Análise Estatística de dados

Inteligência Artificial

PECE Programa de Educação Continuada
Escola Politécnica da USP

USP

AULA 10 – MARKOV

Arturo Forner-Cordero
Larissa Driemeier

PROGRAMA DO CURSO

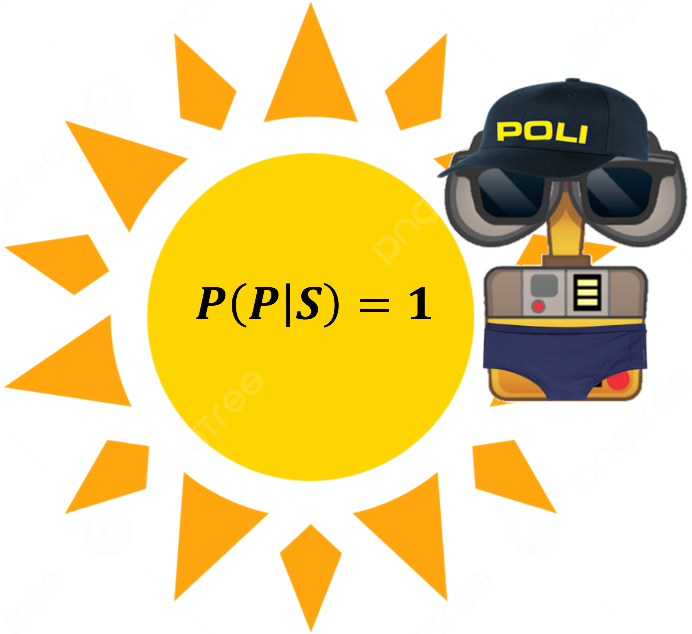
Aula	Data	Conteúdo da Aula
01	27/02	Aula Inaugural
02	05/03	Introdução ao Curso. Noções de Álgebra Linear , Geometria Analítica Parte I
03	12/03	Noções de Álgebra Linear , Geometria Analítica Parte II
04	19/03	Decomposição de valor singular (SVD)
05	26/03	Otimização: derivadas, derivadas parciais (operadores gradiente, Jacobiano, Hessiano e Laplaciano), algoritmos de gradiente
06	02/04	Variáveis independentes e não independentes. Estatística Descritiva e Indutiva. Definições de medidas de dispersão e tendência central.
07	09/04	Probabilidade e Teorema de Bayes
08	16/04	Modelos de probabilidade discretos
09	23/04	Modelos de probabilidade contínuos
10	30/04	Modelo de Markov. Modelo de Markov oculto.

PROCESSOS ESTOCÁSTICOS

Frequentemente, são feitas observações em um período de tempo, influenciadas por efeitos aleatórios, não só em um único instante, mas por todo o intervalo de tempo ou sequência de tempos que se está considerando. Essa situação é denominada um **Processo Estocástico.**

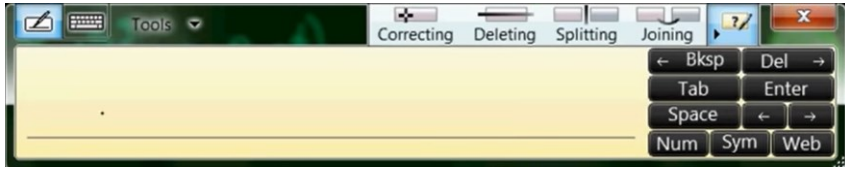
Processos Estocásticos são usados para descrever um sistema operando sobre algum período de tempo. Existem vários modelos de Processos Estocásticos, porém, abordaremos o **Modelo de Markov**.

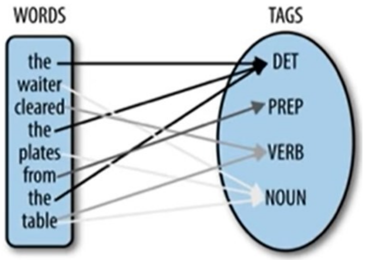
Andrei Andreyevich Markov
(*1856, Ryazan, Russia; †1922, São Petersburgo, Russia).

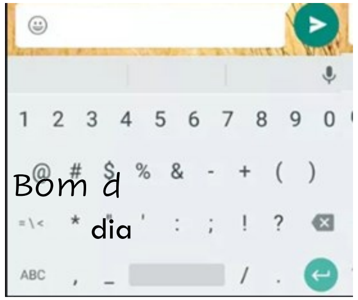


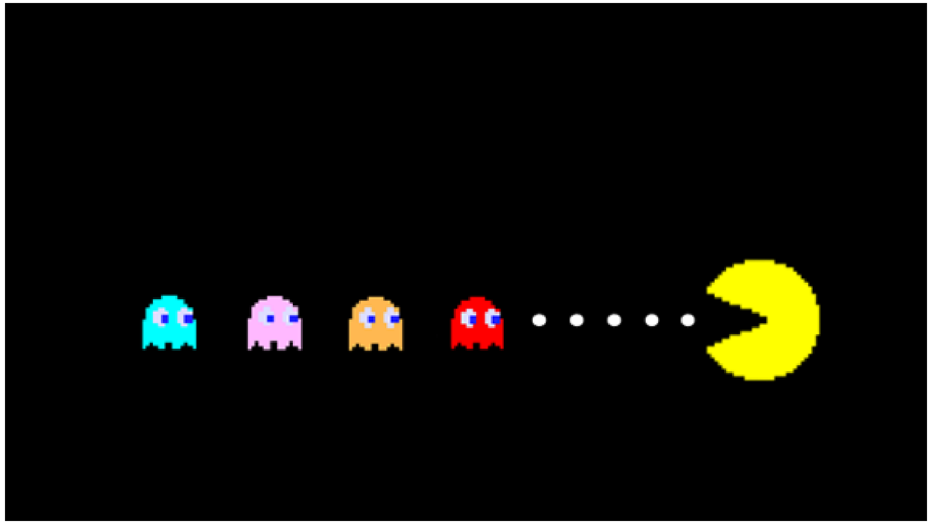
MARKOV

A aplicação do Modelo de Markov oculto (*Hidden Markov Models*) inclui aprendizado por reforço e reconhecimento de padrões *temporais*, como fala, escrita, reconhecimento de gestos, marcação de parte do discurso, acompanhamento de partituras, ...



