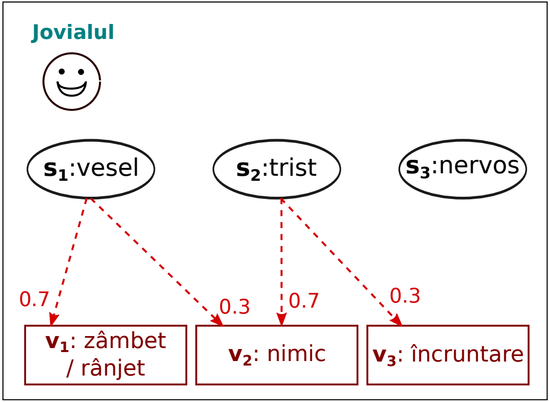
$$\lambda^1 = (A^1, B^1, \Pi^1)$$

$$A^1 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.25 & 0.25 & 0.5 \\ 0.8 & 0.2 & 0 \end{pmatrix} \end{array}$$

$$\Pi^1 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.7 & 0.2 & 0.1 \end{pmatrix} \end{array}$$

$$B^1 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{pmatrix} 0.7 & 0.3 & 0 \\ & & \\ & & \end{pmatrix} \end{array}$$

Jovialul



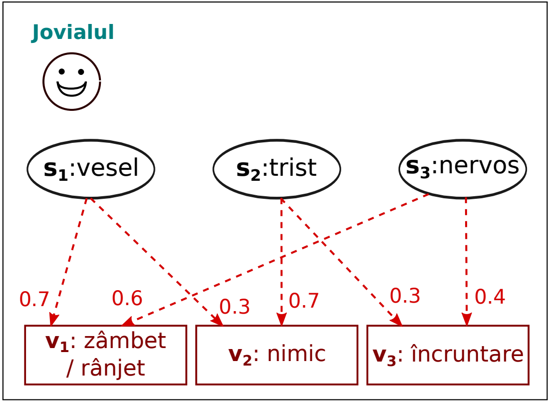
$$\lambda^1 = (A^1, B^1, \Pi^1)$$

$$A^1 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.25 & 0.25 & 0.5 \\ 0.8 & 0.2 & 0 \end{pmatrix} \end{array}$$

$$\Pi^1 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.7 & 0.2 & 0.1 \end{pmatrix} \end{array}$$

$$B^1 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{pmatrix} 0.7 & 0.3 & 0 \\ 0 & 0.7 & 0.3 \\ & & \end{pmatrix} \end{array}$$

Jovialul



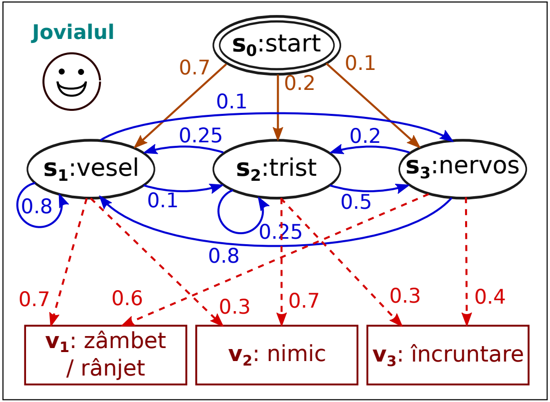
$$\lambda^1 = (A^1, B^1, \Pi^1)$$

$$A^1 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.25 & 0.25 & 0.5 \\ 0.8 & 0.2 & 0 \end{pmatrix} \end{array}$$

$$\Pi^1 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.7 & 0.2 & 0.1 \end{pmatrix} \end{array}$$

$$B^1 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{pmatrix} 0.7 & 0.3 & 0 \\ 0 & 0.7 & 0.3 \\ 0.6 & 0 & 0.4 \end{pmatrix} \end{array}$$

Jovialul



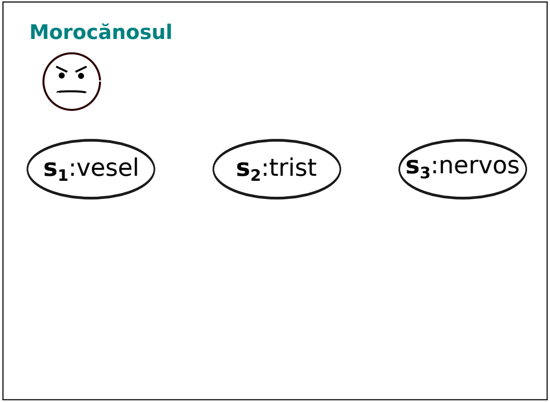
$$\lambda^1 = (A^1, B^1, \Pi^1)$$

$$A^1 = \begin{array}{cc} & \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.25 & 0.25 & 0.5 \\ 0.8 & 0.2 & 0 \end{pmatrix} \end{array}$$

$$\Pi^1 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.7 & 0.2 & 0.1 \end{pmatrix} \end{array}$$

$$B^1 = \begin{array}{cc} & \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} 0.7 & 0.3 & 0 \\ 0 & 0.7 & 0.3 \\ 0.6 & 0 & 0.4 \end{pmatrix} \end{array}$$

Morocănosul



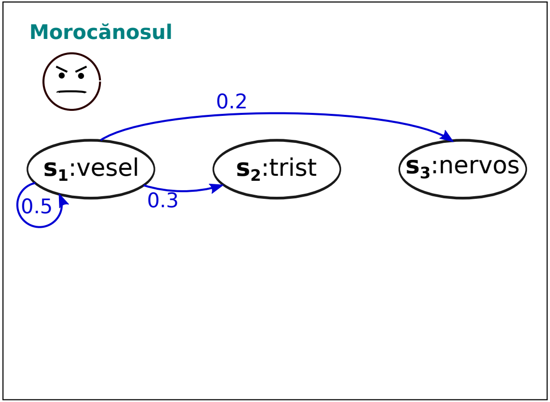
$$\lambda^2 = (A^2, B^2, \Pi^2)$$

$$A^2 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \left(\begin{array}{ccc} \quad & \quad & \quad \\ & & \\ & & \end{array} \right) \end{array}$$

$$\Pi^2 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \left(\begin{array}{ccc} \quad & \quad & \quad \end{array} \right) \end{array}$$

$$B^2 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \left(\begin{array}{ccc} \quad & \quad & \quad \\ & & \\ & & \end{array} \right) \end{array}$$

Morocănosul



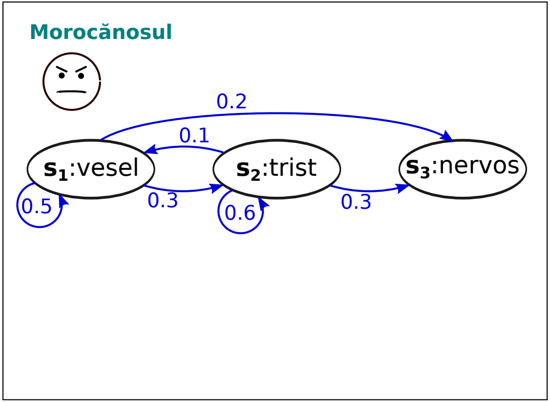
$$\lambda^2 = (A^2, B^2, \Pi^2)$$

$$A^2 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ & & \\ & & \end{pmatrix} \end{array}$$

$$\Pi^2 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} \quad & \quad & \quad \end{pmatrix} \end{array}$$

$$B^2 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{pmatrix} & & \\ & & \\ & & \end{pmatrix} \end{array}$$

Morocănosul



$$\lambda^2 = (A^2, B^2, \Pi^2)$$

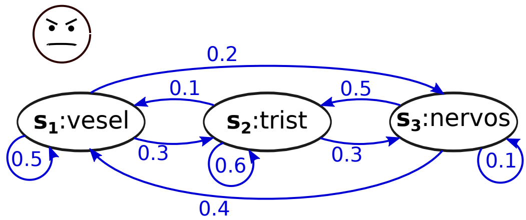
$$A^2 = \begin{array}{c c} & \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ & & \end{pmatrix} \end{array}$$

$$\Pi^2 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} & & \end{pmatrix} \end{array}$$

$$B^2 = \begin{array}{c c} & \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} & & \\ & & \\ & & \end{pmatrix} \end{array}$$

Morocănosul

Morocănosul



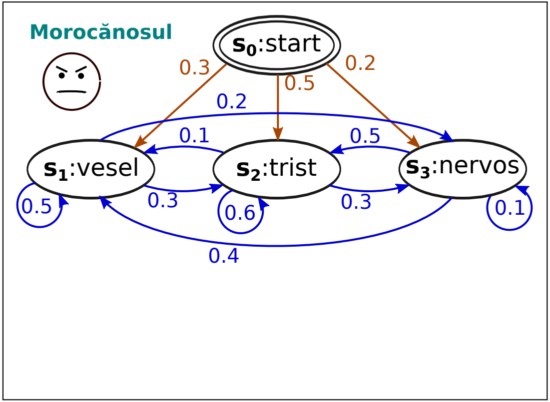
$$\lambda^2 = (A^2, B^2, \Pi^2)$$

$$A^2 = \begin{array}{c c} & \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.4 & 0.5 & 0.1 \end{pmatrix} \end{array}$$

$$\Pi^2 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} \quad & \quad & \quad \end{pmatrix} \end{array}$$

$$B^2 = \begin{array}{c c} & \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} \quad & \quad & \quad \\ & & \\ & & \end{pmatrix} \end{array}$$

Morocănosul



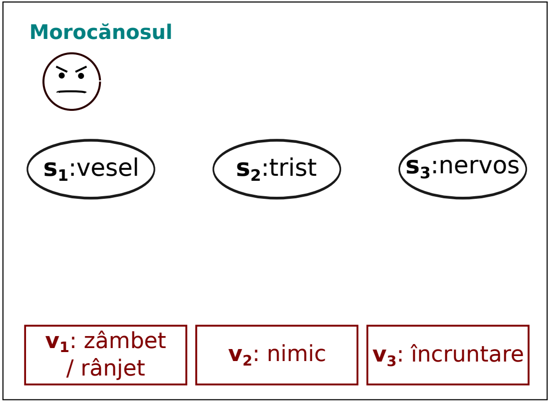
$$\lambda^2 = (A^2, B^2, \Pi^2)$$

$$A^2 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.4 & 0.5 & 0.1 \end{pmatrix} \end{array}$$

$$\Pi^2 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.3 & 0.5 & 0.2 \end{pmatrix} \end{array}$$

$$B^2 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{pmatrix} & & \\ & & \\ & & \end{pmatrix} \end{array}$$

Morocănosul



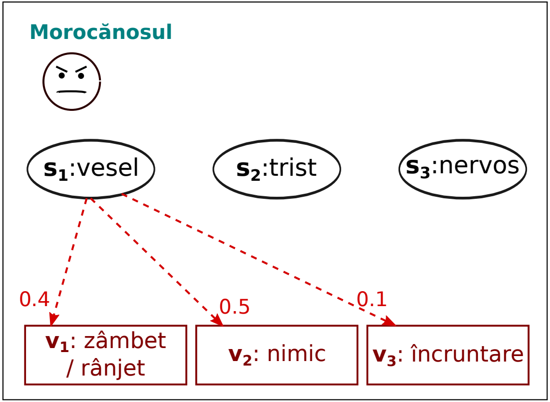
$$\lambda^2 = (A^2, B^2, \Pi^2)$$

$$A^2 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.4 & 0.5 & 0.1 \end{pmatrix} \end{array}$$

$$\Pi^2 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.3 & 0.5 & 0.2 \end{pmatrix} \end{array}$$

$$B^2 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{pmatrix} \quad & \quad & \quad \\ & & \\ & & \end{pmatrix} \end{array}$$

Morocănosul



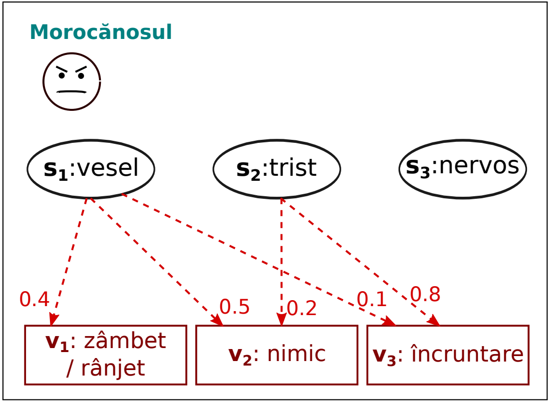
$$\lambda^2 = (A^2, B^2, \Pi^2)$$

$$A^2 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.4 & 0.5 & 0.1 \end{pmatrix} \end{array}$$

$$\Pi^2 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.3 & 0.5 & 0.2 \end{pmatrix} \end{array}$$

$$B^2 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{pmatrix} 0.4 & 0.5 & 0.1 \\ & & \\ & & \end{pmatrix} \end{array}$$

Morocănosul



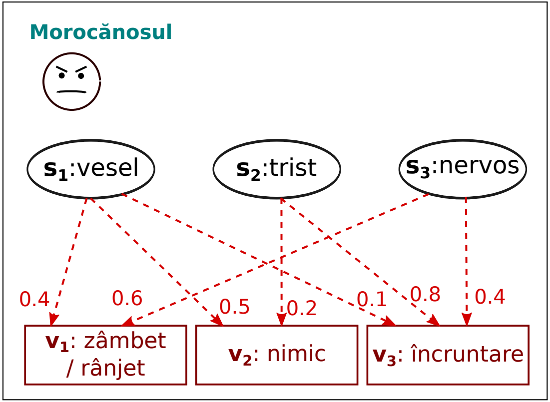
$$\lambda^2 = (A^2, B^2, \Pi^2)$$

$$A^2 = \begin{array}{cc} & \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.4 & 0.5 & 0.1 \end{pmatrix} \end{array}$$

$$\Pi^2 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.3 & 0.5 & 0.2 \end{pmatrix} \end{array}$$

$$B^2 = \begin{array}{cc} & \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{pmatrix} 0.4 & 0.5 & 0.1 \\ 0 & 0.2 & 0.8 \\ & & \end{pmatrix} \end{array}$$

Morocănosul



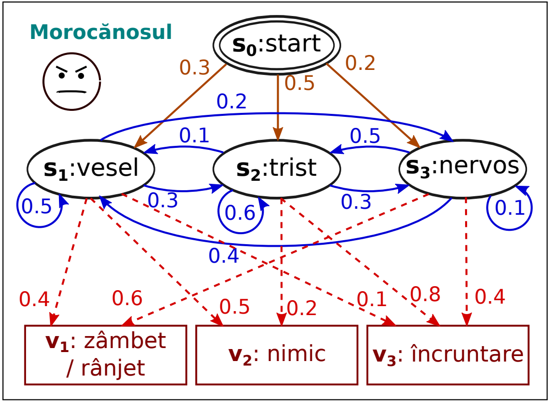
$$\lambda^2 = (A^2, B^2, \Pi^2)$$

$$A^2 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.4 & 0.5 & 0.1 \end{pmatrix} \end{array}$$

$$\Pi^2 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.3 & 0.5 & 0.2 \end{pmatrix} \end{array}$$

$$B^2 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{pmatrix} 0.4 & 0.5 & 0.1 \\ 0 & 0.2 & 0.8 \\ 0.6 & 0 & 0.4 \end{pmatrix} \end{array}$$