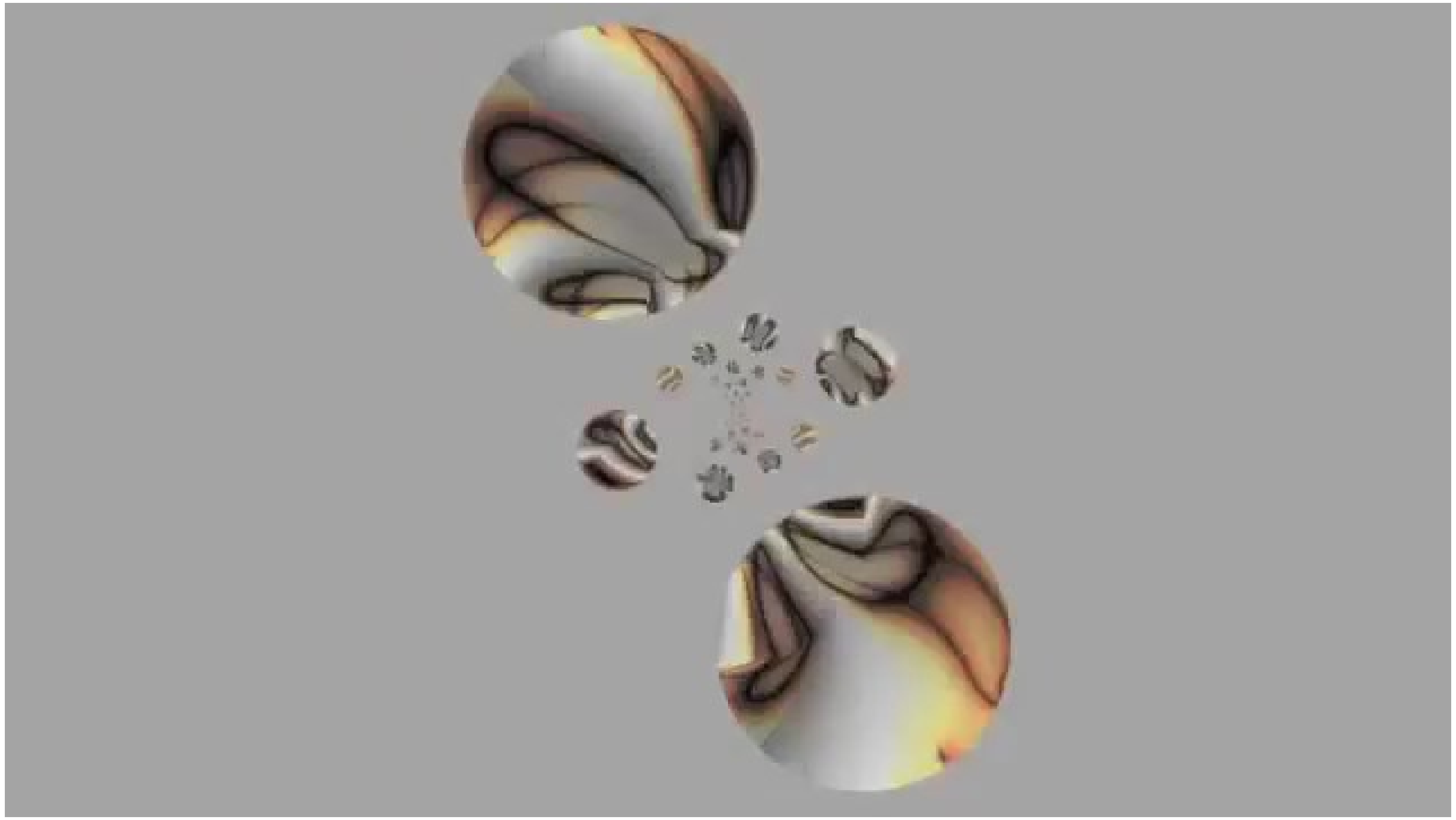


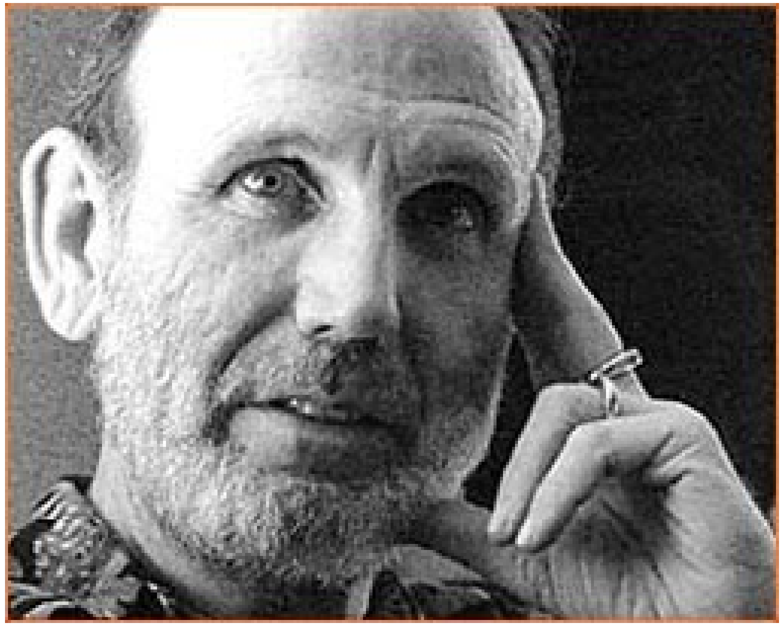




MÚSICA?

David Cope usou seu *programa Experiments in Musical Intelligence* para compor Zodiac, doze obras curtas para orquestra de cordas no estilo de Vivaldi. Este é o Touro. O vídeo também é criado algoritmicamente.





https://www.youtube.com/watch?v=2kuY3BrmTfQ&feature=youtu.be

http://www.thesoundstew.com/2010/07/interview-with-david-cope.html

MODELO DE MARKOV

Na teoria da probabilidade, um modelo de Markov é um modelo estocástico usado para modelar sistemas de mudança aleatória onde se supõe que os estados futuros dependem apenas do estado atual e não dos eventos que ocorreram antes dele.

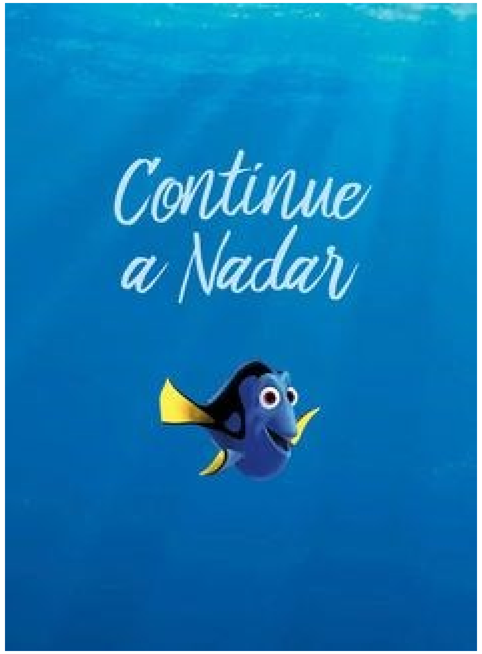
Markov é baseado em um fato intrínseco da vida: **o futuro é independente do passado, dado o presente.**

EM OUTRAS PALAVRAS...

Um processo estocástico tem a propriedade de Markov se a distribuição de probabilidade condicional dos estados futuros do processo (condicional nos estados passado e presente) depende apenas do estado presente, não da sequência de eventos que o precederam.

Um processo com essa propriedade é chamado de processo de Markov.

Este tipo de Processo Estocástico é também denominado memoryless process (processo sem memória), uma vez que o passado é "esquecido" (desprezado).

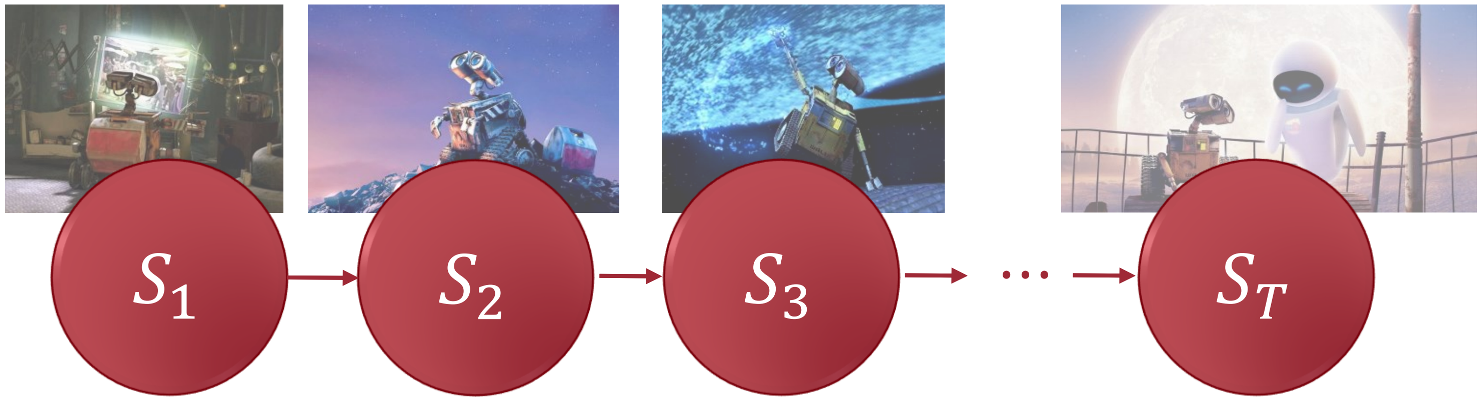


PECE Programa de Educação Continuada
Escola Politécnica da USP

TEMPO DISCRETO

Vamos pensar em uma sequência de estados de tempo discretos estocásticos $S_1, S_2, \ldots, S_T$ **com a propriedade de Markov.** O conjunto de estados define um episódio finito com dimensão T ,

$$\boldsymbol{S} = [\, S_1 \quad S_2 \quad \cdots \quad S_T \,]$$



UM POUCO DE NOTAÇÃO MATEMÁTICA

$$S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow \cdots \rightarrow S_T$$

$$\boldsymbol{S} = [\, S_1 \quad S_2 \quad \cdots \quad S_T]$$

$$P(\boldsymbol{S}) = P(\, S_1, S_2, \cdots, S_T)?$$

$$P(\, S_1, S_2) = P(\, S_2 |\, S_1) P(\, S_1)$$

$$P(\, S_1, S_2, \cdots, S_T) = P(\, S_T | S_1, S_2, \cdots, S_{T-1}) {\color{red} P(S_1, S_2, \cdots, S_{T-1})}$$

$${\color{red} P(S_1, S_2, \cdots, S_{T-1}) = P(\, S_{T-1} | S_1, S_2, \cdots, S_{T-2})} {\color{green} P(S_1, S_2, \cdots, S_{T-2})}$$

$${\color{green} P(S_1, S_2, \cdots, S_{T-2}) = P(\, S_{T-2} | S_1, S_2, \cdots, S_{T-3})} {\color{blue} P(S_1, S_2, \cdots, S_{T-3})}$$

$$P(\boldsymbol{S}) = P(\, S_T | S_1, S_2, \cdots, S_{T-1}) P(\, S_{T-1} | S_1, S_2, \cdots, S_{T-2}) \ldots P(S_1)$$

MARKOV DE PRIMEIRA ORDEM

Um estado S_t é de Markov se, e somente se, satisfaz a **propriedade de Markov:**

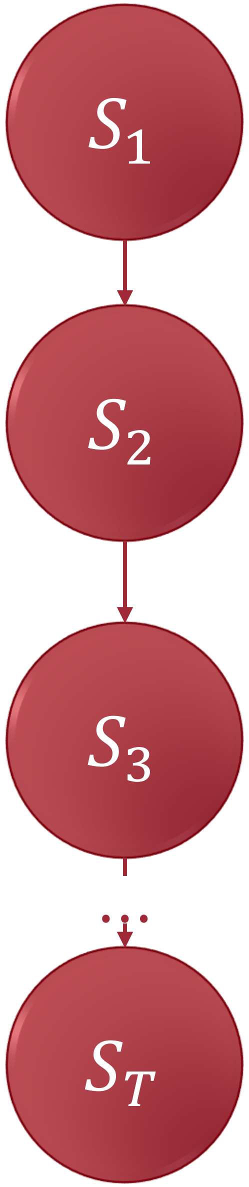
$$P(S_t|S_1, S_2, \cdots, S_{t-1}) = P(S_t|S_{t-1})$$

Então... Se havíamos definido

$$P(\boldsymbol{S}) = P(S_T|S_1, S_2, \cdots, S_{T-1})P(S_{T-1}|S_1, S_2, \cdots, S_{T-2}) \ldots P(S_1)$$

Podemos reescrever como

$$P(\boldsymbol{S}) = P(S_T|S_{T-1})P(S_{T-1}|S_{T-2}) \ldots P(S_1)$$



PROBABILIDADE DE EMISSÃO

$$S_1 \to S_2 \to S_3 \to \cdots \to S_T$$

Probabilidade de emissão do estado $\boldsymbol{S}$

$$P(\boldsymbol{S}) = P(S_1) \prod_{i=2}^{n} P(S_i \mid S_{i-1})$$

Condição Inicial

Probabilidade de emissão $a_{S_i, S_{i-1}} = P(S_i \mid S_{i-1})$

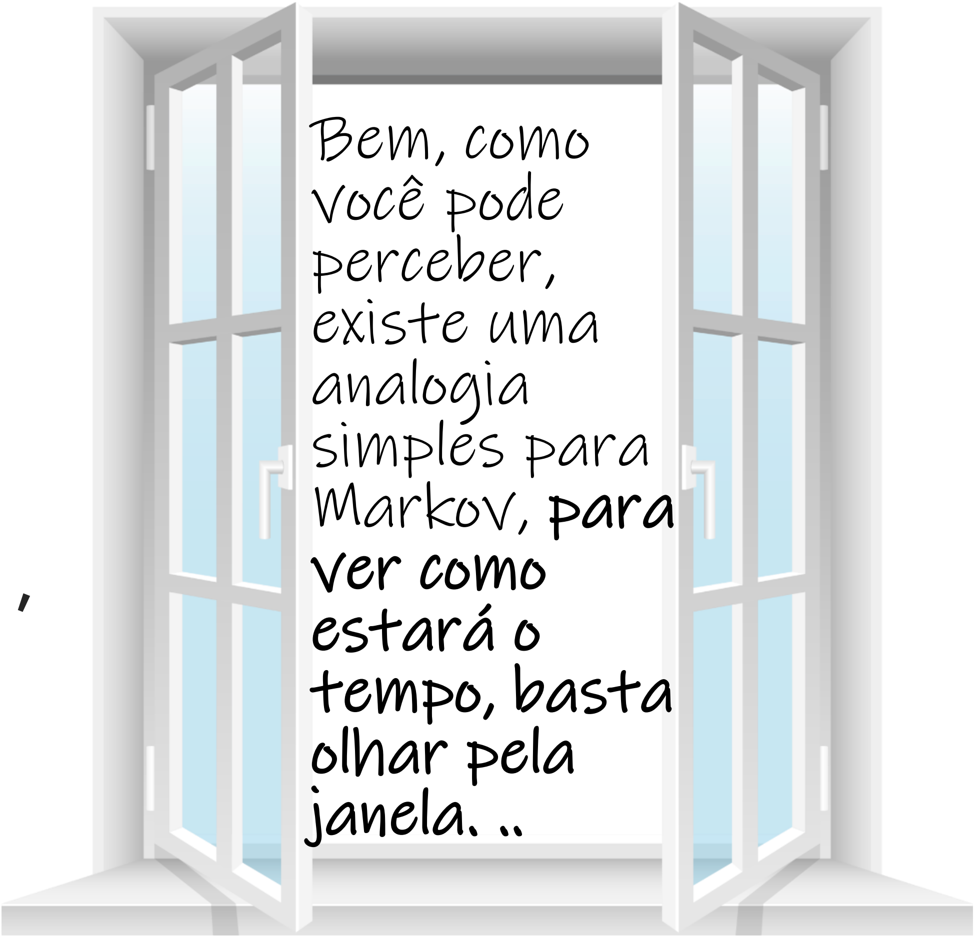
PECE Programa de Educação Continuada
Escola Politécnica da USP

PROPRIEDADE DE MARKOV

O estado S_t é uma estatística suficiente do futuro pois captura todas as informações relevantes do histórico.

Uma vez que o estado é conhecido, a história pode ser jogada fora.

Isto é, pode-se tomar decisões com base apenas em S_t , sem a necessidade de conhecer como o estado S_t foi alcançado.



‘E AGORA, JOSÉ?’
CARLOS DRUMMOND DE ANDRADE

Se você gritasse,
se você gemesse,
se você tocasse
a valsa vienense,
se você dormisse,
se você cansasse,
se você morresse...
Mas você não morre,
você é duro, José!

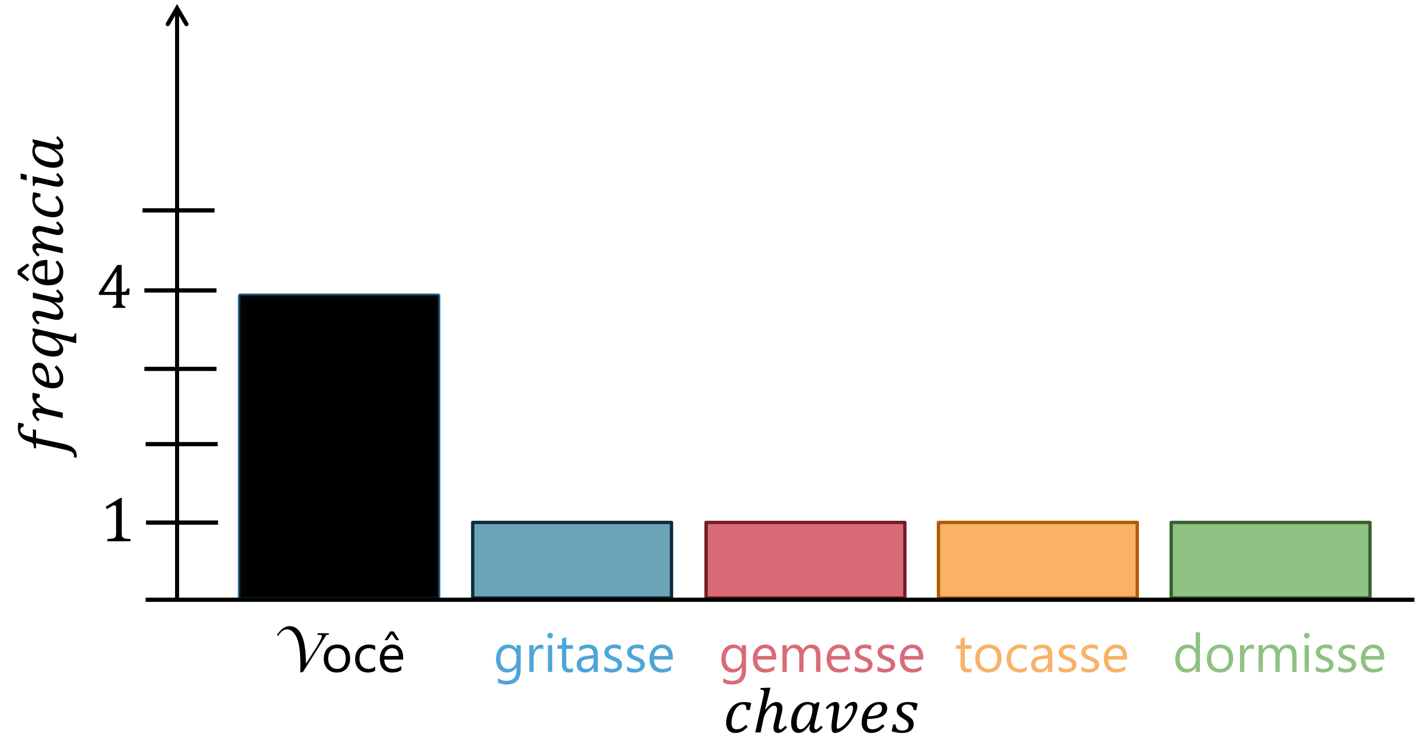
As cadeias de Markov são geralmente definidas por um conjunto de estados e pelas probabilidades de transição entre cada estado.

$\mathcal{V}$ocê gritasse $\mathcal{V}$ocê gemesse $\mathcal{V}$ocê tocasse $\mathcal{V}$ocê dormisse

Você gritasse Você gemesse Você tocasse Você dormisse

oito palavras (**tokens**), mas apenas cinco palavras únicas (**chaves**)

Você	4
Gritasse	1
Gemesse	1
Tocasse	1
dormisse	1



Você gritasse Você gemesse Você tocasse Você dormisse

START	1
Você	4
Gritasse	1
Gemesse	1
Tocasse	1
Dormisse	1
END	1

Cada frase é precedida por um símbolo invisível *START* e sempre termina com um símbolo *END*.

VAMOS APLICAR O MODELO DE MARKOV

START Você gritasse Você gemesse Você tocasse Você dormisse *END*

(*START*, Você)
(Você, gritasse)
(gritasse, Você)
(Você, gemesse)
(gemesse, Você)
(Você, tocasse)
(tocasse, Você)
(Você,dormisse)
(dormisse,*END*)
(*END*, none)

(*START*,Você)
(Você, gritasse) (Você, gemesse) (Você, tocasse) (Você, dormisse)
(gritasse, Você)
(gemesse, Você)
(tocasse, Você)
(dormisse, *END*)
(*END*, none)

Cada chave tem palavras possíveis que podem segui-la. Se déssemos essa estrutura para alguém, esse alguém poderia, com certeza, recriar nossa frase original???