Morocănosul



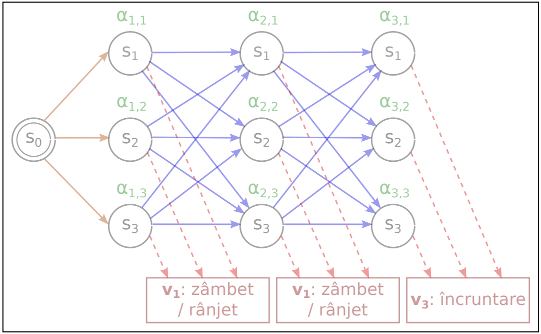
$$\lambda^2 = (A^2, B^2, \Pi^2)$$

$$A^2 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.4 & 0.5 & 0.1 \end{pmatrix} \end{array}$$

$$\Pi^2 = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0.3 & 0.5 & 0.2 \end{pmatrix} \end{array}$$

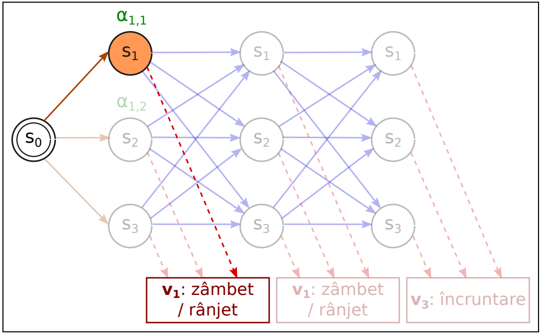
$$B^2 = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{pmatrix} 0.4 & 0.5 & 0.1 \\ 0 & 0.2 & 0.8 \\ 0.6 & 0 & 0.4 \end{pmatrix} \end{array}$$

Calculul variabilelor α:



$$\hat{\alpha} = \begin{bmatrix} \quad \end{bmatrix} \quad c = \begin{bmatrix} \quad \end{bmatrix}$$

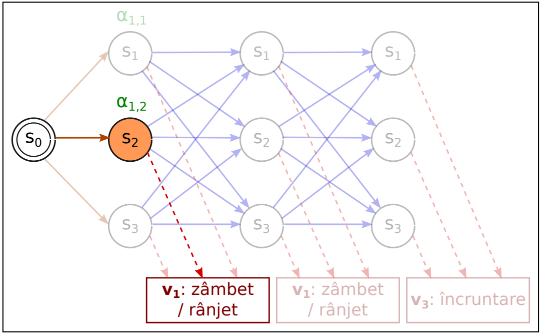
Calculul variabilelor α:



$\ddot{\alpha}_{1,1} = \pi_{1,1} \cdot b_1(v_1) = 0.7 \cdot 0.7 = 0.49$

$$\hat{\alpha} = \begin{bmatrix} \quad \end{bmatrix} \; c = \begin{bmatrix} \quad \end{bmatrix}$$

Calculul variabilelor α:

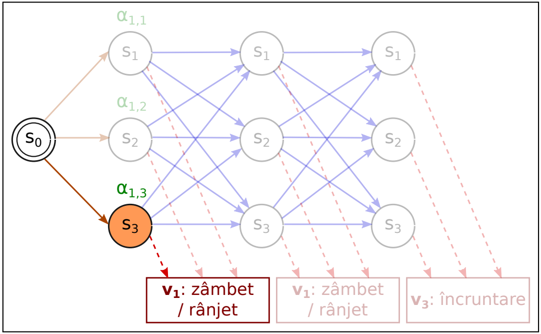


$$\ddot{\alpha}_{1,1} = \pi_{1,1} \cdot b_1(v_1) = 0.7 \cdot 0.7 = 0.49$$

$$\ddot{\alpha}_{1,2} = \pi_{1,2} \cdot b_2(v_1) = 0.2 \cdot 0 = 0$$

$$\hat{\alpha} = \left[\quad \right] \quad c = \left[\quad \right]$$

Calculul variabilelor α:



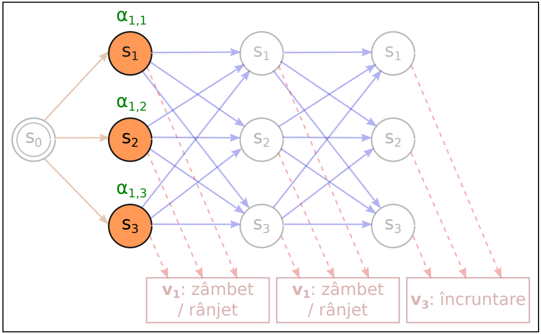
$$\ddot{\alpha}_{1,1}=\pi_{1,1}\cdot b_1(v_1)=0.7\cdot 0.7=0.49$$

$$\ddot{\alpha}_{1,2}=\pi_{1,2}\cdot b_2(v_1)=0.2\cdot 0=0$$

$$\ddot{\alpha}_{1,3}=\pi_{1,3}\cdot b_3(v_1)=0.1\cdot 0.6=0.6$$

$$\hat{\alpha} = \left[\quad \right] \quad c = \left[\quad \right]$$

Calculul variabilelor α:



$$\ddot{\alpha}_{1,1} = \pi_{1,1} \cdot b_1(v_1) = 0.7 \cdot 0.7 = 0.49$$

$$\ddot{\alpha}_{1,2} = \pi_{1,2} \cdot b_2(v_1) = 0.2 \cdot 0 = 0$$

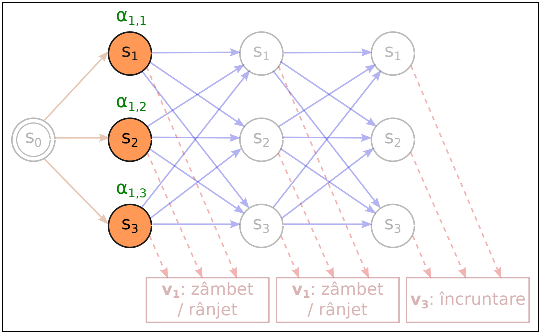
$$\ddot{\alpha}_{1,3} = \pi_{1,3} \cdot b_3(v_1) = 0.1 \cdot 0.6 = 0.6$$

$$c_1 = \frac{1}{\ddot{\alpha}_{1,1} + \ddot{\alpha}_{1,2} + \ddot{\alpha}_{1,3}}$$

$$c_1 = 1/0.54 \approx 1.8182$$

$$\hat{\alpha} = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix} \quad c = \begin{bmatrix} 1.8182 \\ \\ \end{bmatrix}$$

Calculul variabilelor α:



$$\ddot{\alpha}_{1,1} = \pi_{1,1} \cdot b_1(v_1) = 0.7 \cdot 0.7 = 0.49$$

$$\ddot{\alpha}_{1,2} = \pi_{1,2} \cdot b_2(v_1) = 0.2 \cdot 0 = 0$$

$$\ddot{\alpha}_{1,3} = \pi_{1,3} \cdot b_3(v_1) = 0.1 \cdot 0.6 = 0.6$$

$$c_1 = \frac{1}{\ddot{\alpha}_{1,1} + \ddot{\alpha}_{1,2} + \ddot{\alpha}_{1,3}}$$

$$c_1 = 1/0.54 \approx 1.8182$$

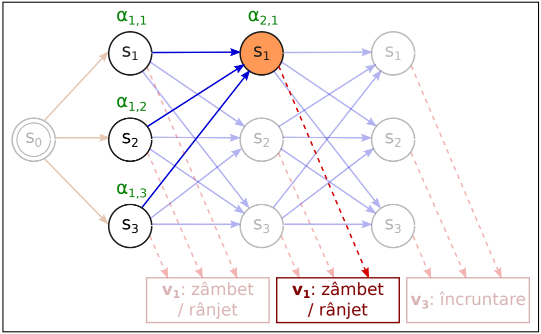
$$\hat{\alpha}_{1,1} = c_1 \cdot \ddot{\alpha}_{1,1} = 0.8909$$

$$\hat{\alpha}_{1,2} = c_1 \cdot \ddot{\alpha}_{1,2} = 0$$

$$\hat{\alpha}_{1,3} = c_1 \cdot \ddot{\alpha}_{1,3} = 0.1091$$

$$\hat{\alpha} = \begin{bmatrix} 0.8909 & 0 & 0.1091 \\ & & \\ & & \end{bmatrix} \quad c = \begin{bmatrix} 1.8182 \\ \\ \end{bmatrix}$$

Calculul variabilelor α:

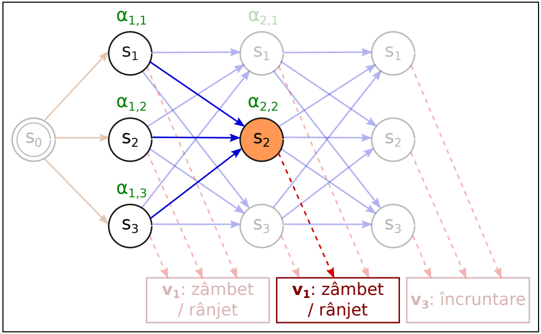


$$\ddot{\alpha}_{2,1} = b_1(v_1) \sum \hat{\alpha}_{1,i} a_{i,1}$$

$$= \ldots = 0.56$$

$$\hat{\alpha} = \begin{bmatrix} 0.8909 & 0 & 0.1091 \\ & & \\ & & \end{bmatrix} \quad c = \begin{bmatrix} 1.8182 \\ \\ \end{bmatrix}$$

Calculul variabilelor α:



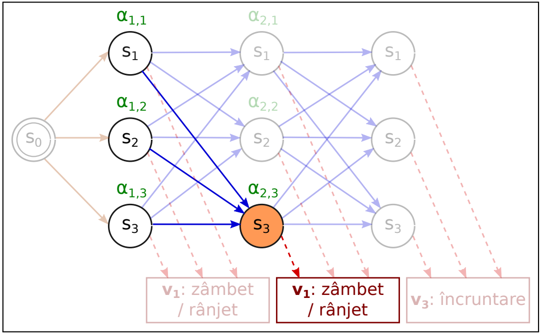
$$\ddot{\alpha}_{2,1} = b_1(v_1) \sum \hat{\alpha}_{1,i} a_{i,1}$$

$$= \ldots = 0.56$$

$$\ddot{\alpha}_{2,2} = b_2(v_1) \sum \hat{\alpha}_{1,i} a_{i,2} = 0$$

$$\hat{\alpha} = \begin{bmatrix} 0.8909 & 0 & 0.1091 \\ & & \\ & & \end{bmatrix} \quad c = \begin{bmatrix} 1.8182 \\ \\ \end{bmatrix}$$

Calculul variabilelor α:



$$\ddot{\alpha}_{2,1} = b_1(v_1) \sum \hat{\alpha}_{1,i} a_{i,1}$$

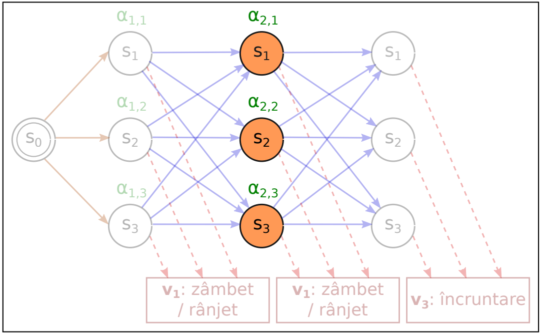
$$= \ldots = 0.56$$

$$\ddot{\alpha}_{2,2} = b_2(v_1) \sum \hat{\alpha}_{1,i} a_{i,2} = 0$$

$$\ddot{\alpha}_{2,3} = b_3(v_1) \sum \hat{\alpha}_{1,i} a_{i,3} = 0.0535$$

$$\hat{\alpha} = \begin{bmatrix} 0.8909 & 0 & 0.1091 \\ & & \\ & & \end{bmatrix} \quad c = \begin{bmatrix} 1.8182 \\ \\ \end{bmatrix}$$

Calculul variabilelor α:



$$\ddot{\alpha}_{2,1} = b_1(v_1) \sum \hat{\alpha}_{1,i} a_{i,1}$$

$$= \ldots = 0.56$$

$$\ddot{\alpha}_{2,2} = b_2(v_1) \sum \hat{\alpha}_{1,i} a_{i,2} = 0$$

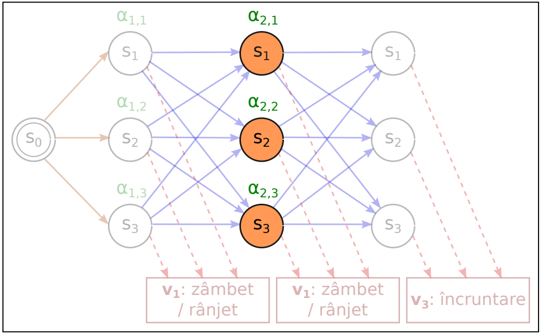
$$\ddot{\alpha}_{2,3} = b_3(v_1) \sum \hat{\alpha}_{1,i} a_{i,3} = 0.0535$$

$$c_2 = \frac{1}{\ddot{\alpha}_{2,1} + \ddot{\alpha}_{2,2} + \ddot{\alpha}_{2,3}}$$

$$c_2 = \frac{1}{0.484 + 0.0776 + 0} = 1.63$$

$$\hat{\alpha} = \begin{bmatrix} 0.8909 & 0 & 0.1091 \\ & & \\ & & \end{bmatrix} \quad c = \begin{bmatrix} 1.8182 \\ 1.63 \\ \\ \end{bmatrix}$$