VAMOS TENTAR...

START Você dormisse *END*

(*START*, Você)
(Você, gritasse) (Você, gemesse) (Você, tocasse) (Você, dormisse)
(gritasse, Você)
(gemesse, Você)
(tocasse, Você)
(dormisse, *END*)
(*END*, none)

(*START*, Você)

(Você, gritasse) (Você, gemesse) (Você, tocasse) (Você, dormisse)

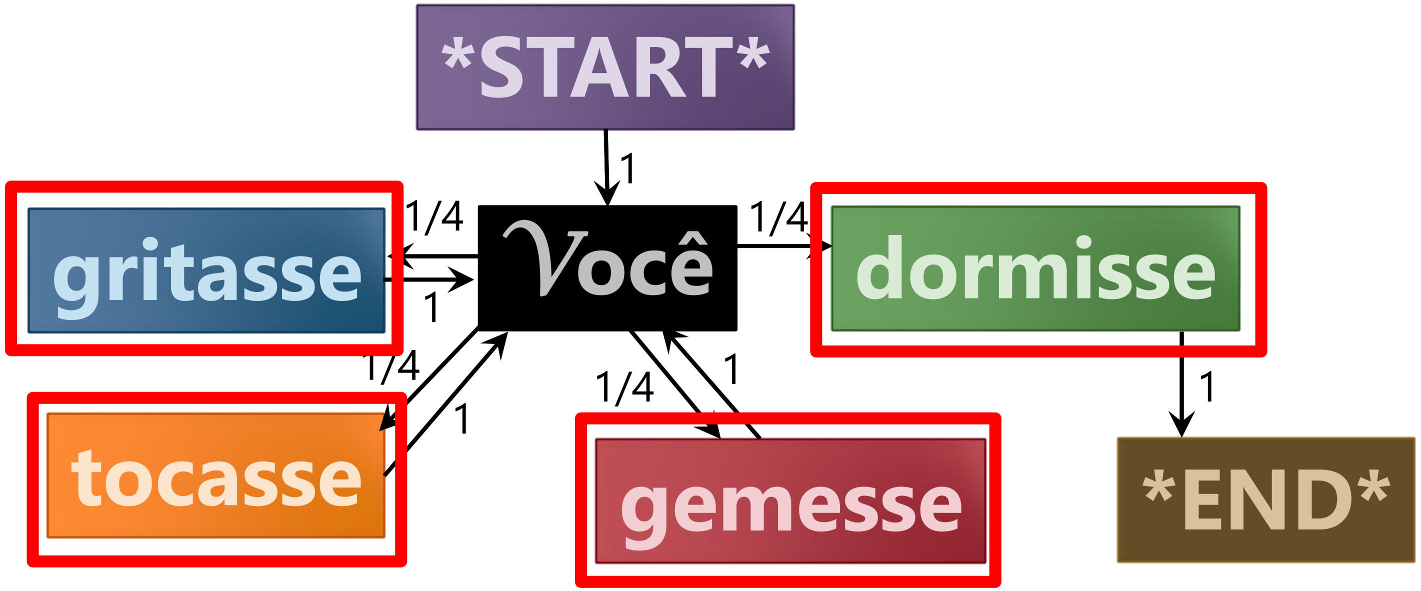
(gritasse, Você)

(gemesse, Você)

(tocasse, Você)

(dormisse, *END*)

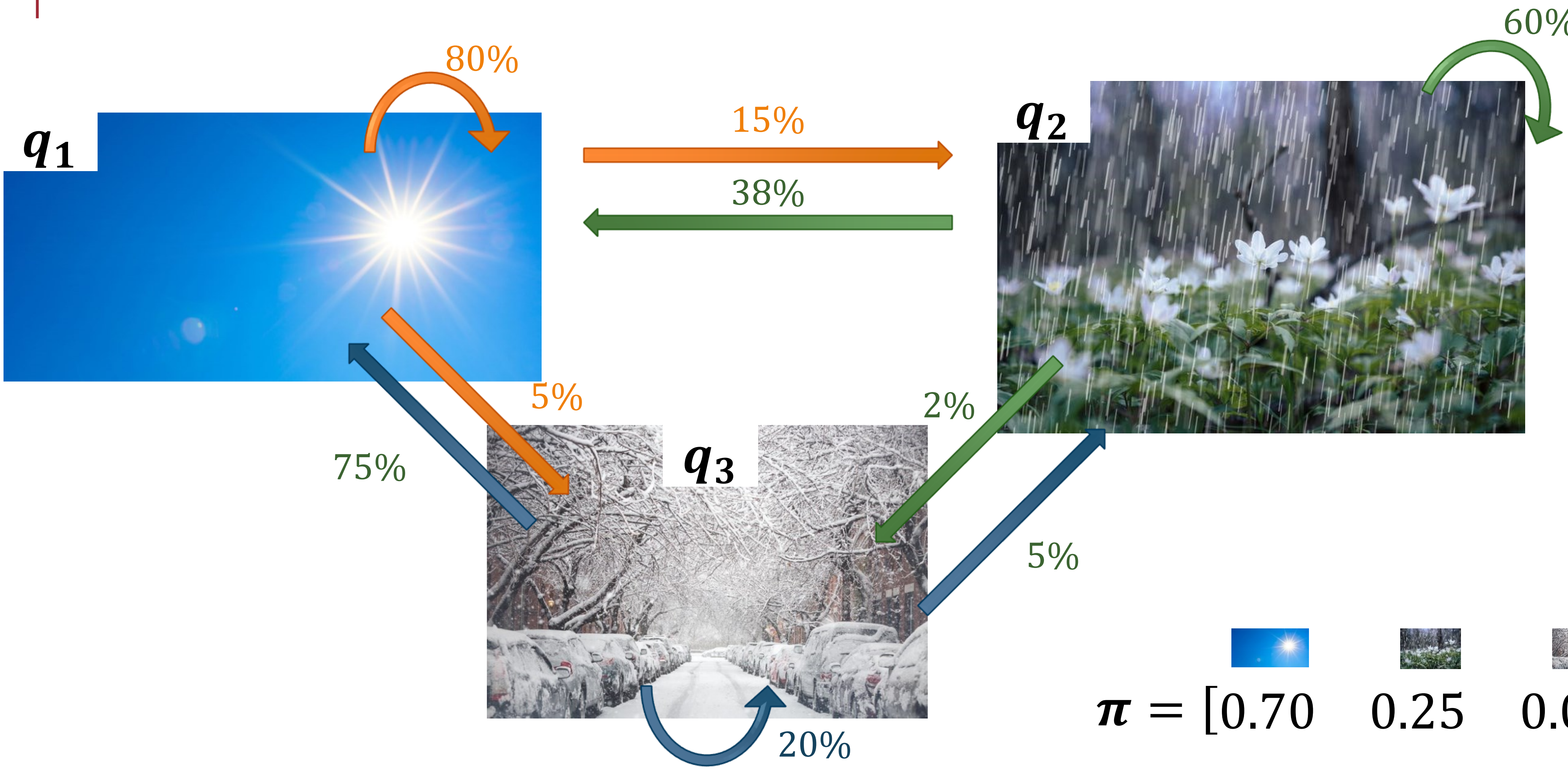
(*END*, none)



START Você gritasse Você gemesse Você tocasse Você dormisse *END*

PECE Programa de Educação Continuada
Escola Politécnica da USP

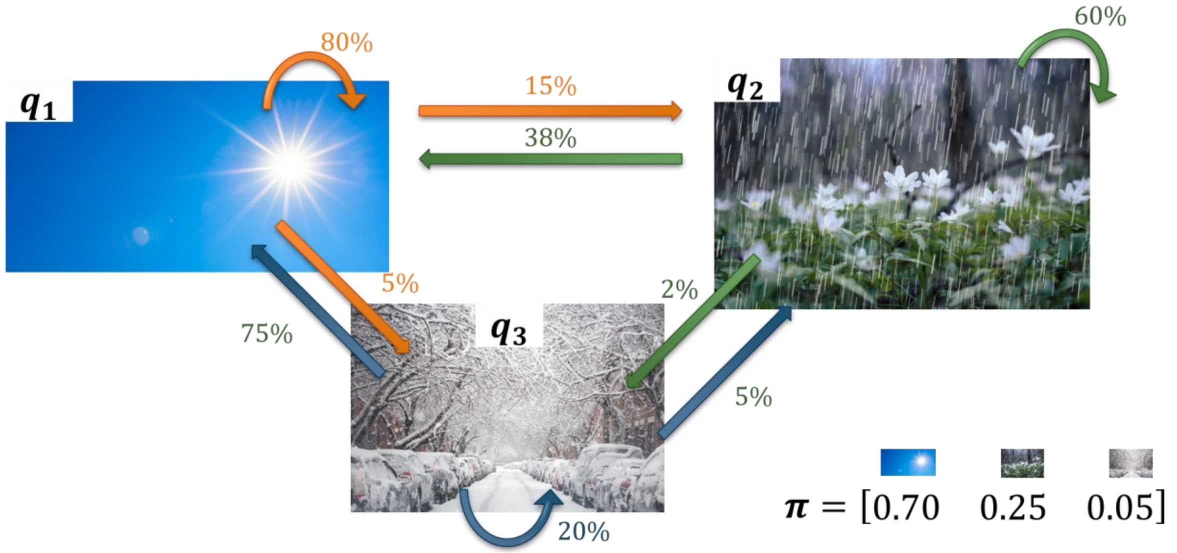
VAMOS EM FRENTE, COM OUTRO EXEMPLO



$$\pi = [0.70 \quad 0.25 \quad 0.05]$$

POR EXEMPLO... QUAL A PROBABILIDADE DA SÉRIE ABAIXO OCORRER?

$$P(s_0) \quad \times \quad P(s_1|s_0) \quad \times \quad P(s_2|s_1) \quad \times \quad P(s_2|s_2) \quad \times \quad P(s_2|s_2) \quad \times \quad P(s_3|s_2)$$



Resposta:

$$P = 0,7 \times 0,15 \times 0,6 \times 0,6 \times 0,02 \times 0,2 = 0,0001512$$

PECE Programa de Educação Continuada
Escola Politécnica da USP

INGREDIENTES DO MODELO DE MARKOV

Um Processo de Markov é uma tupla $\langle Q, A \rangle$ onde Q é um conjunto (finito) de estados e A é a matriz de probabilidade de transição de estado.