Calculul variabilelor α:



$$\ddot{\alpha}_{2,1} = b_1(v_1)\sum \hat{\alpha}_{1,i} a_{i,1}$$

$$= \ldots = 0.56$$

$$\ddot{\alpha}_{2,2} = b_2(v_1)\sum \hat{\alpha}_{1,i} a_{i,2} = 0$$

$$\ddot{\alpha}_{2,3} = b_3(v_1)\sum \hat{\alpha}_{1,i} a_{i,3} = 0.0535$$

$$c_2 = \frac{1}{\ddot{\alpha}_{2,1} + \ddot{\alpha}_{2,2} + \ddot{\alpha}_{2,3}}$$

$$c_2 = \frac{1}{0.484 + 0.0776 + 0} = 1.63$$

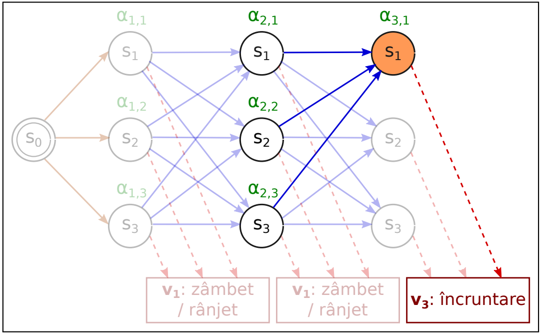
$$\hat{\alpha} = \begin{bmatrix} 0.8909 & 0 & 0.1091 \\ 0.9128 & 0 & 0.0872 \\ & & \end{bmatrix} \quad c = \begin{bmatrix} 1.8182 \\ 1.63 \\ \ \end{bmatrix}$$

$$\hat{\alpha}_{2,1} = c_2 \cdot \ddot{\alpha}_{2,1} = 0.9128$$

$$\hat{\alpha}_{2,2} = c_2 \cdot \ddot{\alpha}_{2,2} = 0$$

$$\hat{\alpha}_{2,3} = c_2 \cdot \ddot{\alpha}_{2,3} = 0.0872$$

Calculul variabilelor α:

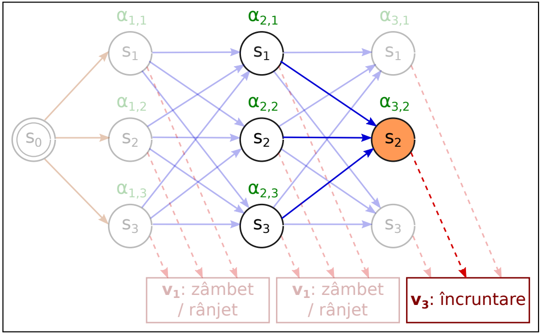


$$\ddot{\alpha}_{3,1} = b_1(v_3) \sum \hat{\alpha}_{2,i} a_{i,1}$$

$$= \ldots = 0$$

$$\hat{\alpha} = \begin{bmatrix} 0.8909 & 0 & 0.1091 \\ 0.9128 & 0 & 0.0872 \\ & & \end{bmatrix} \quad c = \begin{bmatrix} 1.8182 \\ 1.63 \\ \end{bmatrix}$$

Calculul variabilelor α:



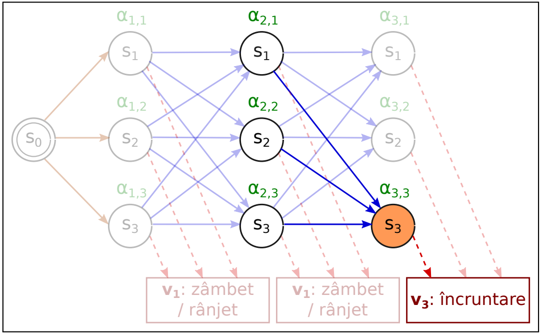
$$\ddot{\alpha}_{3,1} = b_1(v_3) \sum \hat{\alpha}_{2,i} a_{i,1}$$

$$= \ldots = 0$$

$$\ddot{\alpha}_{3,2} = b_2(v_3) \sum \hat{\alpha}_{2,i} a_{i,2} = 0.0326$$

$$\hat{\alpha} = \begin{bmatrix} 0.8909 & 0 & 0.1091 \\ 0.9128 & 0 & 0.0872 \\ & & \end{bmatrix} \quad c = \begin{bmatrix} 1.8182 \\ 1.63 \\ \end{bmatrix}$$

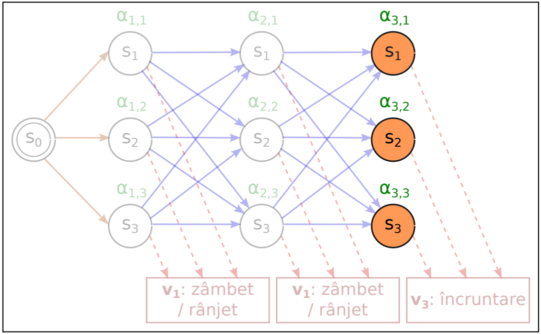
Calculul variabilelor α:



$$\ddot{\alpha}_{3,1} = b_1(v_3) \sum \hat{\alpha}_{2,i} a_{i,1}$$
$$= \ldots = 0$$
$$\ddot{\alpha}_{3,2} = b_2(v_3) \sum \hat{\alpha}_{2,i} a_{i,2} = 0.0326$$
$$\ddot{\alpha}_{3,3} = b_3(v_3) \sum \hat{\alpha}_{2,i} a_{i,3} = 0.0365$$

$$\hat{\alpha} = \begin{bmatrix} 0.8909 & 0 & 0.1091 \\ 0.9128 & 0 & 0.0872 \\ & & \end{bmatrix} \quad c = \begin{bmatrix} 1.8182 \\ 1.63 \\ \ \end{bmatrix}$$

Calculul variabilelor α:



$$\ddot{\alpha}_{3,1} = b_1(v_3) \sum \hat{\alpha}_{2,i} a_{i,1}$$

$$= \ldots = 0$$

$$\ddot{\alpha}_{3,2} = b_2(v_3) \sum \hat{\alpha}_{2,i} a_{i,2} = 0.0326$$

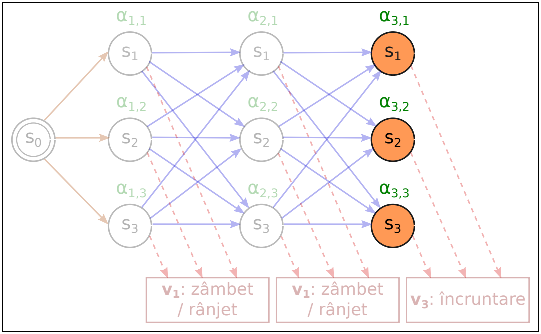
$$\ddot{\alpha}_{3,3} = b_3(v_3) \sum \hat{\alpha}_{2,i} a_{i,3} = 0.0365$$

$$c_3 = \frac{1}{\ddot{\alpha}_{3,1} + \ddot{\alpha}_{3,2} + \ddot{\alpha}_{,3}}$$

$$c_3 = \frac{1}{x} = 14.4718$$

$$\hat{\alpha} = \begin{bmatrix} 0.8909 & 0 & 0.1091 \\ 0.9128 & 0 & 0.0872 \\ & & \end{bmatrix} \quad c = \begin{bmatrix} 1.8182 \\ 1.63 \\ 14.4718 \end{bmatrix}$$

Calculul variabilelor α:



$$\ddot{\alpha}_{3,1} = b_1(v_3) \sum \hat{\alpha}_{2,i} a_{i,1}$$
$$= \ldots = 0$$
$$\ddot{\alpha}_{3,2} = b_2(v_3) \sum \hat{\alpha}_{2,i} a_{i,2} = 0.0326$$
$$\ddot{\alpha}_{3,3} = b_3(v_3) \sum \hat{\alpha}_{2,i} a_{i,3} = 0.0365$$

$$c_3 = \frac{1}{\ddot{\alpha}_{3,1} + \ddot{\alpha}_{3,2} + \ddot{\alpha}_{,3}}$$
$$c_3 = \frac{1}{x} = 14.4718$$

$$\hat{\alpha}_{3,1} = c_3 \cdot \ddot{\alpha}_{3,1} = 0$$
$$\hat{\alpha}_{3,2} = c_3 \cdot \ddot{\alpha}_{3,2} = 0.4718$$
$$\hat{\alpha}_{3,3} = c_3 \cdot \ddot{\alpha}_{3,3} = 0.5282$$

$$\hat{\alpha} = \begin{bmatrix} 0.8909 & 0 & 0.1091 \\ 0.9128 & 0 & 0.0872 \\ 0 & 0.4718 & 0.5282 \end{bmatrix} \quad c = \begin{bmatrix} 1.8182 \\ 1.63 \\ 14.4718 \end{bmatrix}$$

Morocănos sau jovial?

- $c = \begin{bmatrix} 1.8182 & 1.63 & 14.4718 \end{bmatrix}$
- Probabilitatea ca secvența observată să fi fost generată de un *jovial*: $\log(P(O|\lambda^1)) = -\sum \log(c_i) = -3.7583$

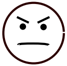
Morocănos sau jovial?

- $c = \begin{bmatrix} 1.8182 & 1.63 & 14.4718 \end{bmatrix}$
- Probabilitatea ca secvența observată să fi fost generată de un *jovial*: $\log(P(O|\lambda^1)) = -\sum \log(c_i) = -3.7583$
- Probabilitatea ca secvența observată să fi fost generată de un *morocănos*: $\log(P(O|\lambda^2)) = -3.6362$
- $\log(P(O|\lambda^2)) > \log(P(O|\lambda^1))$
- Este mai probabil să avem de-aface cu...

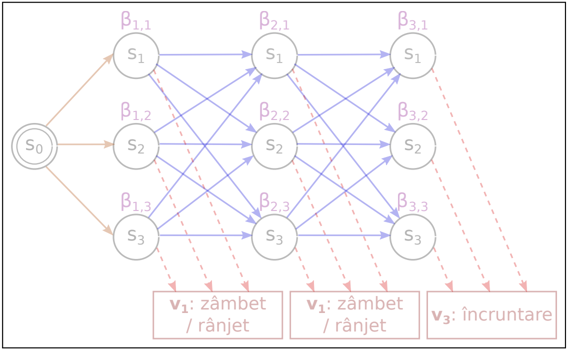
Morocănos sau jovial?

- $c = \begin{bmatrix} 1.8182 & 1.63 & 14.4718 \end{bmatrix}$
- Probabilitatea ca secvența observată să fi fost generată de un *jovial*:
  $\log(P(O|\lambda^1)) = -\sum \log(c_i) = -3.7583$
- Probabilitatea ca secvența observată să fi fost generată de un *morocănos*:
  $\log(P(O|\lambda^2)) = -3.6362$
- $\log(P(O|\lambda^2)) > \log(P(O|\lambda^1))$
- Este mai probabil să avem de-aface cu...

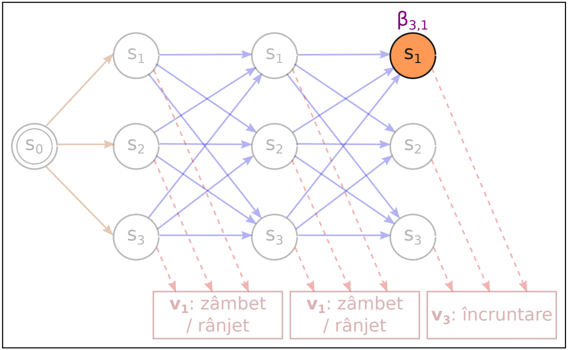
Morocănos



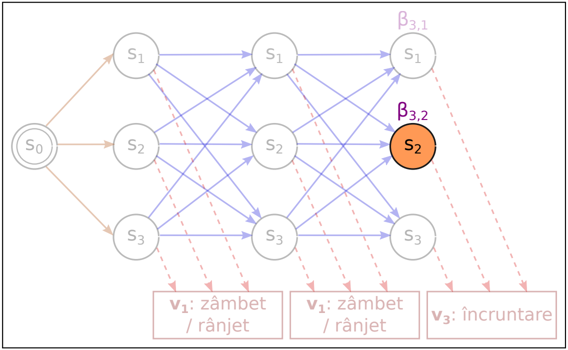
Calculul variabilelor β:



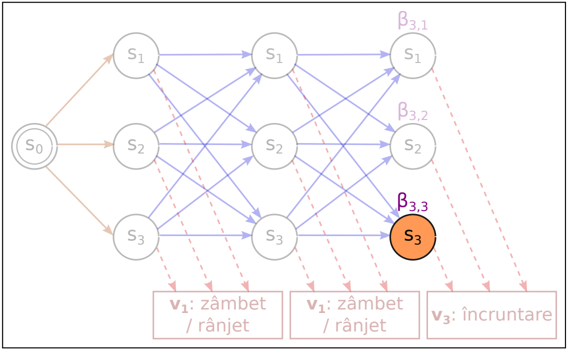
Calculul variabilelor β:



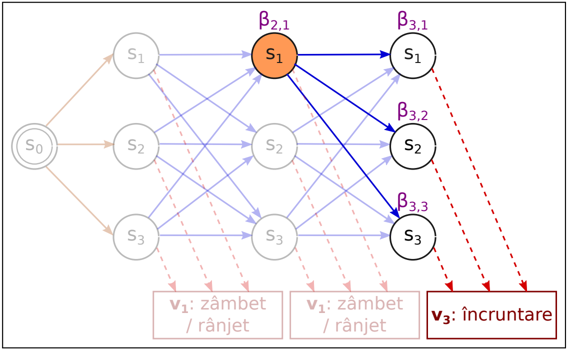
Calculul variabilelor β:



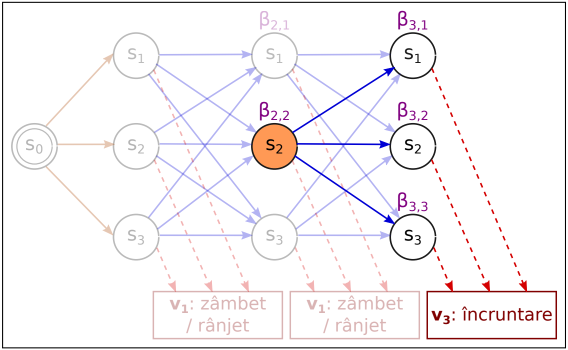
Calculul variabilelor β:



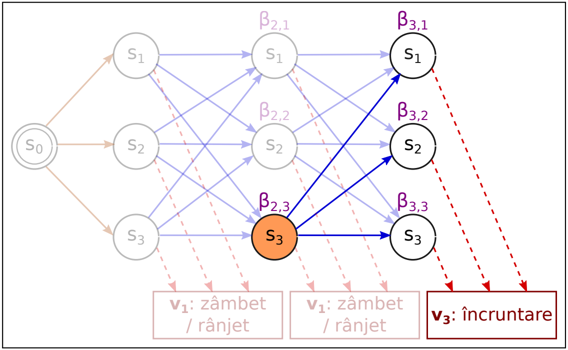
Calculul variabilelor β:



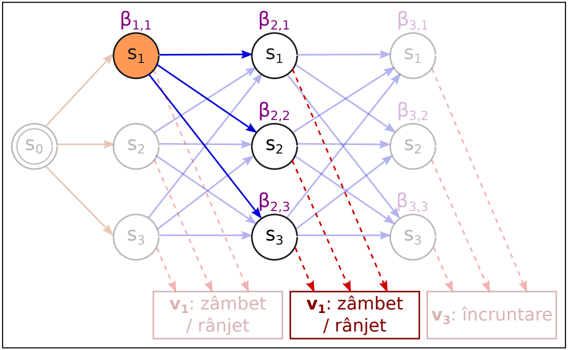
Calculul variabilelor β:



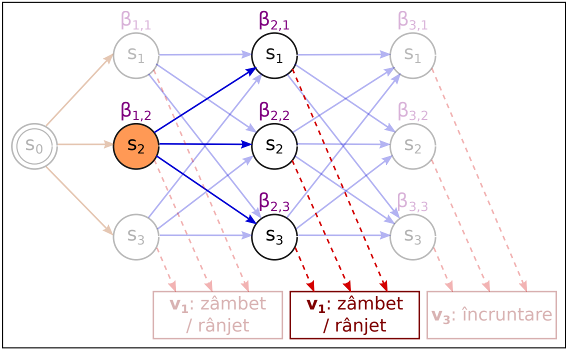
Calculul variabilelor β:



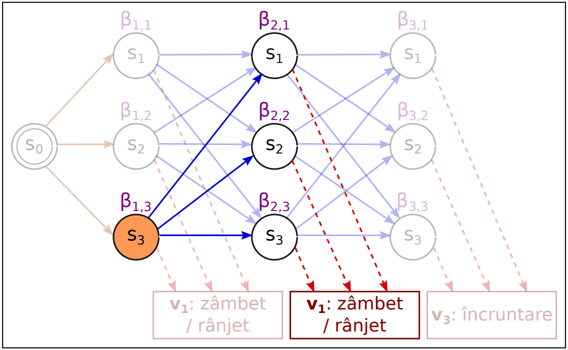
Calculul variabilelor β:



Calculul variabilelor β:



Calculul variabilelor β:



Algoritmul Forward-Backward

Algoritmul 1 Calculul variabilelor α

for $i = 1$ to N **do**
  $\ddot{\alpha}_{1,i} \leftarrow \pi_i \cdot b_i(o_1)$
end for
$c_1 \leftarrow (\sum_{i=1}^{N} \ddot{\alpha}_{1,i})^{-1}$
for $i = 1$ to N **do**
  $\hat{\alpha}_{1,i} \leftarrow c_1 \cdot \ddot{\alpha}_{1,i}$
end for
for $t = 1$ to $T - 1$ **do**
  for $i = 1$ to N **do**
    $$\ddot{\alpha}_{t+1,i} \leftarrow \left[\sum_{i=1}^{N} \hat{\alpha}_{t,i} a_{i,j}\right] b_j(o_{t+1})$$
  end for
  $$c_{t+1} \leftarrow (\sum_{i=1}^{N} \ddot{\alpha}_{t+1,i})^{-1}$$
  for $i = 1$ to N **do**
    $\hat{\alpha}_{t+1,i} \leftarrow c_{t+1} \cdot \ddot{\alpha}_{t+1,i}$
  end for
end for

Algoritmul 2 Calculul $P(O|\lambda)$

$$logP \leftarrow -\sum_{t=1}^{T} c_t$$

Algoritmul 3 Calculul variabilelor β

for $i = 1$ to N **do**
  $\hat{\beta}_{T,i} \leftarrow \cdot c_T$
end for
for $t = (T-1)$ to 1 **do**
  for $i = 1$ to N **do**
    $$\hat{\beta}_{t,i} \leftarrow \sum_{j=1}^{N} a_{i,j} b_j(o_{t+1}) \hat{\beta}_{t+1,j} \cdot c_t$$
  end for
end for

A venit vremea să scriem cod

1. Faceți o copie a fișierului `forward_backward_disc.m.stub` și denumiți-o `forward_backward_disc.m` (eliminați sufixul `.stub`). Veți implementa funcția:

```
function [logP, Alpha, Beta, Scale] = ...
forward_backward_disc(O,Pi, A, B)
```

A venit vremea să scriem cod

1. Faceți o copie a fișierului `forward_backward_disc.m.stub` și denumiți-o `forward_backward_disc.m` (eliminați sufixul `.stub`). Veți implementa funcția:

```
function [logP, Alpha, Beta, Scale] = ...
forward_backward_disc(O,Pi, A, B)
```

2. Completați cele trei secțiuni.

```
...
Scale = zeros (1, T); % Scale is an 1 x T matrix
Alpha = zeros (T, N); % Alpha is a T x N matrix
Beta = ones (T, N); % Beta is a T x N matrix
%% Forward variables
% alpha_disc-start - Write code below

% alpha_disc-end - Write code above
...
```

3. Folosiți `hmm_test` pentru a vă testa codul.

Outline

1. Aplicații în Învățarea Automată pentru MMA
   - Învățarea Automată
   - MMA în Învățarea Automată
2. Teoria MMA
   - Cele Trei Probleme ale MMA
   - Fundamente Matematice
3. Implementarea MMA
   - Problema Evaluării: Algoritmul Forward-Backward
   - Problema Interpretării: Algoritmul Viterbi
   - Problema Estimării: Algoritmul Baum-Welch
4. Demo: Recunoașterea Simbolurilor
5. Tipuri de MMA
6. Discuții și Concluzii

Problema interpretării

Problema interpretării unei secvențe de observații

Date fiind un model $\lambda = (A, B, \Pi)$ și o secvență de observații $O = [o_1 o_2 \cdots o_T]$, cum alegem o secvență corespunzătoare de stări $Q_{\text{best}} = [q1_{\text{best}} q2_{\text{best}} \cdots q_{T_{\text{best}}}]$ care *să dea un înțeles* observațiilor? Cum *descoperim* partea ascunsă a modelului?

Problema interpretării

Problema interpretării unei secvențe de observații

Date fiind un model $\lambda = (A, B, \Pi)$ și o secvență de observații $O = [o_1 o_2 \cdots o_T]$, cum alegem o secvență corespunzătoare de stări $Q_{\text{best}} = [q_{1_{\text{best}}} q_{2_{\text{best}}} \cdots q_{T_{\text{best}}}]$ care *să dea un înțeles* observațiilor? Cum *descoperim* partea ascunsă a modelului?

- Răspunsul depinde de criteriul cu care alegem *cea mai bună* secvență

Problema interpretării

Problema interpretării unei secvențe de observații

Date fiind un model $\lambda = (A, B, \Pi)$ și o secvență de observații $O = [o_1 o_2 \cdots o_T]$, cum alegem o secvență corespunzătoare de stări $Q_{\text{best}} = [q1_{\text{best}} q2_{\text{best}} \cdots q_{T_{\text{best}}}]$ care *să dea un înțeles* observațiilor? Cum *descoperim* partea ascunsă a modelului?

- Răspunsul depinde de criteriul cu care alegem *cea mai bună* secvență
  - secvența celor mai probabile stări $q_{t_{\text{best}}}$ (luate individual), date fiind modelul și secvența observată:
    $q_{t_{\text{best}}} = \underset{s_i}{\text{argmax}}\, P(q_t = s_i | O, \lambda)$
  - cea mai probabilă secvență de stări Q (per ansamblu), date fiind modelul și secvența observată:
    $Q_{\text{best}} = \underset{\text{all } Q}{\text{argmax}}\, P(Q | O, \lambda)$

Secvența celor mai probabile stări

Notație: $P(q_t = s_i | O, \lambda) = \gamma_{\mathbf{t,i}} = \frac{\alpha_{t,i}\beta_{t,i}}{\sum_{j=1}^{N} \alpha_{t,j}\beta_{t,j}}$

- Este un criteriu satisfăcător?

Secvența celor mai probabile stări

Notație: $P(q_t = s_i | O, \lambda) = \gamma_{\mathbf{t,i}} = \frac{\alpha_{t,i}\beta_{t,i}}{\displaystyle\sum_{j=1}^{N} \alpha_{t,j}\beta_{t,j}}$

- Este un criteriu satisfăcător?
- **NU!**
  Pot exista q_t și q_{t+1} astfel încât $a_{q_t,q_{t+1}} = 0$

Algoritmul Viterbi

- Criteriu care ia în considerare distribuțiile de probabilitate ale tranzițiilor între stări
- Cea mai bună cale: $Q_{\text{best}} = [q_{1_{\text{best}}} q_{2_{\text{best}}} \cdots q_{T_{\text{best}}}]$

$$Q_{\text{best}} = \underset{Q}{\operatorname{argmax}}\, P(Q|O,\lambda) = \underset{Q}{\operatorname{argmax}}\, P(Q,O|\lambda) \tag{6}$$

- **Algoritmul Viterbi** - programare dinamică
- Vom introduce variabilele *delta*.

Variabilele δ - intuiție

Întrebare

Dacă $Q = [q_1, q_2, \ldots, q_{t-1}, q_t]$ este cea mai bună secvență care explică $O = [o_1, o_2, \ldots o_{t-1}, o_t]$, atunci putem afirma că $Q[1 : t-1]$ este cea mai bună secvență de stări care explică $O[1 : t-1]$?

Variabilele δ - intuiție

Întrebare

Dacă $Q = [q_1, q_2, \ldots, q_{t-1}, q_t]$ este cea mai bună secvență care explică $O = [o_1, o_2, \ldots o_{t-1}, o_t]$, atunci putem afirma că $Q[1 : t-1]$ este cea mai bună secvență de stări care explică $O[1 : t-1]$?

- **NU!**

$$P(q_1 = s_{i_1}, q_2 = s_{i_2}, \ldots, q_{t-1} = s_{i_{t-1}}, q_t = s_{i_t} | O, \lambda) =$$
$$P(q_1, q_2, \ldots, q_{t-1} | O, \lambda) \cdot P(q_t = s_{i_t} | q_{t-1} = s_{i_{t-1}}, \lambda) \cdot P(o_t | q_t = s_{i_t}, \lambda)$$

Variabilele δ

- Vom numi variabile δ:

$$\delta_{t,i} = \max_{q_1,\ldots,q_{t-1}} P([q_1 q_2 \ldots q_{t-1} s_i], [o_1, o_2, \ldots o_t] | \lambda) \tag{26}$$

- $\delta_{\mathbf{t},\mathbf{i}}$ - cea mai mare probabilitate a unei secvențe de stări de lungime t care ajunge în s_i și explică primele t valori observate

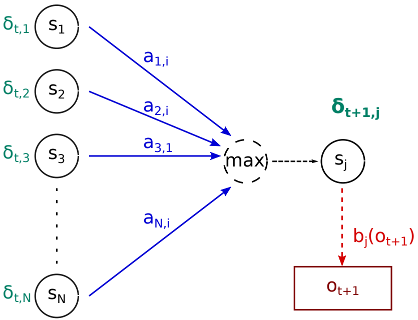
Variabilele δ

- Vom numi variabile δ:

$$\delta_{t,i} = \max_{q_1,\ldots,q_{t-1}} P([q_1 q_2 \ldots q_{t-1} s_i], [o_1, o_2, \ldots o_t] | \lambda) \tag{26}$$

- $\delta_{\mathbf{t,i}}$ - cea mai mare probabilitate a unei secvențe de stări de lungime t care ajunge în s_i și explică primele t valori observate
- relația dintre variabilele δ:

$$\delta_{t+1,j} = [\max_i \delta_{t,i} \cdot a_{i,j}] \cdot b_j(o_{t+1}) \tag{27}$$



Algoritmul Viterbi - Privire de ansamblu

Pași înainte

Calculăm cea mai mare probabilitate de a ajunge în starea s_i la momentul t pe baza celor mai mari probabilități de a fi ajuns în toate stările s_j la $t-1$.

Algoritmul Viterbi - Privire de ansamblu

Pași înainte

Calculăm cea mai mare probabilitate de a ajunge în starea s_i la momentul t pe baza celor mai mari probabilități de a fi ajuns în toate stările s_j la $t-1$.

La final ($t = T$)

Starea finală este acea stare s_i cu cea mai mare probabilitate de a ajunge la ea după observarea tuturor valorilor.

Algoritmul Viterbi - Privire de ansamblu

Pași înainte

Calculăm cea mai mare probabilitate de a ajunge în starea s_i la momentul t pe baza celor mai mari probabilități de a fi ajuns în toate stările s_j la $t-1$.

La final ($t = T$)

Starea finală este acea stare s_i cu cea mai mare probabilitate de a ajunge la ea după observarea tuturor valorilor.

Pași înapoi

Refacem *cea mai bună cale* alegând pentru fiecare moment t starea care a dus la probabilitatea maximă pentru starea de la $t+1$.

Algoritmul Viterbi (I)

1 Inițializare:

$$\begin{aligned} \delta_{1,i} &= \pi_i b_i(o_1), \quad 1 \leq i \leq N \\ \psi_{1,i} &= 0 \end{aligned} \tag{28}$$

2 Recursivitate:

$$\begin{aligned} \delta_{t,j} &= [\max_i \delta_{t-1,i} \cdot a_{i,j}] \cdot b_j(o_t) \quad {\scriptstyle 2 \leq t \leq T, 1 \leq j \leq N} \\ \psi_{t,i} &= \operatorname*{argmax}_i \delta_{t-1,i} \cdot a_{i,j} \quad {\scriptstyle 2 \leq t \leq T, 1 \leq j \leq N} \end{aligned} \tag{29}$$

Algoritmul Viterbi (II)

3 Terminare:

$$P(Q_{\text{best}}|O,\lambda) = \max_i \delta_{T,i}$$
$$q_{T_{\text{best}}} = \operatorname*{argmax}_i \delta_{T,i} \tag{30}$$

4 Backtracking:

$$q_{t_{\text{best}}} = \psi_{t+1}(q_{t+1_{\text{best}}}), \quad {\scriptstyle t=T-1,T-2,\cdots,1} \tag{31}$$

Probleme numerice

- Cine este $P(Q_{\text{best}}|O,\lambda)$?

$$P(Q_{\text{best}}|O,\lambda) = \delta_{T,i_T}$$
$$P(Q_{\text{best}}|O,\lambda) = \delta_{T-1,i_{T-1}} \cdot a_{i_{T-1},i_T} \cdot b_T(o_T)$$
$$P(Q_{\text{best}}|O,\lambda) = \delta_{T-2,i_{T-2}} \cdot a_{i_{T-2},i_{T-1}} \cdot b_{T-1}(o_{T-1}) \cdot a_{i_{T-1},j_T} \cdot b_T(o_T)$$
$$P(Q_{\text{best}}|O,\lambda) = \prod \dots$$

- Cum putem evita apropierea rapidă de zero?