Estados possíveis

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

Probabilidade de transição de estado (matriz $n \times n$)

$$p_{ij} = \mathbb{P}\left(s_{t+1} = q_j \middle| s_t = q_i\right)$$

Estado inicial

$$\boldsymbol{\pi} = \mathbb{P}\left[s_{0_i} = q_i\right]$$

DESTINO

$$\boldsymbol{A} = \begin{array}{cc} \begin{array}{cccc} q_1 & q_2 & \dots & q_n \end{array} & \\ \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} & \begin{array}{c} q_1 \\ q_2 \\ \dots \\ q_n \end{array} \end{array}$$

ORIGEM

$a_{ij} =$ probabilidade do estado passar de a_i para o estado a_j

INGREDIENTES DO MODELO DE MARKOV

Estados possíveis $\{q_1, q_2, \dots, q_n\}$

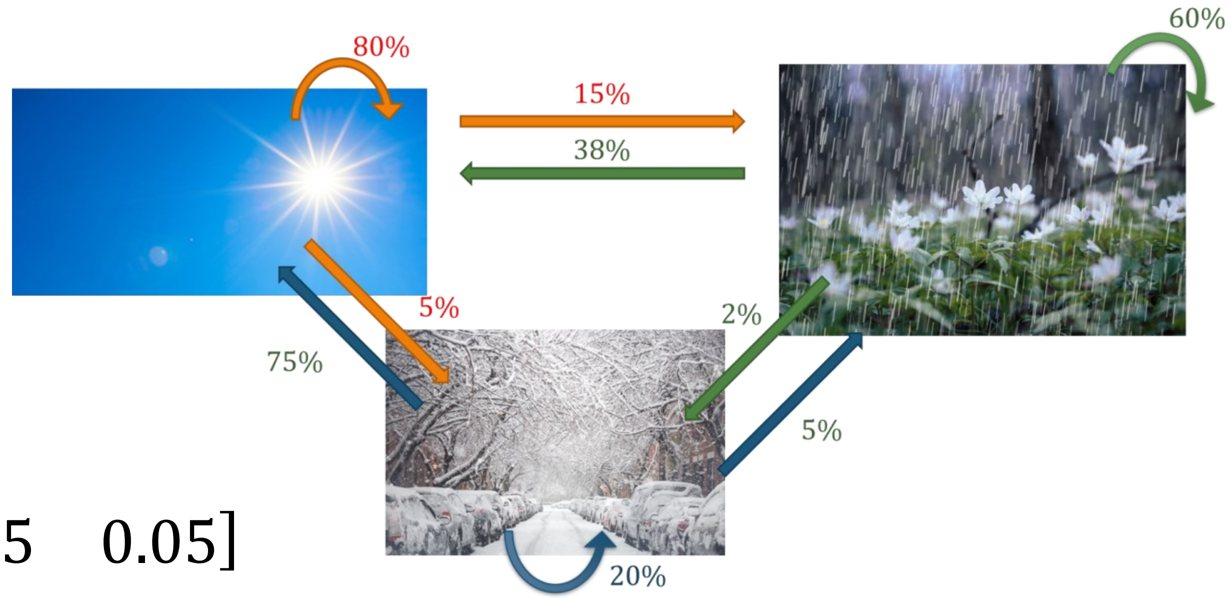
$$\boldsymbol{Q} = \{q_{sol}, q_{chuva}, q_{neve}\}$$

Probabilidade de transição de estado $a_{ij} = \mathbb{P}(s_{t+1} = q_j | s_t = q_i)$

DESTINO

$$\boldsymbol{A} = \begin{array}{cc} \begin{array}{ccc} \text{SOL} & \text{CHUVA} & \text{NEVE} \end{array} & \\ \begin{bmatrix} 0.80 & 0.15 & 0.05 \\ 0.38 & 0.60 & 0.02 \\ 0.75 & 0.05 & 0.20 \end{bmatrix} & \begin{array}{c} \text{SOL} \\ \text{CHUVA} \\ \text{NEVE} \end{array} \end{array}$$

ORIGEM



Estado inicial $\pi = \mathbb{P}[S_{0_i} = s_i]$

$$\boldsymbol{\pi} = [0.7 \quad 0.25 \quad 0.05]$$

IMPORTANTE: MATRIZ ESTOCÁSTICA E REGULAR

$$A = \begin{bmatrix} 0.80 & 0.15 & 0.05 \\ 0.38 & 0.60 & 0.02 \\ 0.75 & 0.05 & 0.20 \end{bmatrix}$$

matriz estocástica:
SOMA = 1

- Uma matriz estocástica é também regular, se $A^n, n > 1$ possui apenas entradas não nulas positivas;
- Se a matriz de transição A para uma cadeia de Markov é regular, então a cadeia de Markov tem um vetor limite único (conhecido como vetor de **estado estacionário**), independentemente dos valores do vetor de probabilidade inicial;
- A matriz estocástica regular tem autovalores $\lambda \leq 1$ e um autovalor $\lambda = 1$.

COMO ACHAR UM ESTADO FUTURO

Estado FUTURO

Estado INICIAL

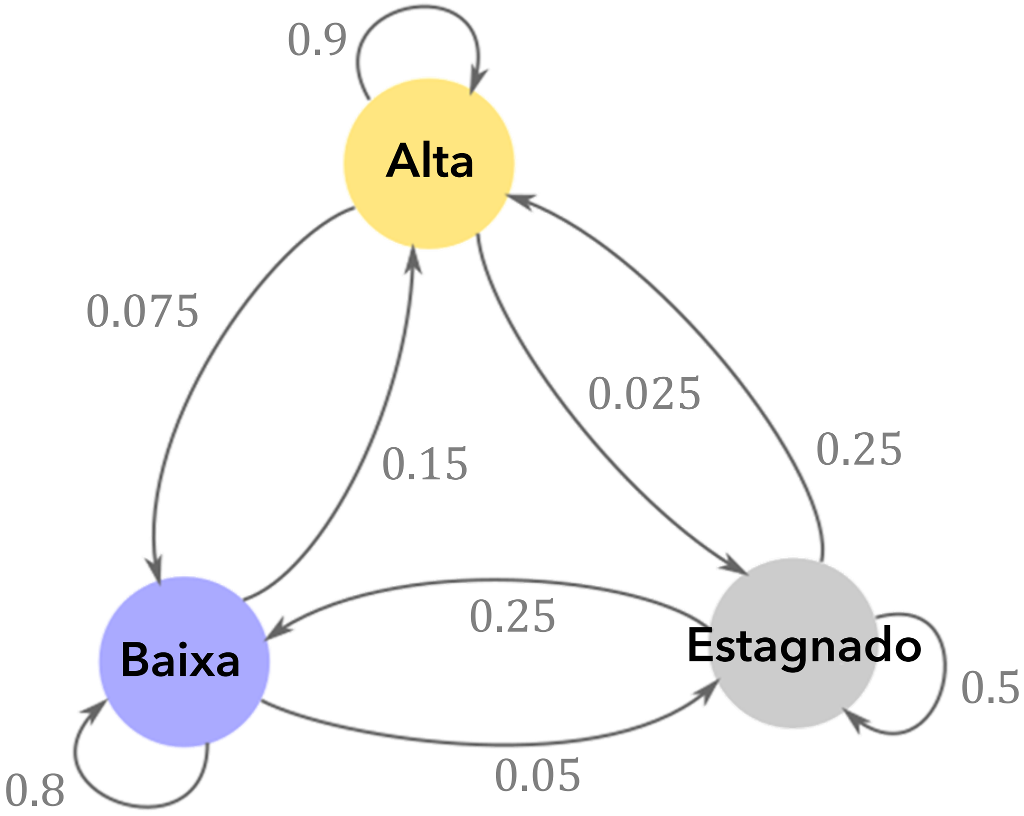
Matriz de Transição

$$[\mathbb{P}(q_{sol}) \quad \mathbb{P}(q_{chuva}) \quad \mathbb{P}(q_{neve})] = [0.7 \quad 0.25 \quad 0.05] \begin{array}{c c} \begin{array}{ccc} & \text{ORIGEM} & \\ \text{SOL} & \text{CHUVA} & \text{NEVE} \end{array} & \\ \begin{bmatrix} 0.80 & 0.15 & 0.05 \\ 0.38 & 0.60 & 0.02 \\ 0.75 & 0.05 & 0.20 \end{bmatrix} & \begin{array}{c} \text{SOL} \\ \text{CHUVA} \\ \text{NEVE} \end{array} \ \text{DESTINO} \end{array} = [0.69 \quad 0.26 \quad 0.05]$$

$$\begin{aligned} \mathbb{P}(s_1 = q_{sol}) = {} & \mathbb{P}(s_1 = q_{sol} | s_0 = q_{sol}) \mathbb{P}(s_0 = q_{sol}) + \\ & \mathbb{P}(s_1 = s_{sol} | s_0 = q_{chuva}) \mathbb{P}(s_0 = q_{chuva}) + \\ & \mathbb{P}(s_1 = s_{sol} | s_0 = q_{neve}) \mathbb{P}(s_0 = q_{neve}) \end{aligned}$$