Probleme numerice

- Cine este $P(Q_{\text{best}}|O,\lambda)$?

$$P(Q_{\text{best}}|O,\lambda) = \delta_{T,i_T}$$
$$P(Q_{\text{best}}|O,\lambda) = \delta_{T-1,i_{T-1}} \cdot a_{i_{T-1},i_T} \cdot b_T(o_T)$$
$$P(Q_{\text{best}}|O,\lambda) = \delta_{T-2,i_{T-2}} \cdot a_{i_{T-2},i_{T-1}} \cdot b_{T-1}(o_{T-1}) \cdot a_{i_{T-1},j_T} \cdot b_T(o_T)$$
$$P(Q_{\text{best}}|O,\lambda) = \prod \ldots$$

- Cum putem evita apropierea rapidă de zero?
- Calculăm *log(P)*

$$\log(P(Q_{\text{best}}|O,\lambda)) = \log(\delta_{T,i_T})$$
$$\log(P(Q_{\text{best}}|O,\lambda)) = \log(\delta_{T-1,i_{T-1}}) + \log(a_{i_{T-1},i_T}) + \log(b_T(o_T))$$
$$\log(P(Q_{\text{best}}|O,\lambda)) = \log(\delta_{T-2,i_{T-2}}) + \log(a_{i_{T-2},i_{T-1}}) + \log(b_{T-1}(o_{T-1})$$
$$\log(P(Q_{\text{best}}|O,\lambda)) = \sum \ldots$$

Probleme numerice - Rezolvare

- Notație:

$$\phi_{t,i} = \max_{q_1,\ldots,q_{t-1}} \log(P(q_1,\ldots,q_{t-1},q_t = s_i, o_1,\ldots,o_t|\lambda)) = \log(\delta_{t,i})$$

- Matricele $\log(\Pi)$, $\log(A)$ și $\log(B)$ pot fi precalculate.

Algoritmul Viterbi - $log(P)$ (I)

1 Inițializare:

$$\begin{aligned} &\phi_{1,i} = \log(\pi_i) + \log(b_i(o_1)), \quad 1 \leq i \leq N \\ &\psi_{1,i} = 0 \end{aligned} \tag{32}$$

2 Recursivitate:

$$\begin{aligned} &\phi_{t,j} = [\max_i \phi_{t-1,i} + log(a_{i,j})] + \log(b_j(o_t)) \quad {\scriptstyle 2 \leq t \leq T, 1 \leq j \leq N} \\ &\psi_{t,i} = \operatorname*{argmax}_i \phi_{t-1,i} + \log(a_{i,j}) \quad {\scriptstyle 2 \leq t \leq T, 1 \leq j \leq N} \end{aligned} \tag{33}$$

Algoritmul Viterbi - $log(P)$ (II)

3 Terminare:

$$\log(P(Q_{\text{best}}|O,\lambda)) = \max_i \phi_{T,i}$$
$$q_{T_{\text{best}}} = \operatorname*{argmax}_i \phi_{T,i} \tag{34}$$

4 Backtracking:

$$q_{t_{\text{best}}} = \psi_{t+1}(q_{t+1_{\text{best}}}), \quad {\scriptstyle t=T-1,T-2,\cdots,1} \tag{35}$$

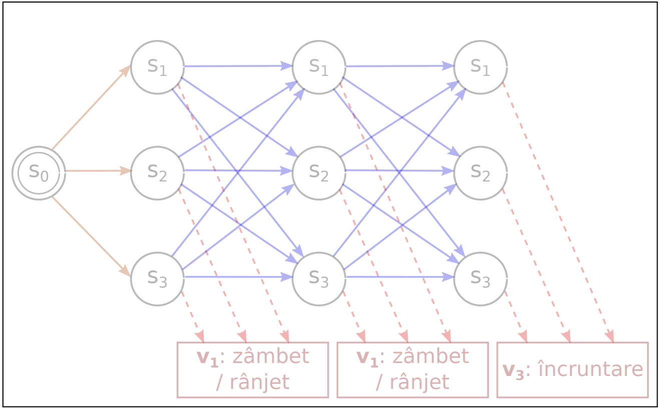
Înapoi la exemplu

- Să reluăm exemplul cu robotul.
- Robotul a recunoscut secvența de gesturi:
  (zâmbet | rânjet) ⟶ (zâmbet | rânjet) ⟶ încruntare
- Am stabilit: $P(O|\lambda^1) > P(O|\lambda^2)$

Înapoi la exemplu

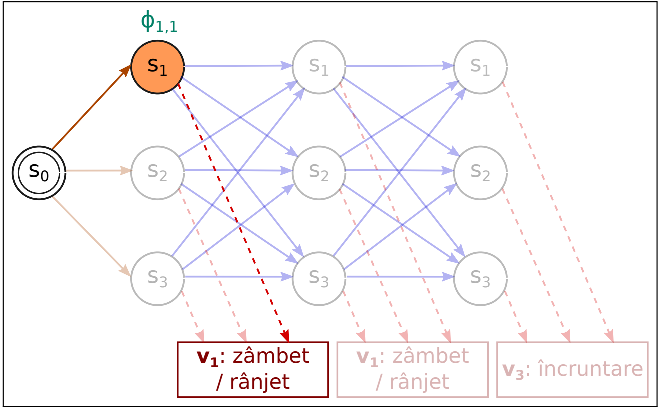
- Să reluăm exemplul cu robotul.
- Robotul a recunoscut secvența de gesturi:
  (zâmbet | rânjet) $\longrightarrow$ (zâmbet | rânjet) $\longrightarrow$ încruntare
- Am stabilit: $P(O|\lambda^1) > P(O|\lambda^2)$
  Deci, robotul se confruntă *[, probabil]* cu un morocănos!
- A doua întrebare: **Prin ce stări a trecut acesta?**

Jovialul



$$\Phi = \left[\qquad \qquad \qquad \right] \quad \Psi = \left[\qquad \right] \quad Q_{best} = [\qquad]$$

Jovialul

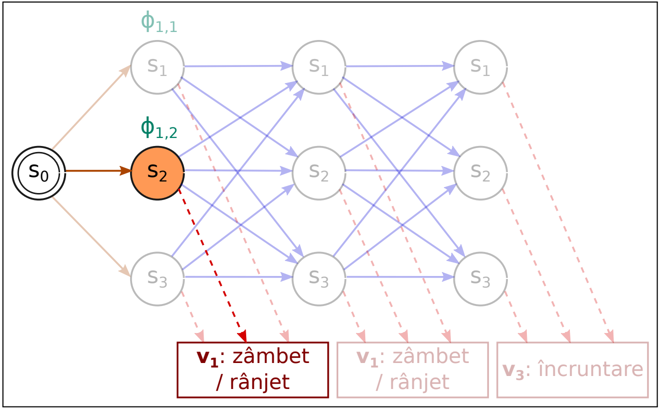


$\phi_{1,1} = \log(\pi_1) + \log(b_{1,1})$

$\psi_{1,1} = 0$

$$\Phi = \begin{bmatrix} -2.1202 & & \\ & & \\ & & \end{bmatrix} \quad \Psi = \begin{bmatrix} 0 & & \\ & & \\ & & \end{bmatrix} \quad Q_{best} = \begin{bmatrix} & & \end{bmatrix}$$

Jovialul

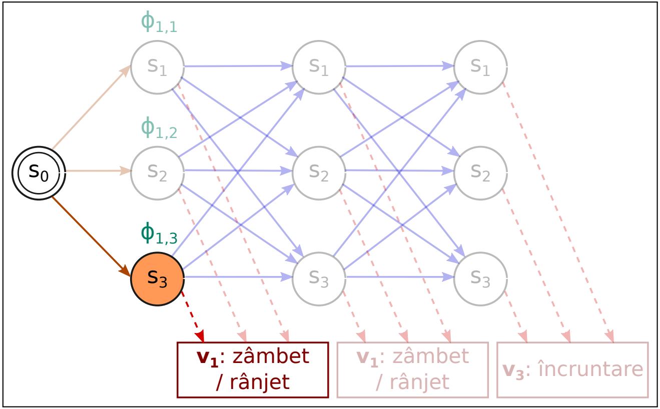


$\phi_{1,2}$=log(π_2)+log($b_{2,1}$)

$\psi_{1,2}$=0

$$\Phi = \begin{bmatrix} -2.1202 & -Inf & \\ & & \\ & & \end{bmatrix} \quad \Psi = \begin{bmatrix} 0 & 0 \\ & \\ & \end{bmatrix} \quad Q_{best} = [\quad]$$

Jovialul

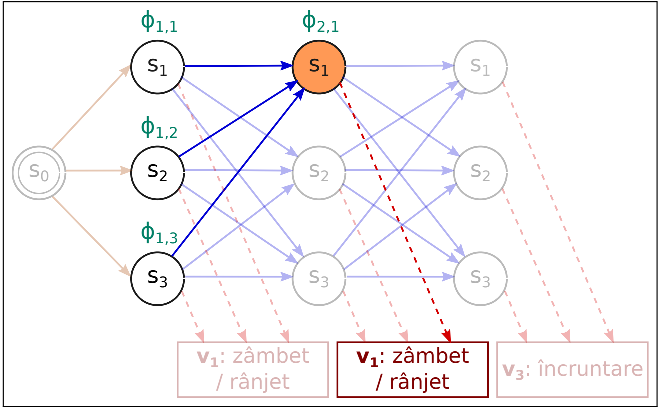


$\phi_{1,3}$=log(π_3)+log$(b_{3,1})$

$\psi_{1,3}$=0

$$\Phi = \begin{bmatrix} -2.1202 & -Inf & -2.1202 \\ & & \\ & & \end{bmatrix} \quad \Psi = \begin{bmatrix} 0 & 0 & 0 \\ & & \\ & & \end{bmatrix} \quad Q_{best} = [\quad]$$

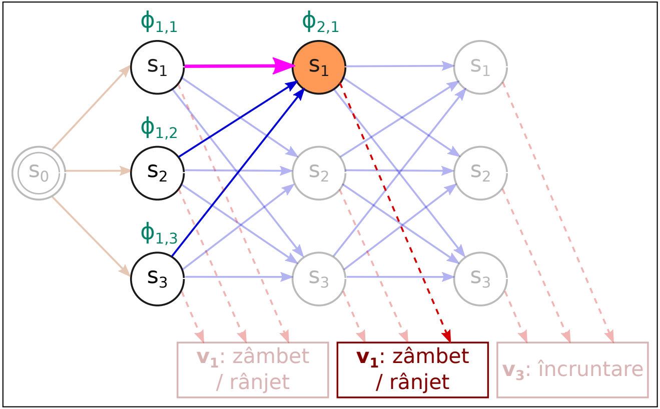
Jovialul



$$\phi_{1,3} = \max \left\{ \begin{array}{l} \log(\phi_{1,1}) + \log(a_{1,1}), \\ \log(\phi_{2,1}) + \log(a_{2,1}), \\ \log(\phi_{1,3}) + \log(a_{3,1}) \end{array} \right\} + \log(b_{1,1})$$

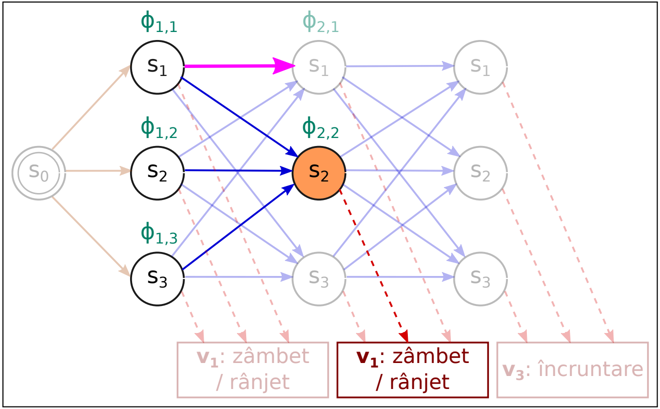
$$\Phi = \begin{bmatrix} -2.1202 & -Inf & -2.1202 \\ -3.7297 & & \\ & & \end{bmatrix} \quad \Psi = \begin{bmatrix} 0 & 0 & 0 \\ & & \\ & & \end{bmatrix} \quad Q_{best} = \left[\quad \right]$$

Jovialul



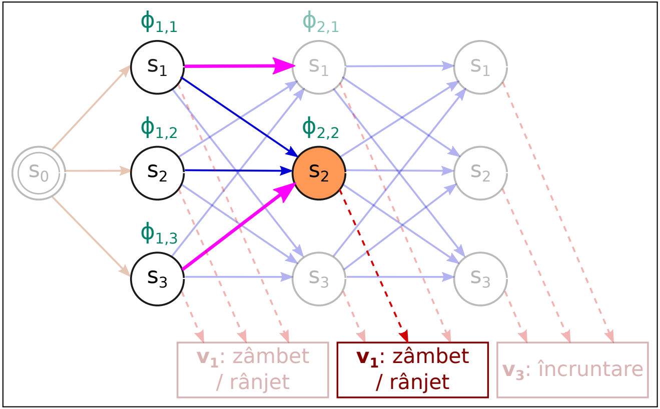
$$\Phi = \begin{bmatrix} -2.1202 & -Inf & -2.1202 \\ -3.7297 & & \\ & & \end{bmatrix} \quad \Psi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & & \\ & & \end{bmatrix} \quad Q_{best} = [\quad]$$

Jovialul



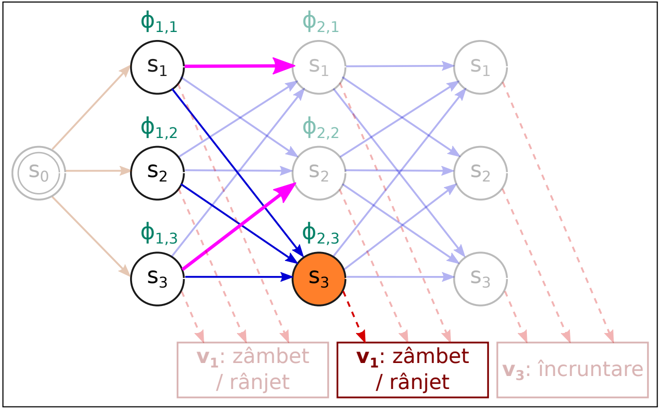
$$\Phi = \begin{bmatrix} -2.1202 & -Inf & -2.1202 \\ -3.7297 & -Inf & \\ & & \end{bmatrix} \quad \Psi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & & \\ & & \end{bmatrix} \quad Q_{best} = [\quad]$$

Jovialul



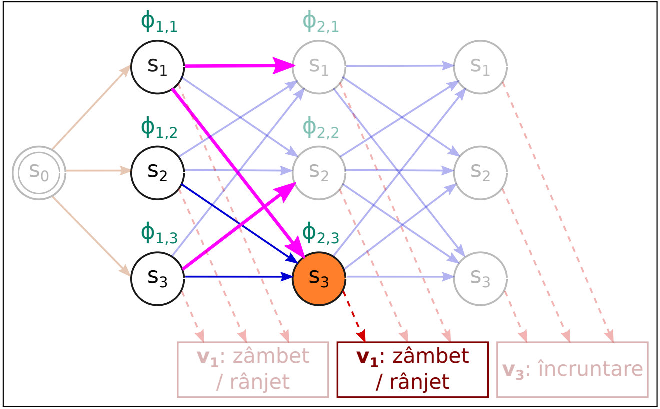
$$\Phi = \begin{bmatrix} -2.1202 & -Inf & -2.1202 \\ -3.7297 & -Inf & \\ & & \end{bmatrix} \quad \Psi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 3 & \\ & & \end{bmatrix} \quad Q_{best} = [\quad\quad]$$

Jovialul



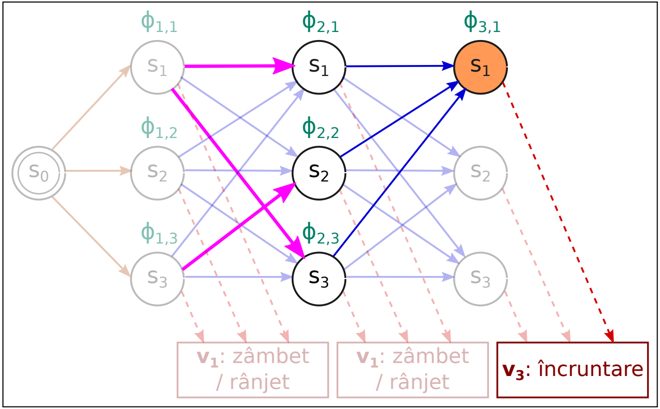
$$\Phi = \begin{bmatrix} -2.1202 & -Inf & -2.1202 \\ -3.7297 & -Inf & -4.2205 \\ & & \end{bmatrix} \quad \Psi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 3 & \\ & & \end{bmatrix} \quad Q_{best} = [\quad]$$

Jovialul



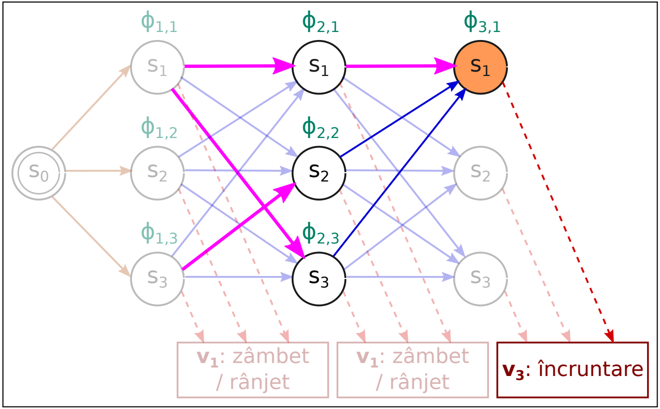
$$\Phi = \begin{bmatrix} -2.1202 & -Inf & -2.1202 \\ -3.7297 & -Inf & -4.2205 \\ & & \end{bmatrix} \Psi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 3 & 1 \\ & & \end{bmatrix} Q_{best} = [\quad]$$

Jovialul



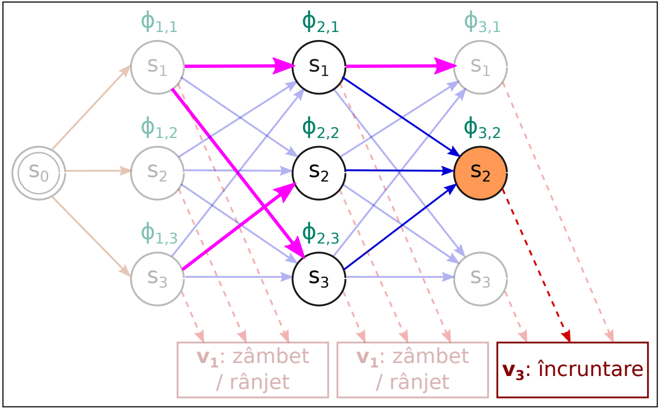
$$\Phi = \begin{bmatrix} -2.1202 & -Inf & -2.1202 \\ -3.7297 & -Inf & -4.2205 \\ -6.7254 & & \end{bmatrix} \quad \Psi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 3 & 1 \\ & & \end{bmatrix} \quad Q_{best} = [\quad]$$

Jovialul



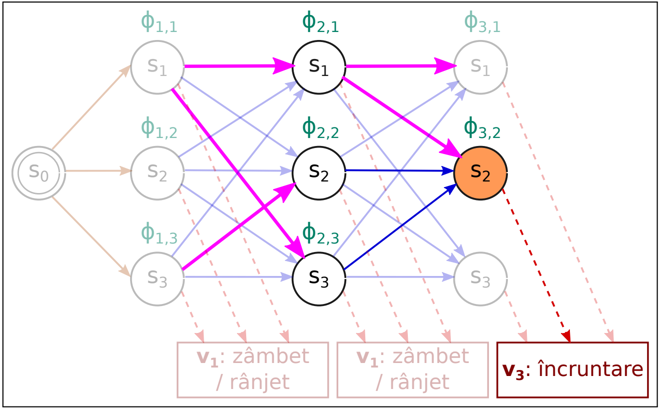
$$\Phi = \begin{bmatrix} -2.1202 & -Inf & -2.1202 \\ -3.7297 & -Inf & -4.2205 \\ -6.7254 & & \end{bmatrix} \quad \Psi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 3 & 1 \\ 1 & & \end{bmatrix} \quad Q_{best} = [\qquad]$$

Jovialul



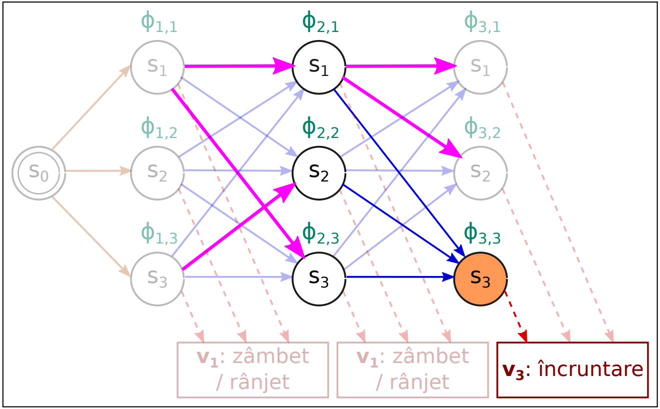
$$\Phi = \begin{bmatrix} -2.1202 & -Inf & -2.1202 \\ -3.7297 & -Inf & -4.2205 \\ -6.7254 & -5.1567 & \end{bmatrix} \quad \Psi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 3 & 1 \\ 1 & & \end{bmatrix} \quad Q_{best} = [\quad]$$

Jovialul



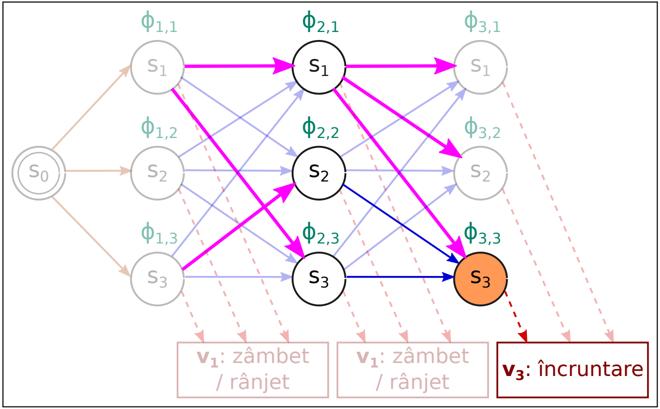
$$\Phi = \begin{bmatrix} -2.1202 & -Inf & -2.1202 \\ -3.7297 & -Inf & -4.2205 \\ -6.7254 & -5.1567 & \end{bmatrix} \quad \Psi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 3 & 1 \\ 1 & 3 & \end{bmatrix} \quad Q_{best} = [\qquad]$$

Jovialul



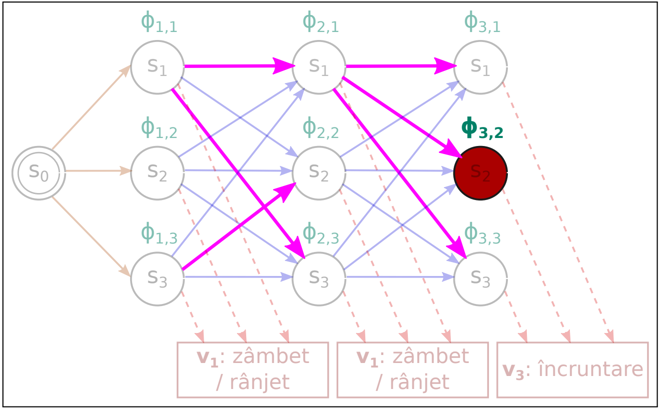
$$\Phi = \begin{bmatrix} -2.1202 & -Inf & -2.1202 \\ -3.7297 & -Inf & -4.2205 \\ -6.7254 & -5.1567 & -5.3391 \end{bmatrix} \quad \Psi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 3 & 1 \\ 1 & 3 & \end{bmatrix} \quad Q_{best} = [\qquad\qquad]$$

Jovialul



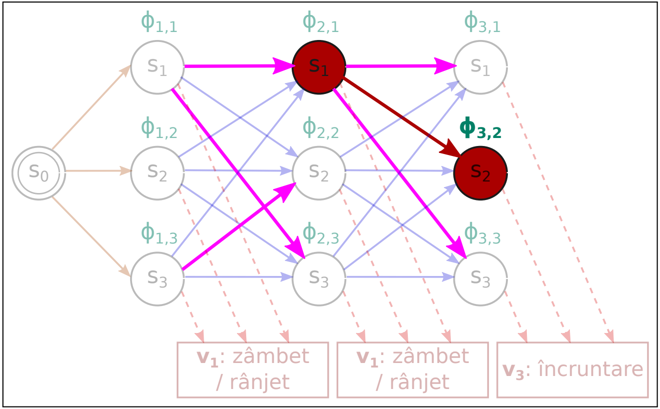
$$\Phi = \begin{bmatrix} -2.1202 & -Inf & -2.1202 \\ -3.7297 & -Inf & -4.2205 \\ -6.7254 & -5.1567 & -5.3391 \end{bmatrix} \quad \Psi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 3 & 1 \\ 1 & 3 & 1 \end{bmatrix} \quad Q_{best} = [\qquad\qquad]$$

Jovialul



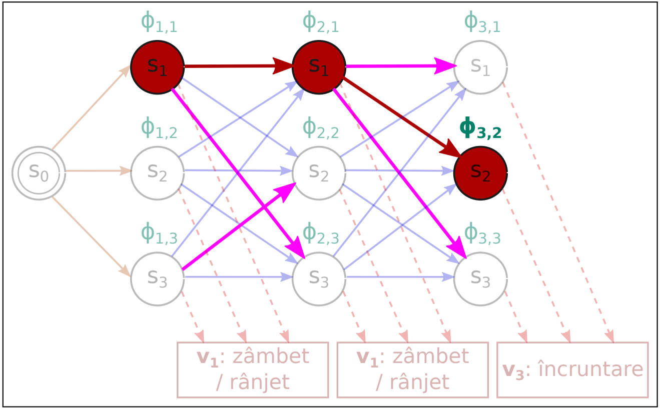
$$\Phi = \begin{bmatrix} -2.1202 & -Inf & -2.1202 \\ -3.7297 & -Inf & -4.2205 \\ -6.7254 & -5.1567 & -5.3391 \end{bmatrix} \quad \Psi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 3 & 1 \\ 1 & 3 & 1 \end{bmatrix} \quad Q_{best} = \begin{bmatrix} & & 2 \end{bmatrix}$$

Jovialul



$$\Phi = \begin{bmatrix} -2.1202 & -Inf & -2.1202 \\ -3.7297 & -Inf & -4.2205 \\ -6.7254 & -5.1567 & -5.3391 \end{bmatrix} \quad \Psi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 3 & 1 \\ 1 & 3 & 1 \end{bmatrix} \quad Q_{best} = \begin{bmatrix} & 1 & 2 \end{bmatrix}$$

Jovialul



$$\Phi = \begin{bmatrix} -2.1202 & -Inf & -2.1202 \\ -3.7297 & -Inf & -4.2205 \\ -6.7254 & -5.1567 & -5.3391 \end{bmatrix} \quad \Psi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 3 & 1 \\ 1 & 3 & 1 \end{bmatrix} \quad Q_{best} = \begin{bmatrix} 1 & 1 & 2 \end{bmatrix}$$

Algoritmul Viterbi

Algoritmul 4 Calculul celei mai probabile secvențe Q_{best}

for $i = 1$ to N **do**
    $\phi_{1,i} \leftarrow \log(\pi_i) + \log(b_i(o_1))$
    $\psi_{1,i} \leftarrow 0$
end for
for $t = 2$ to T **do**
    for $i = 1$ to N **do**
        $\phi_{t,j} \leftarrow [\max_i \phi_{t-1,i} + log(a_{i,j})] + \log(b_j(o_t))$
        $\psi_{t,i} \leftarrow \operatorname{argmax}_i \phi_{t-1,i} + \log(a_{i,j})$
    end for
end for
$\log(P(Q_{\text{best}}|O,\lambda)) \leftarrow \max_i \phi_{T,i}$
$q_{T_{\text{best}}} \leftarrow \operatorname{argmax}_i \phi_{T,i}$
for $t = T - 1$ to 1 **do**
    $q_{t_{\text{best}}} \leftarrow \psi_{t+1}(q_{t+1_{\text{best}}})$
end for

A venit vremea să scriem cod

1. Faceți o copie a fișierului `viterbi_disc.m.stub` și denumiți-o `viterbi_disc.m` (eliminați sufixul `.stub`).
   Veți implementa funcția:
   ```
   function [logP, Q] = viterbi_disc(O,Pi, A, B)
   ```

A venit vremea să scriem cod

1. Faceți o copie a fișierului `viterbi_disc.m.stub` și denumiți-o `viterbi_disc.m` (eliminați sufixul `.stub`).
   Veți implementa funcția:
   `function [logP, Q] = viterbi_disc(O,Pi, A, B)`
2. Completați cele două secțiuni.

```
%% Recursion
% phi_psi_disc-start - Write code below

% phi_psi_disc-end - Write code above
%% logP
[logP, Q(T)] = max(Phi(T, :));
%% Backtracking to compute the path Q
% path_disc-start - Write code below

% path_disc-end - Write code above
```

3. Folosiți `hmm_test` pentru a vă testa codul.

Outline

1. Aplicații în Învățarea Automată pentru MMA
   - Învățarea Automată
   - MMA în Învățarea Automată
2. Teoria MMA
   - Cele Trei Probleme ale MMA
   - Fundamente Matematice
3. Implementarea MMA
   - Problema Evaluării: Algoritmul Forward-Backward
   - Problema Interpretării: Algoritmul Viterbi
   - Problema Estimării: Algoritmul Baum-Welch
4. Demo: Recunoașterea Simbolurilor
5. Tipuri de MMA
6. Discuții și Concluzii

Învătarea din observații - Amintire

Problema Estimării (Antrenării) Modelului

Dându-se un set de secvențe observate $\mathcal{O} = [O_1 O_2 \cdots O_L]$, cum ajustam parameterii $\lambda = (A, B, \Pi)$ ai unui MMA care incearcă să explice cel mai bine acele observații?