EXEMPLO

Ache a matriz de transição do problema de Mercado abaixo



$$\boldsymbol{A} = \begin{bmatrix} 0.9 & & \\ & 0.8 & \\ & & 0.5 \end{bmatrix}$$

PECE Programa de Educação Continuada
Escola Politécnica da USP

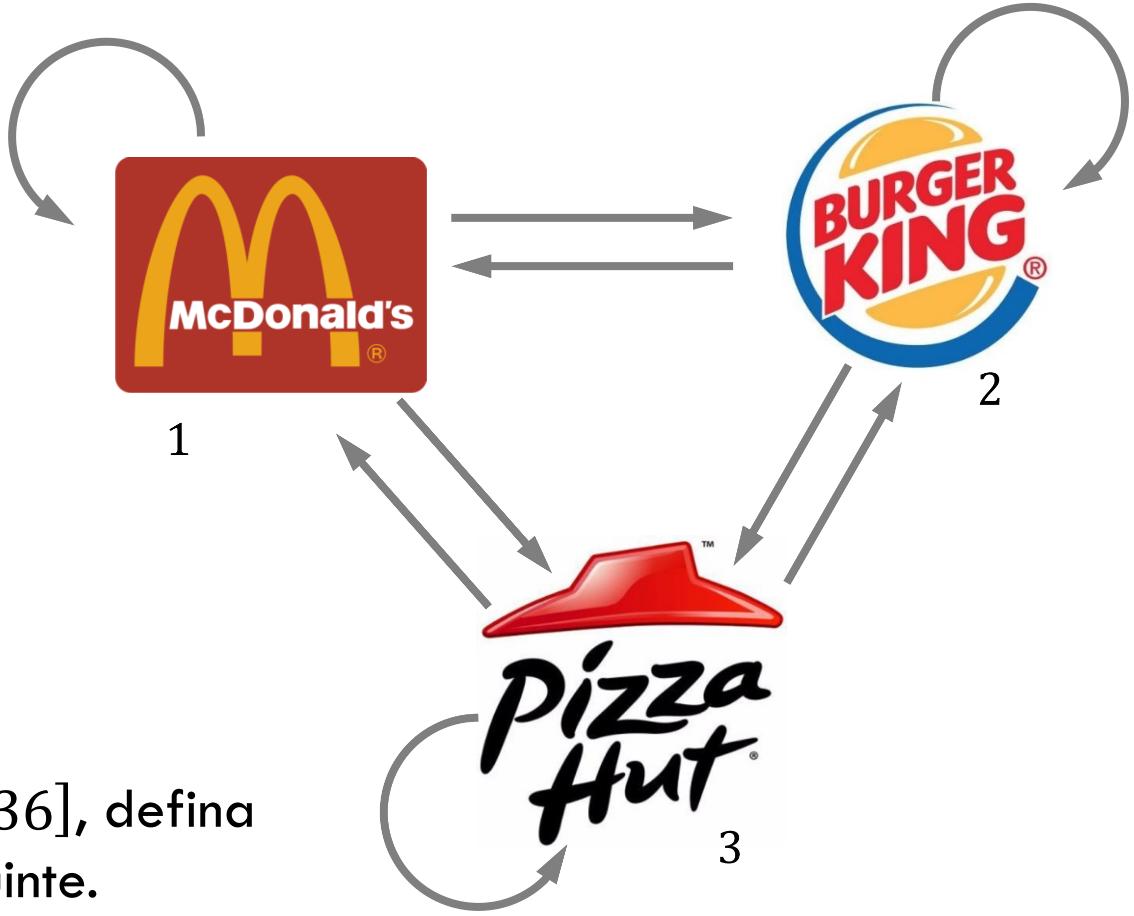
AGORA, O CONTRÁRIO.

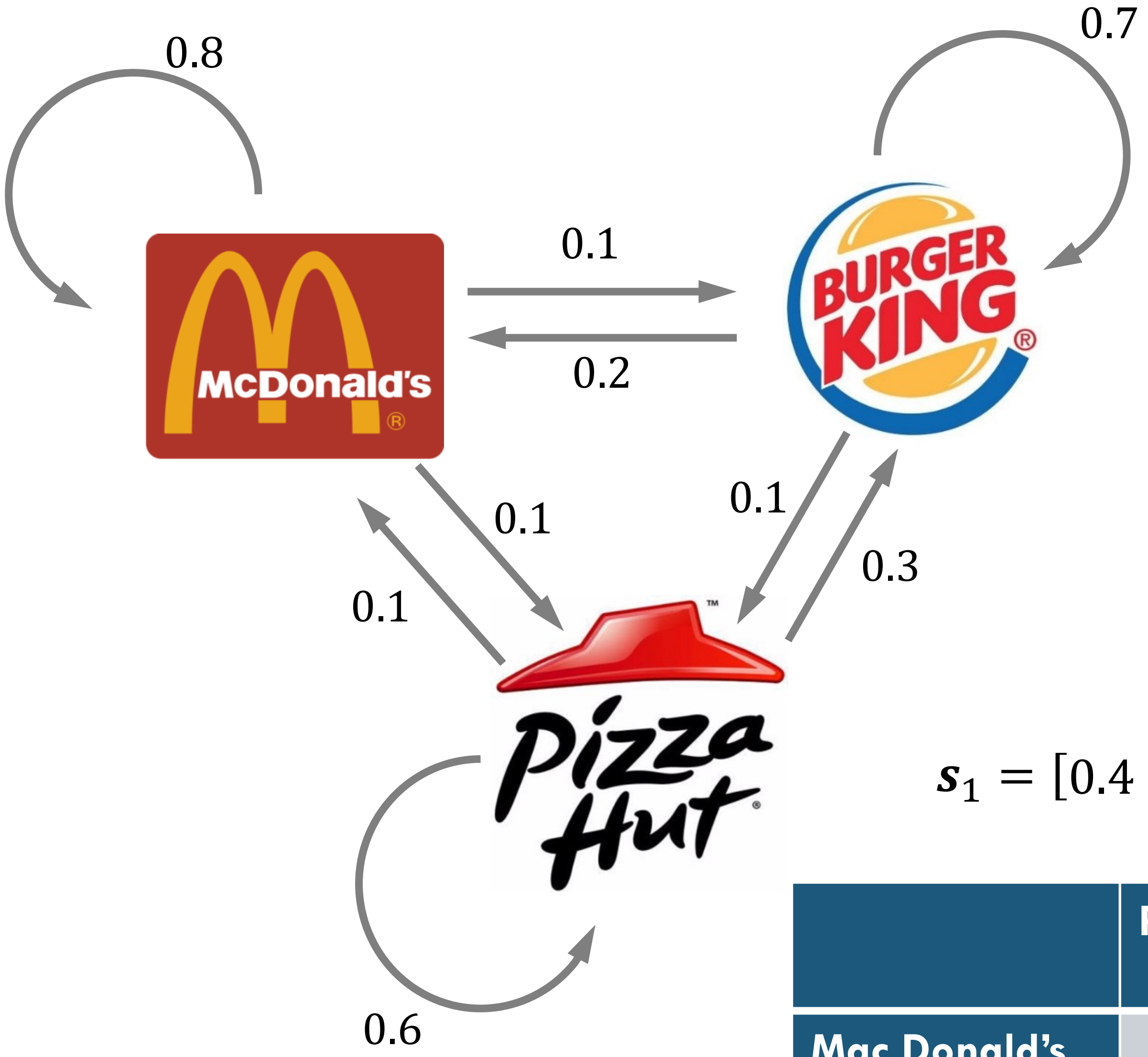
Queremos analisar a transação de clientes em uma área de alimentação de um shopping. Analisamos os clientes almoçando nos três lugares mostrados: Mac Donald's (q_1), Burger King (q_2) e Pizza Hut (q_3), respectivamente. A probabilidade do cliente voltar ou ir para outro lugar é definida conforme a matriz de transição $a_{ij} = \mathbb{P}\left(s_{t+1} = q_j \middle| s_t = q_i\right)$

$$\boldsymbol{A} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.3 & 0.6 \end{bmatrix}$$

OBS: A matriz de transição tem um ciclo de 24 horas.

1. Complete a cadeia de Markov ao lado;
2. Dada a probabilidade inicial $\boldsymbol{\pi} = [0.4 \quad 0.24 \quad 0.36]$, defina a distribuição provável de 500 clientes no dia seguinte.





PECE Programa de Educação Continuada
Escola Politécnica da USP

DESTINO

$$A = \begin{array}{c} \\ \text{Mac} \\ \text{Burger} \\ \text{Pizza} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Mac} & \text{Burger} & \text{Pizza} \end{array} \\ \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.3 & 0.6 \end{bmatrix} \end{array} \text{ORIGEM}$$

$$s_1 = \begin{bmatrix} 0.4 & 0.24 & 0.36 \end{bmatrix} \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.3 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.404 & 0.316 & 0.280 \end{bmatrix}$$

	Probabilidade	# Clientes dia 0	Probabilidade	# Clientes dia 1
Mac Donald's	0.4	200	0.404	202
Burger King	0.24	120	0.316	158
Pizza Hut	0.36	180	0.280	140

E O QUE ACONTECERÁ NO 2^0, 3^0 ... DIA??

$$S_0 \rightarrow S_1 = S_0 A \rightarrow S_2 = S_1 A \rightarrow S_3 = S_2 A$$

$$\boldsymbol{S}_1 = [0.40 \quad 0.240 \quad 0.360]\begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.3 & 0.6 \end{bmatrix} = [0.404 \quad 0.316 \quad 0.280]$$

$$\boldsymbol{S}_2 = [0.404 \quad 0.316 \quad 0.280]\begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.3 & 0.6 \end{bmatrix} = [0.414 \quad 0.346 \quad 0.240]$$

$$\boldsymbol{S}_3 = [0.414 \quad 0.346 \quad 0.240]\begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.3 & 0.6 \end{bmatrix} = [0.424 \quad 0.356 \quad 0.220]$$

S_0	S_1	S_2	S_3
200	202	207	212
120	158	173	179
180	140	120	110

ENTÃO... QUAL SERÁ O ESTADO DEPOIS DE n DIAS?

Multiplicando-se a matriz de transição $\boldsymbol{A}$ elevada à potência n pelo estado inicial $\boldsymbol{s}_0$ (a matriz $\boldsymbol{\pi}$) tem-se a distribuição de probabilidade do estado $\boldsymbol{s}_n$,

$$\boldsymbol{S}_{t+n} = \boldsymbol{S}_t\boldsymbol{A}^n \text{ ou } \boldsymbol{S}_n = \boldsymbol{\pi}\boldsymbol{A}^n$$

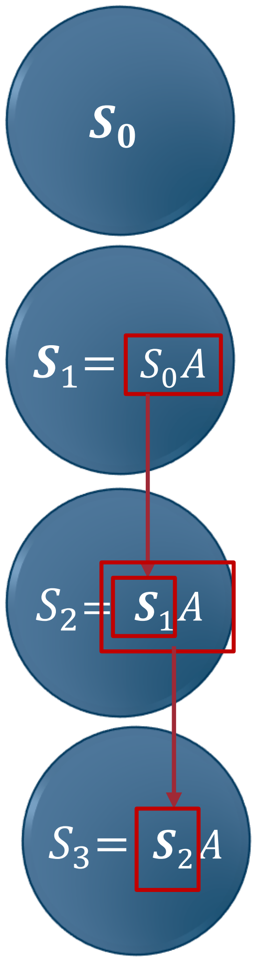
Por exemplo:

$$\boldsymbol{S}_3 = \underbrace{\boxed{\begin{bmatrix}0.414 & 0.346 & 0.240\end{bmatrix}}}_{\boldsymbol{S}_2}\underbrace{\boxed{\begin{bmatrix}0.8 & 0.1 & 0.1\\ 0.2 & 0.7 & 0.1\\ 0.1 & 0.3 & 0.6\end{bmatrix}}}_{\boldsymbol{A}} = \begin{bmatrix}0.424 & 0.356 & 0.220\end{bmatrix}$$

ou

$$\boldsymbol{S}_3 = \underbrace{\boxed{\begin{bmatrix}0.40 & 0.240 & 0.360\end{bmatrix}}}_{\boldsymbol{\pi}}\underbrace{\boxed{\begin{bmatrix}0.8 & 0.1 & 0.1\\ 0.2 & 0.7 & 0.1\\ 0.1 & 0.3 & 0.6\end{bmatrix}^3}}_{\boldsymbol{A}^3} = \begin{bmatrix}0.424 & 0.356 & 0.220\end{bmatrix}$$

$$\begin{bmatrix}0.8 & 0.1 & 0.1\\ 0.2 & 0.7 & 0.1\\ 0.1 & 0.3 & 0.6\end{bmatrix}^3 = \begin{bmatrix}0.8 & 0.1 & 0.1\\ 0.2 & 0.7 & 0.1\\ 0.1 & 0.3 & 0.6\end{bmatrix}\begin{bmatrix}0.8 & 0.1 & 0.1\\ 0.2 & 0.7 & 0.1\\ 0.1 & 0.3 & 0.6\end{bmatrix}\begin{bmatrix}0.8 & 0.1 & 0.1\\ 0.2 & 0.7 & 0.1\\ 0.1 & 0.3 & 0.6\end{bmatrix}$$



$$S_1 = [0.40 \quad 0.240 \quad 0.360]\begin{bmatrix}0.8 & 0.1 & 0.1\\ 0.2 & 0.7 & 0.1\\ 0.1 & 0.3 & 0.6\end{bmatrix} = [0.404 \quad 0.316 \quad 0.280]$$

$$S_2 = [0.404 \quad 0.316 \quad 0.280]\begin{bmatrix}0.8 & 0.1 & 0.1\\ 0.2 & 0.7 & 0.1\\ 0.1 & 0.3 & 0.6\end{bmatrix} = [0.414 \quad 0.346 \quad 0.240]$$

$$S_3 = [0.414 \quad 0.346 \quad 0.240]\begin{bmatrix}0.8 & 0.1 & 0.1\\ 0.2 & 0.7 & 0.1\\ 0.1 & 0.3 & 0.6\end{bmatrix} = [0.424 \quad 0.356 \quad 0.220]$$

$$S_4 = [0.424 \quad 0.356 \quad 0.220]\begin{bmatrix}0.8 & 0.1 & 0.1\\ 0.2 & 0.7 & 0.1\\ 0.1 & 0.3 & 0.6\end{bmatrix} = [0.432 \quad 0.358 \quad 0.210]$$

$$S_5 = [0.432 \quad 0.358 \quad 0.210]\begin{bmatrix}0.8 & 0.1 & 0.1\\ 0.2 & 0.7 & 0.1\\ 0.1 & 0.3 & 0.6\end{bmatrix} = [0.438 \quad 0.357 \quad 0.205]$$

$$S_6 = [0.438 \quad 0.357 \quad 0.205]\begin{bmatrix}0.8 & 0.1 & 0.1\\ 0.2 & 0.7 & 0.1\\ 0.1 & 0.3 & 0.6\end{bmatrix} = [0.442 \quad 0.355 \quad 0.203]$$

$$S_7 = [0.442 \quad 0.355 \quad 0.203]\begin{bmatrix}0.8 & 0.1 & 0.1\\ 0.2 & 0.7 & 0.1\\ 0.1 & 0.3 & 0.6\end{bmatrix} = [0.445 \quad 0.354 \quad 0.202] = S_8$$

Distribuição estável $\overline{S}$

200 (40%)

120 (24%)

180 (36%)

222 (45%)

177 (20%)

101(35%)

PORÉM...

Dessa forma, $\bar{s} = \boldsymbol{\pi}\,\bar{A}$.

$$\boldsymbol{S}_7 = \bar{\boldsymbol{S}} = [0.40 \quad 0.240 \quad 0.360]\boxed{\begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.3 & 0.6 \end{bmatrix}^7} = \boldsymbol{\pi}\,\boxed{\bar{A}}$$

Existem alguma maneira eficiente de achar $\bar{A}$ que não seja por tentativa????

Veja, que, se $\boldsymbol{A}_{t+n} = \boldsymbol{S}_t \boldsymbol{A}^n$ e o modelo está estabilizado, então, para $n = 1$

$$\bar{\boldsymbol{S}} = \bar{\boldsymbol{S}}\bar{\boldsymbol{A}}$$

PECE Programa de Educação Continuada
Escola Politécnica da USP

$A^T \bar{S}^T = \bar{S}^T$

$$Au = \lambda u, \quad u \neq 0$$
$$(A - \lambda I)u = 0$$
$$|(\lambda I - A)| = 0$$

Se está em dúvidas, reveja a aula de autovalores e autovetores...

Podemos ver que $\bar{S}^T$ deve ser um autovetor de $\bar{A}^T$ com um autovalor 1.