Învătarea din observații - Amintire

Problema Estimării (Antrenării) Modelului

Dându-se un set de secvențe observate $\mathcal{O} = [O_1 O_2 \cdots O_L]$, cum ajustam parameterii $\lambda = (A, B, \Pi)$ ai unui MMA care incearcă să explice cel mai bine acele observații?

Secvențele de observații folosite pentru ajustarea parametrilor modelului se numesc secvențe de antrenare.
Problema antrenării este esențiala - ea permite crearea celor mai bune modele pentru fenomene reale.

Învătarea din observații - Abordare

Învătarea din observații - Abordare

Problemă

Nu se cunoaște o metodă analitică de cautare a parametrilor modelului care *maximizeaza* probabilitatea secvențelor observate.

Învătarea din observații - Abordare

Problemă

Nu se cunoaște o metodă analitică de cautare a parametrilor modelului care *maximizeaza* probabilitatea secvențelor observate.

Soluție

Putem totuși găsi $\lambda = (A, B, \Pi)$, astfel încât $\max_\lambda P(O|\lambda)$ corespunde unui maxim local, utilizând o procedură iterativă precum *algoritmul Baum-Welch*. Această metodă este o instanță a *algoritmului EM (Expectation Maximization) [DLR77]* pentru cazul MMA.

Algoritmul Baum-Welch (I)

Procedura în descriere conceptuală:

1. Avem MMA $\lambda = (A, B, \Pi)$ și o secvență observată O

Algoritmul Baum-Welch (I)

Procedura în descriere conceptuală:

1. Avem MMA $\lambda = (A, B, \Pi)$ și o secvență observată O
2. Calculăm folosindu-ne de parametrii $\alpha_t(i)$ și $\beta_t(i)$
   - nr. estimat de tranziții din S_i, pentru fiecare $1 \leq i \leq N$
   - nr. estimat de tranziții din S_i la S_j, pentru fiecare $1 \leq i \leq N, 1 \leq j \leq N$
   - nr. estimat de vizite în S_j observând simbolul v_k, pentru fiecare $1 \leq j \leq N, 1 \leq k \leq M$

Algoritmul Baum-Welch (I)

Procedura în descriere conceptuală:

1. Avem MMA $\lambda = (A, B, \Pi)$ și o secvență observată O
2. Calculăm folosindu-ne de parametrii $\alpha_t(i)$ și $\beta_t(i)$
   - nr. estimat de tranziții din S_i, pentru fiecare $1 \leq i \leq N$
   - nr. estimat de tranziții din S_i la S_j, pentru fiecare $1 \leq i \leq N, 1 \leq j \leq N$
   - nr. estimat de vizite în S_j observând simbolul v_k, pentru fiecare $1 \leq j \leq N, 1 \leq k \leq M$
3. Daca modelul este corect ne așteptăm ca
   (a) Π_i=nr. estimat de vizite în starea S_i la momentul $(t=1) = \bar{\Pi}_i$

   (b) $a_{i,j} = \dfrac{\text{nr. estimat de tranziții din } s_i \text{ la } s_j}{\text{nr. estimat de tranziții din } s_i} = \bar{a_{i,j}}$

   (c) $b_{j,k} = \dfrac{\text{nr. estimat de vizite în } s_j \text{ observând simbolul } v_k}{\text{nr. estimat de vizite în } s_j} = \bar{b_{j,k}}$

Algoritmul Baum-Welch (I)

Procedura în descriere conceptuală:

1. Avem MMA $\lambda = (A, B, \Pi)$ și o secvență observată O
2. Calculăm folosindu-ne de parametrii $\alpha_t(i)$ și $\beta_t(i)$
   - nr. estimat de tranziții din S_i, pentru fiecare $1 \leq i \leq N$
   - nr. estimat de tranziții din S_i la S_j, pentru fiecare $1 \leq i \leq N, 1 \leq j \leq N$
   - nr. estimat de vizite în S_j observând simbolul v_k, pentru fiecare $1 \leq j \leq N, 1 \leq k \leq M$
3. Daca modelul este corect ne așteptăm ca
   (a) Π_i=nr. estimat de vizite în starea S_i la momentul $(t=1) = \bar{\Pi}_i$

   (b) $a_{i,j} = \dfrac{\text{nr. estimat de tranziții din } s_i \text{ la } s_j}{\text{nr. estimat de tranziții din } s_i} = \bar{a}_{i,j}$

   (c) $b_{j,k} = \dfrac{\text{nr. estimat de vizite în } s_j \text{ observând simbolul } v_k}{\text{nr. estimat de vizite în } s_j} = \bar{b}_{j,k}$
4. Vedem că raporturile calculate din presupunerea noastră explică mai bine observația decât parametrii anteriori, i.e. $P(O|\bar{\lambda}) > P(O|\lambda)$

Algoritmul Baum-Welch (I)

Procedura în descriere conceptuală:

1. Avem MMA $\lambda = (A, B, \Pi)$ și o secvență observată O
2. Calculăm folosindu-ne de parametrii $\alpha_t(i)$ și $\beta_t(i)$
   - nr. estimat de tranziții din S_i, pentru fiecare $1 \leq i \leq N$
   - nr. estimat de tranziții din S_i la S_j, pentru fiecare $1 \leq i \leq N, 1 \leq j \leq N$
   - nr. estimat de vizite în S_j observând simbolul v_k, pentru fiecare $1 \leq j \leq N, 1 \leq k \leq M$
3. Daca modelul este corect ne așteptăm ca
   (a) Π_i=nr. estimat de vizite în starea S_i la momentul $(t=1) = \bar{\Pi}_i$

   (b) $a_{i,j} = \dfrac{\text{nr. estimat de tranziții din } s_i \text{ la } s_j}{\text{nr. estimat de tranziții din } s_i} = \bar{a_{i,j}}$

   (c) $b_{j,k} = \dfrac{\text{nr. estimat de vizite în } s_j \text{ observând simbolul } v_k}{\text{nr. estimat de vizite în } s_j} = \bar{b_{j,k}}$
4. Vedem că raporturile calculate din presupunerea noastră explică mai bine observația decât parametrii anteriori, i.e. $P(O|\bar{\lambda}) > P(O|\lambda)$
5. Atunci repetăm procesul până ce suntem mulțumiți (convergență: $P(O|\bar{\lambda}) - P(O|\lambda) \leq \epsilon$)

Algoritmul Baum-Welch (II)

Algoritmul Baum-Welch (II)

Definim întâi niște variabile auxiliare:

$\xi_{t,i,j} = \xi_t(i,j) = P(q_t = s_i, q_{t+1} = s_j | O, \lambda)$
Probabilitatea de a fi în starea s_i la momentul t **și** în starea s_j la momentul $t+1$, condiționat de parametrii modelului curent și secvența observată.

Algoritmul Baum-Welch (II)

Definim întâi niște variabile auxiliare:

$\xi_{t,i,j} = \xi_t(i,j) = P(q_t = s_i, q_{t+1} = s_j | O, \lambda)$
Probabilitatea de a fi în starea s_i la momentul t **și** în starea s_j la momentul $t+1$, condiționat de parametrii modelului curent și secvența observată.

$\gamma_{t,i} = \gamma_t(i) = P(q_t = s_i | O, \lambda)$
Probabilitatea de a fi în starea s_i la momentul t, condiționat de parametrii modelului curent și secvența observată.

Algoritmul Baum-Welch (II)

Definim întâi niște variabile auxiliare:

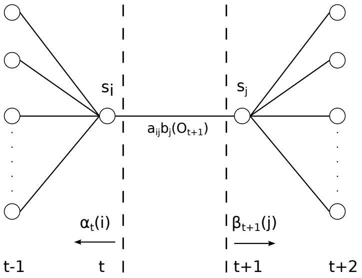
$\xi_{t,i,j} = \xi_t(i,j) = P(q_t = s_i, q_{t+1} = s_j | O, \lambda)$
Probabilitatea de a fi în starea s_i la momentul t **și** în starea s_j la momentul $t+1$, condiționat de parametrii modelului curent și secvența observată.

$\gamma_{t,i} = \gamma_t(i) = P(q_t = s_i | O, \lambda)$
Probabilitatea de a fi în starea s_i la momentul t, condiționat de parametrii modelului curent și secvența observată.

Din definiții rezultă că:

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j)$$

Algoritmul Baum-Welch (III)



Secvența de operații necesară pentru calculul evenimentului mixt ca sistemul se află în starea S_i la momentul t și în starea S_j la momentul $t+1$

$$\alpha_{t,i} = P(o_1, o_2, \ldots, o_t, q_t = S_i|\lambda) \quad (36)$$

$$\beta_{t,i} = P(o_{t+1}o_{t+2}\cdots o_T|q_t = S_i, \lambda) \quad (37)$$

$$\xi_t(i,j) = \frac{\alpha_{t,i}\cdot a_{i,j}\cdot b_j(o_{t+1})\cdot \beta_{t+1,j}}{P(O|\lambda)} = \frac{\alpha_{t,i}\cdot a_{i,j}\cdot b_j(o_{t+1})\cdot \beta_{t+1,j}}{\sum_{k=1}^{N}\sum_{l=1}^{N}\alpha_{t,k}\cdot a_{k,l}\cdot b_l(o_{t+1})\cdot \beta_{t+1,l}} \quad (38)$$

Algoritmul Baum-Welch (IV)

Cum ne ajută aceste variabile auxiliare?

$$\sum_{t=1}^{T-1} \gamma_t(i) = \text{numărul estimat de tranziții din } S_i$$

$$\sum_{t=1}^{T-1} \xi_t(i,j) = \text{numărul estimat de tranziții din } S_i \text{ la } S_j$$

Algoritmul Baum-Welch (V)

$$\bar{\pi}_i = \text{ nr. estimat de vizite în starea } S_i \text{ la momentul } (t = 1) = \gamma_1(i) \tag{39}$$

Algoritmul Baum-Welch (V)

$$\bar{\pi}_i = \text{ nr. estimat de vizite în starea } S_i \text{ la momentul } (t=1) = \gamma_1(i) \tag{39}$$

$$\begin{aligned} \bar{a_{i,j}} &= \frac{\text{nr. estimat de tranziții din } S_i \text{ la } S_j}{\text{nr. estimat de tranziții din } S_i} \\ &= \frac{\displaystyle\sum_{t=1}^{T-1} \xi_t(i,j)}{\displaystyle\sum_{t=1}^{T-1} \gamma_t(i)} \end{aligned} \tag{40}$$

Algoritmul Baum-Welch (V)

$$\bar{\pi}_i = \text{ nr. estimat de vizite în starea } S_i \text{ la momentul } (t = 1) = \gamma_1(i) \tag{39}$$

$$\begin{aligned} \bar{a_{i,j}} &= \frac{\text{nr. estimat de tranziții din } S_i \text{ la } S_j}{\text{nr. estimat de tranziții din } S_i} \\ &= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \end{aligned} \tag{40}$$

$$\begin{aligned} \bar{b_{j,k}} &= \frac{\text{nr. estimat de vizite în } S_j \text{ observând simbolul } v_k}{\text{nr. estimat de vizite în } S_j} \\ &= \frac{\sum_{t=1, O_t = v_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)} \end{aligned} \tag{41}$$

Algoritmul Baum-Welch (VI)

Algoritmul 5 Algoritm Baum-Welch

intrări: $O \leftarrow$ secvența de observații, $\epsilon \leftarrow$ prag de convergență

{*Initializare*}
init. uniformă Π ($\Pi_i = 1/N, 1 \leq i \leq N$)
init. aleatoare $a_{i,j}$, a. î. $\sum_{j=1}^{N} a_{i,j} = 1, \forall i = \overline{1,N}$
init. uniformă $b_{j,k}$ ($b_{j,k} = 1/M, 1 \leq j \leq N, 1 \leq k \leq M$)
init $log(P(O|\bar{\lambda})) = 0$

repeat
    $log(P(O|\lambda)) = log(P(O|\bar{\lambda}))$

    {*E STEP - calculeaza variantele scalate pentru α și β și probabilitatea curentă (log likelihood - $log(P(O|\bar{\lambda}))$) a secvenței observate*}
    $[log(P(O|\bar{\lambda})), \hat{\alpha}, \hat{\beta}, Scale] = forward\_backward(O, \Pi, A, B)$

    {M STEP - recalculeaza Π, A și B}
    $\Pi = update\_pi\_procedure(\hat{\alpha}, \hat{\beta}, Scale)$
    $A = update\_A\_procedure(O, \hat{\alpha}, \hat{\beta}, Scale)$
    $B = update\_B\_procedure(O, \hat{\alpha}, \hat{\beta}, Scale)$
until $log(P(O|\bar{\lambda})) - log(P(O|\lambda)) < \epsilon$

Algoritmul Baum-Welch (VI)

Algoritmul 6 Algoritm Baum-Welch

Function update_pi_procedure($\hat{\alpha}$, $\hat{\beta}$, *Scale*)
for $i = 1$ to N **do**
$$\Pi_i = \frac{\hat{\alpha}_1(i) \cdot \hat{\beta}_1(i) / Scale(1)}{\sum_{j=1}^{N} \hat{\alpha}_1(j) \cdot \hat{\beta}_1(j) / Scale(1)}$$
end for
return Π
EndFunction

Function update_A_procedure(O, $\hat{\alpha}$, $\hat{\beta}$, *Scale*)
for $i = 1$ to N **do**
  for $j = 1$ to N **do**
$$a_{i,j} = \frac{\sum_{t=1}^{T-1} \hat{\alpha}_{t,i} \cdot a_{ij} \cdot b_j(o_{t+1}) \cdot \hat{\beta}_{t+1,j}}{\sum_{t=1}^{T-1} \sum_{j=1}^{N} \hat{\alpha}_{t,i} \cdot a_{i,j} \cdot b_j(o_{t+1}) \cdot \hat{\beta}_{t+1,j}}$$
  end for
end for
return a
EndFunction

Algoritmul Baum-Welch (VI)

Algoritmul 7 Algoritm Baum-Welch

Function update_B_procedure(O, $\hat{\alpha}$, $\hat{\beta}$, *Scale*)
for $j = 1$ to N **do**
    for $k = 1$ to M **do**
        $b_{j,k} = \frac{\sum_{t=1,O(t)=v_k}^{T} \hat{\alpha}_t(j) \cdot \hat{\beta}_t(j) / Scale(t)}{\sum_{t=1}^{T} \hat{\alpha}_t(j) \cdot \hat{\beta}_t(j) / Scale(t)}$
    end for
end for
return b
EndFunction

Baum-Welch - Să scriem niște cod

LET'S WRITE SOME CODE :-)

Aplicație de recunoaștere a simbolurilor

Features:

Aplicație de recunoaștere a simbolurilor

Features:

Definire

Definire, organizare și vizualizare a unui set de date de simboluri captate prin mișcarea mouse-ului.

Aplicație de recunoaștere a simbolurilor

Features:

Definire

Definire, organizare și vizualizare a unui set de date de simboluri captate prin mișcarea mouse-ului.

Antrenare

Antrenarea unui motor de recunoaștere a simbolurilor prin metoda MMA.

Aplicație de recunoaștere a simbolurilor

Features:

Definire

Definire, organizare și vizualizare a unui set de date de simboluri captate prin mișcarea mouse-ului.

Antrenare

Antrenarea unui motor de recunoaștere a simbolurilor prin metoda MMA.

Recunoaștere

Recunoașterea unui nou simbol si vizualizarea unor metrici de clasificare.

Aplicație de recunoaștere a simbolurilor

Features:

Definire

Definire, organizare și vizualizare a unui set de date de simboluri captate prin mișcarea mouse-ului.

Antrenare

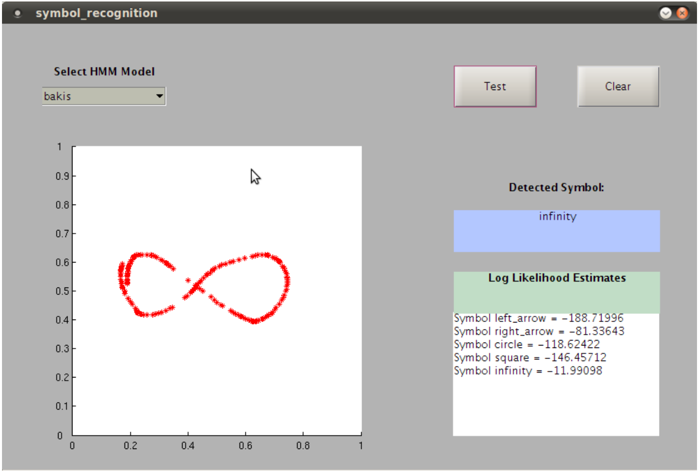
Antrenarea unui motor de recunoaștere a simbolurilor prin metoda MMA.

Recunoaștere

Recunoașterea unui nou simbol si vizualizarea unor metrici de clasificare.

Simboluri default incluse: **săgeată stânga, săgeată dreapta, cerc, pătrat, infinit**

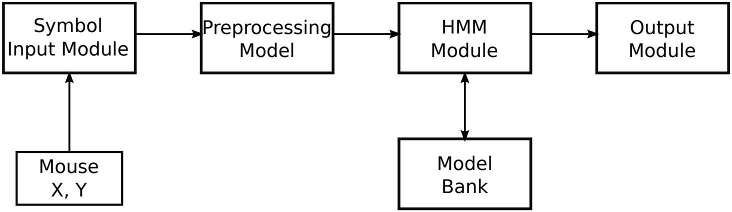
Aplicație de recunoaștere a simbolurilor - Preview



O vizualizare cu GUI-ul aplicației de recunoaștere a simbolurilor

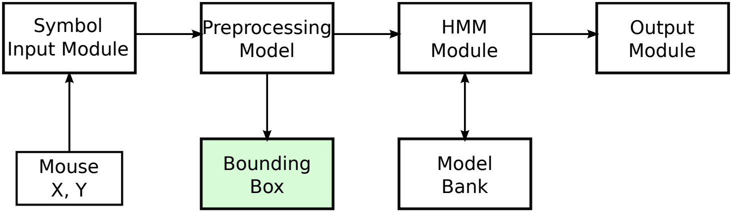
Aplicație de recunoaștere a simbolurilor - Abordare

Adapted from [YX94].



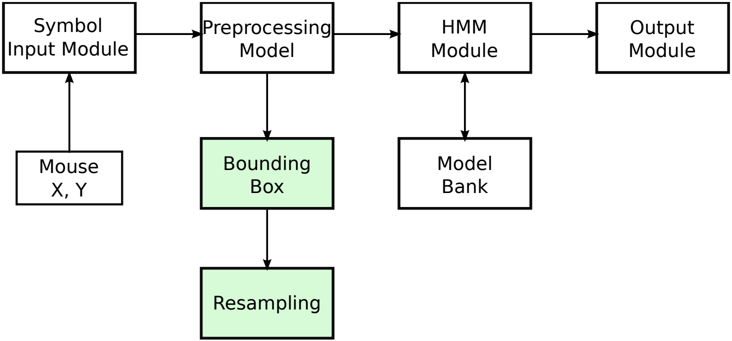
Aplicație de recunoaștere a simbolurilor - Abordare

Adapted from [YX94].



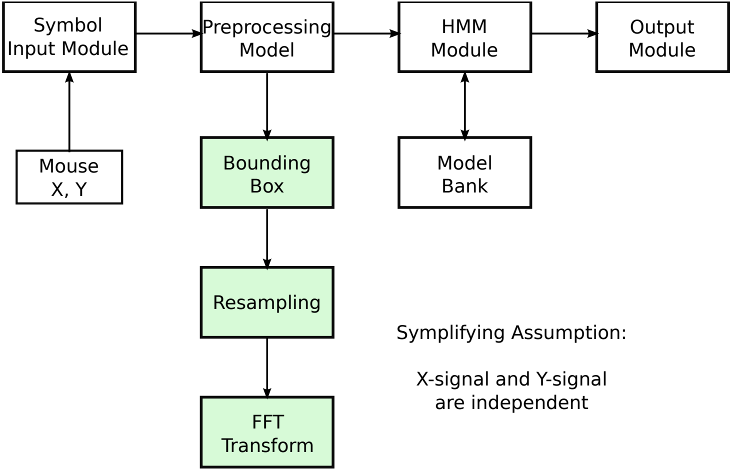
Aplicație de recunoaștere a simbolurilor - Abordare

Adapted from [YX94].



Aplicație de recunoaștere a simbolurilor - Abordare

Adapted from [YX94].



Symplifying Assumption:

X-signal and Y-signal are independent

Aplicație de recunoaștere a simbolurilor - Abordare

Adaptare după [YX94].

Structura MMA

N(număr de stări) $= 8$
2 variabile observabile de tip discrte per stare - $coef_{FFT}(x)$, $coef_{FFT}(y)$
M(număr de valori pentru fiecare variabilă observabilă) $= 256$
Modelul de tranziție:

- Bakis
- Ergodic

Aplicație de recunoaștere a simbolurilor - Abordare

Procedură de recunoaștere:

1. Construim și antrenăm câte un MMA pentru fiecare simbol în parte

Aplicație de recunoaștere a simbolurilor - Abordare

Procedură de recunoaștere:

1. Construim și antrenăm câte un MMA pentru fiecare simbol în parte
2. Folosim un set de date (simboluri) de validare pentru a stabili *praguri* de recunoaștere, i.e. dacă probabilitatea secvenței observate este prea mică pentru fiecare MMA, marcăm simbolul drept *necunoscut*

Aplicație de recunoaștere a simbolurilor - Abordare

Procedură de recunoaștere:

1. Construim și antrenăm câte un MMA pentru fiecare simbol în parte
2. Folosim un set de date (simboluri) de validare pentru a stabili *praguri* de recunoaștere, i.e. dacă probabilitatea secvenței observate este prea mică pentru fiecare MMA, marcăm simbolul drept *necunoscut*
3. Recunoaștere:
   - Calculăm $P(O|\lambda_i)$ pentru fiecare MMA construit pentru simbolurile $i = 1, \cdots, nr\_simboluri$
   - Alegem $\max(P(O|\lambda_i))$ ca și simbol candidat. Daca $P(O|\lambda_i) > prag_i$, atunci am recunoscut *simbolul i*, altfel marcăm *necunoscut*

Aplicație de recunoaștere a simbolurilor - Rezultate

Mărime set de date

5 simboluri: **săgeată stânga, săgeată dreapta, cerc, pătrat, infinit**
100 exemple per simbol: **50** antrenare, **10** validare, **40** testare

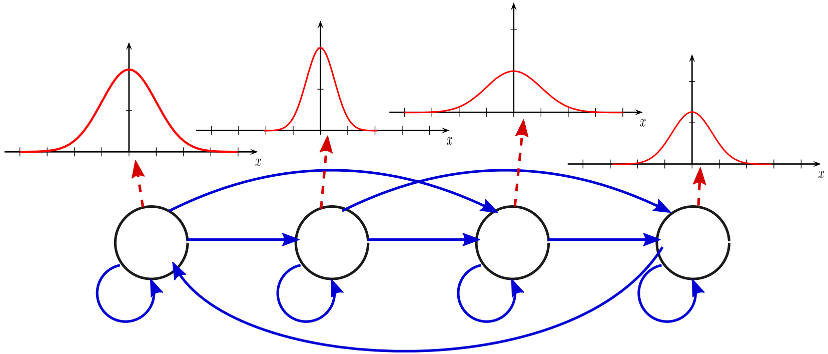
```
>> symbol_performance_test('ergodic')
========= Testing trained HMM models ========
## Results for the model of symbol "left_arrow":
        Accuracy: 0.97500
        Precision: 1.00000
        Recall: 0.97500
        Confusion matrix line:    39    0    1    0    0    0

## Results for the model of symbol "right_arrow":
        Accuracy: 1.00000
        Precision: 1.00000
        Recall: 1.00000
        Confusion matrix line:     0   40    0    0    0    0

## Results for the model of symbol "circle":
        Accuracy: 0.90244
        Precision: 0.97368
        Recall: 0.92500
        Confusion matrix line:     0    0   37    2    1    0

## Results for the model of symbol "square":
        Accuracy: 0.95238
        Precision: 0.95238
        Recall: 1.00000
        Confusion matrix line:     0    0    0   40    0    0

## Results for the model of symbol "infinity":
        Accuracy: 0.97561
        Precision: 0.97561
        Recall: 1.00000
        Confusion matrix line:     0    0    0    0   40    0
```

```
>> symbol_performance_test('bakis')
========= Testing trained HMM models ========
## Results for the model of symbol "left_arrow":
        Accuracy: 0.90000
        Precision: 1.00000
        Recall: 0.90000
        Confusion matrix line:    36    0    1    0    0    3

## Results for the model of symbol "right_arrow":
        Accuracy: 1.00000
        Precision: 1.00000
        Recall: 1.00000
        Confusion matrix line:     0   40    0    0    0    0

## Results for the model of symbol "circle":
        Accuracy: 0.97561
        Precision: 0.97561
        Recall: 1.00000
        Confusion matrix line:     0    0   40    0    0    0

## Results for the model of symbol "square":
        Accuracy: 0.97500
        Precision: 1.00000
        Recall: 0.97500
        Confusion matrix line:     0    0    0   39    0    1

## Results for the model of symbol "infinity":
        Accuracy: 1.00000
        Precision: 1.00000
        Recall: 1.00000
        Confusion matrix line:     0    0    0    0   40    0
```

Emisii continue

- Probabilitatea emisiilor este modelată cu o lege de distribuție continuă (e.g. normală).

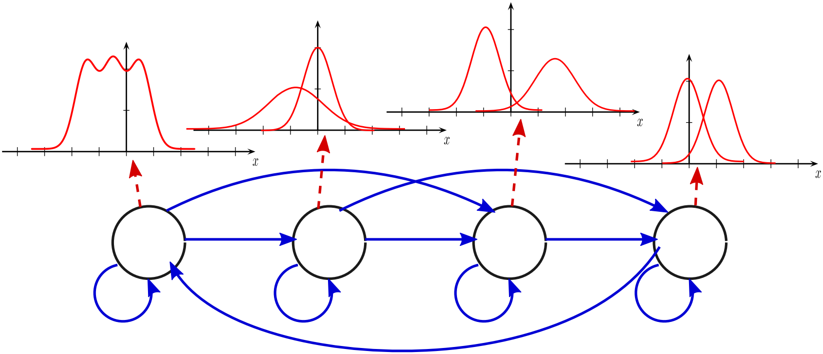
$$b_j(o) = \mathcal{N}(o, \mu_j, \sigma_j^2) \tag{42}$$



Mixturi Gaussiene

- Probabilitatea emisiilor este modelată printr-o mixtură de distribuții normale.

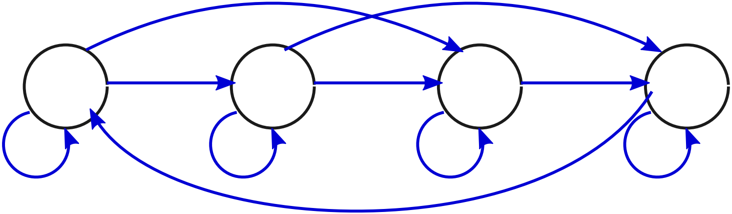
$$b_j(o) = \sum_{k=1}^{M} c_{j,k}\mathcal{N}(o, \mu_{j,k}, \sigma_{j,k}^2) \tag{43}$$



Modele Markov Ergodice

Modele Markov Ergodice

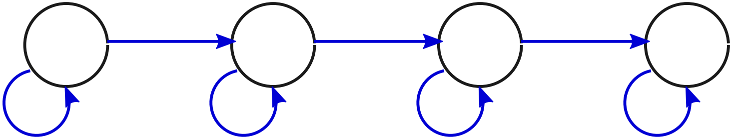
Un lanț Markov se numește ergodic dacă din orice stare se poate ajunge în oricare altă stare (nu neaparat într-o singură mutare).
Lanțurile Markov ergodice se mai numesc și *ireductibile*.



Modelul Bakis

Modele Markov Bakis (stânga ⟶ dreapta)

Un model Bakis este unul în care nu este permisă tranziția dintr-o stare către o altă stare cu un indice mai mic.



[DLR77] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.

[KF09] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

[Mit97] Thomas M. Mitchell. *Machine Learning*. McGraw-Hill, Inc., New York, NY, USA, 1 edition, 1997.

[Rab89] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[YX94] J. Yang and Y. Xu. Hidden markov model for gesture recognition. Technical report, DTIC Document, 1994.

[Zub06] R. Zubek. Introduction to hidden markov models. *AI Game Programming Wisdom*, 3:633–646, 2006.

Thank you!

Baftă, șailor!