Por exemplo, para uma matriz $\bar{A}$ 3×3, tem-se:

$$\begin{vmatrix} a_{11}-1 & a_{21} & a_{31} \\ a_{12} & a_{22}-1 & a_{32} \\ a_{13} & a_{23} & a_{33}-1 \end{vmatrix} = 0$$

$A^T \overline{S}^T = \overline{S}^T$

Ou, ainda, para uma matriz $\overline{A}$ 3×3,

$$\bar{s}_1 a_{11} + \bar{s}_2 a_{21} + \bar{s}_3 a_{31} = \bar{s}_1$$
$$\bar{s}_1 a_{12} + \bar{s}_2 a_{22} + \bar{s}_3 a_{32} = \bar{s}_2$$
$$\bar{s}_1 a_{13} + \bar{s}_2 a_{23} + \bar{s}_3 a_{33} = \bar{s}_3$$

$$\bar{s}_1 + \bar{s}_2 + \bar{s}_3 = 1$$

$\boldsymbol{A}^T\overline{\boldsymbol{S}}^T = \overline{\boldsymbol{S}}^T$

Ou, ainda, para uma matriz $\overline{\boldsymbol{A}}$ 3×3,

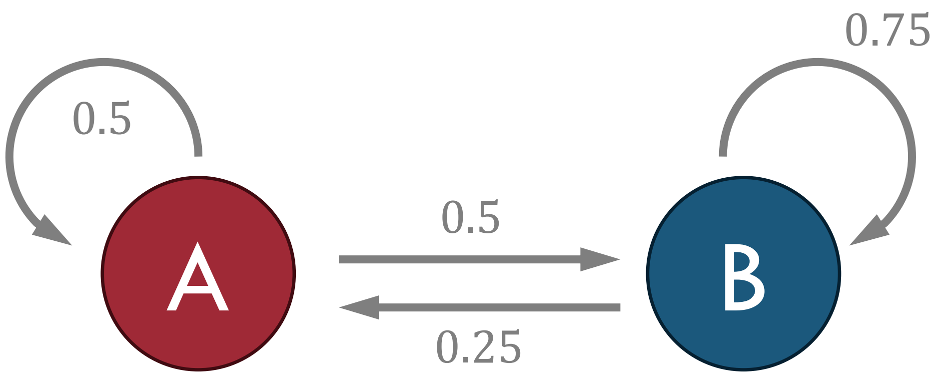
$$
\begin{aligned}
\bar{s}_1\left(a_{11}-1\right) &+ \bar{s}_2 a_{21} &+ \bar{s}_3 a_{31} &= 0 \\
\bar{s}_1 a_{12} &+ \bar{s}_2\left(a_{22}-1\right) &+ \bar{s}_3 a_{32} &= 0 \\
\bar{s}_1 a_{13} &+ \bar{s}_2 a_{23} &+ \bar{s}_3\left(a_{33}-1\right) &= 0 \\
\bar{s}_1 &+ \bar{s}_2 &+ \bar{s}_3 &= 1
\end{aligned}
$$

EXEMPLO

Supondo $\boldsymbol{\pi} = [3/5 \quad 2/5]$ e a matriz $\boldsymbol{A}$ de transição

$$\boldsymbol{A} = \begin{bmatrix} 0.5 & 0.5 \\ 0.25 & 0.75 \end{bmatrix},$$

encontre $\overline{\boldsymbol{S}}$ e $\overline{\boldsymbol{A}}$.



```python
A = np.array([[0.5,0.5],[0.25,0.75]])
w,v=np.linalg.eig(A.T) # autovetores são as colunas de v
for i,autovalor in enumerate(w):
  if autovalor == 1:
    autovetor = v[:,i]
    sum = np.sum(autovetor)
    autovetor = autovetor/sum
sbar = autovetor
print('A distribuição estável é dada por:\n',sbar)
```

```
A distribuição estável é dada por: [0.33333333 0.66666667]
```

```python
A = np.array([[0.5,0.5],[0.25,0.75]])
A2 = np.append(A.T-np.identity(2),[[1,1]],axis=0)
b = (np.array([0,0,1])).T
sbar = np.linalg.solve(A2.T.dot(A2), A2.T.dot(b))
print('A distribuição estável é dada por:\n',sbar)
```

```
A distribuição estável é dada por: [0.33333333 0.66666667]
```

$$\bar{S} = \pi\bar{A}$$

Dedução no Nootebook

$$\bar{\mathbf{A}} = \begin{bmatrix} \bar{\mathbf{S}} \\ \bar{\mathbf{S}} \\ \vdots \end{bmatrix}$$

$$\bar{\mathbf{A}} = \begin{bmatrix} \bar{s}_1 & \bar{s}_2 \\ \bar{s}_1 & \bar{s}_2 \end{bmatrix} = \begin{bmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{bmatrix}$$

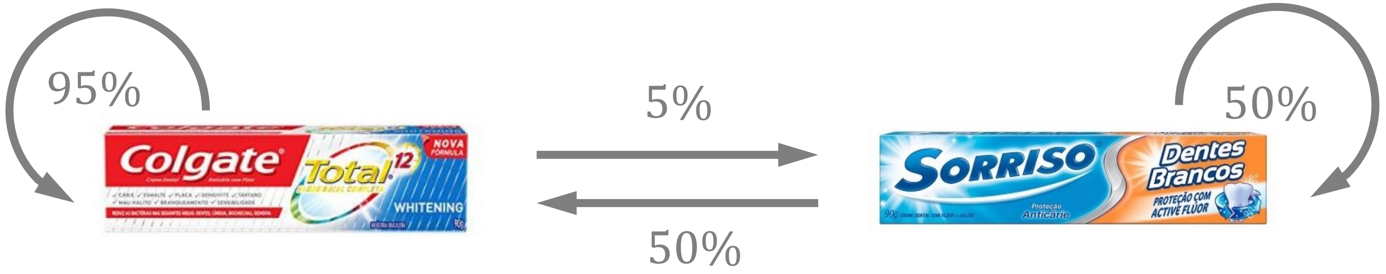
```
S0=np.array([3/5,2/5])
S0.dot(Abar)
```

```
array([0.33333333, 0.66666667])
```

PECE Programa de Educação Continuada
Escola Politécnica da USP

FAÇA SOZINHO

Encontre a matriz de transição estabilizada $\bar{A}$ e o vetor $\overline{\boldsymbol{S}}$.

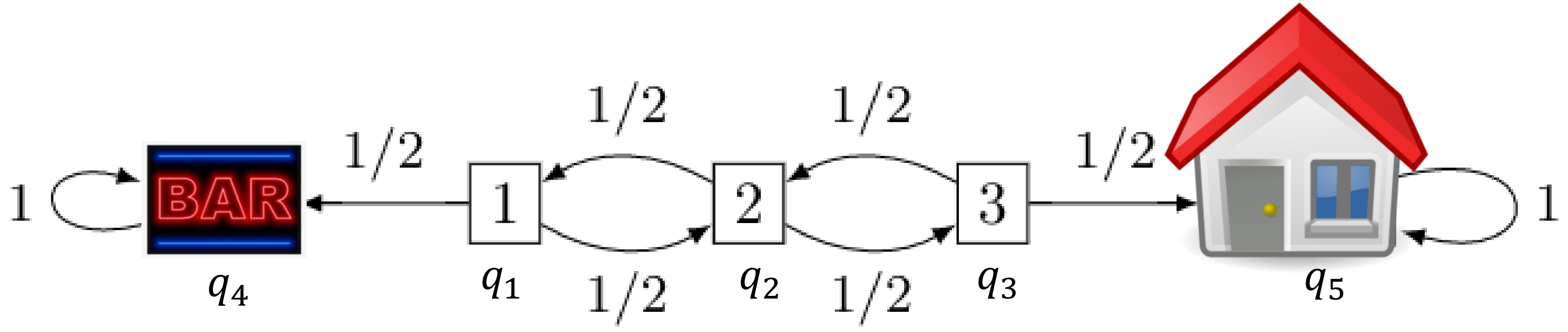


PECE Programa de Educação Continuada
Escola Politécnica da USP

CADEIAS DE MARKOV COM ESTADOS ABSORVENTES

Um estado i da cadeia de Markov é chamado de *estado absorvente* se $a_{ii} = 1$ e, por consequência, qualquer valor da linha $a_{ij} = 0$ para $i \neq j$. Uma cadeia de Markov é dita absorvente se existe ao menos um estado absorvente, ou se for possível, a partir de qualquer estado, atingir um estado absorvente, não necessariamente em um único passo.

$$
\boldsymbol{A} = \begin{bmatrix} 0 & 0.5 & 0 & 0.5 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}
$$



CARACTERÍSTICA DE ESTADO ABSORVENTE

$$\mathbf{A} = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.2 & 0.7 & 0.1 \\ 0.3 & 0.1 & 0.6 \end{bmatrix}$$

$$\mathbf{A}^2 = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.37 & 0.50 & 0.13 \\ 0.50 & 0.13 & 0.37 \end{bmatrix}$$

$$\mathbf{A}^3 = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.509 & 0.363 & 0.128 \\ 0.637 & 0.128 & 0.235 \end{bmatrix}$$

Eventualmente:

$$\bar{\mathbf{A}} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$\bar{s} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

PALAVRA *EVENTUALMENTE*

Veja como exemplo um estudo realizado pela Universidade da Carolina do Norte em pacientes de um determinado hospital. O problema foi modelado por uma cadeia de Markov: 0 (morto), 1 (estado desfavorável), 2(estado favorável). A matriz de transição tem um ciclo de 72 horas,

$$\mathbf{A} = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.085 & 0.779 & 0.136 \\ 0.017 & 0.017 & 0.966 \end{bmatrix}$$

Perceba que $a_{00} = 1$, ou seja, 0 é um estado de absorção, uma vez que o paciente morto, a cada passo ele continuará morto. Os estados 1 e 2 são os estados de transição, e a partir de qualquer um destes é possível chegar no estado de absorção. Daí, a cadeia é absorvente. Porém, não é sensato acreditar que todos os pacientes acabarão mortos. Então, a matriz de transformação estabilizada não se torna .

EXERCÍCIO

Para a cadeia de Markov da rotina do Wall-E ao lado, responda:

1. Qual a matriz de $A_{ss'}$ de probabilidade de transição de estado?

$$a_{ss'} = \mathbb{P}[s_{t+1} = q' | s_t = q]$$

1. Qual a probabilidade do Wall-E ir comer, dado que ele está na academia?

2. Dado $\boldsymbol{\pi} = [0.1 \quad 0.3 \quad 0.3 \quad 0.3 \quad 0 \quad 0]$, qual a probabilidade dos seguintes episódios ocorrerem:

Episódio 1: (Aula, Academia, Comer, Dormir)

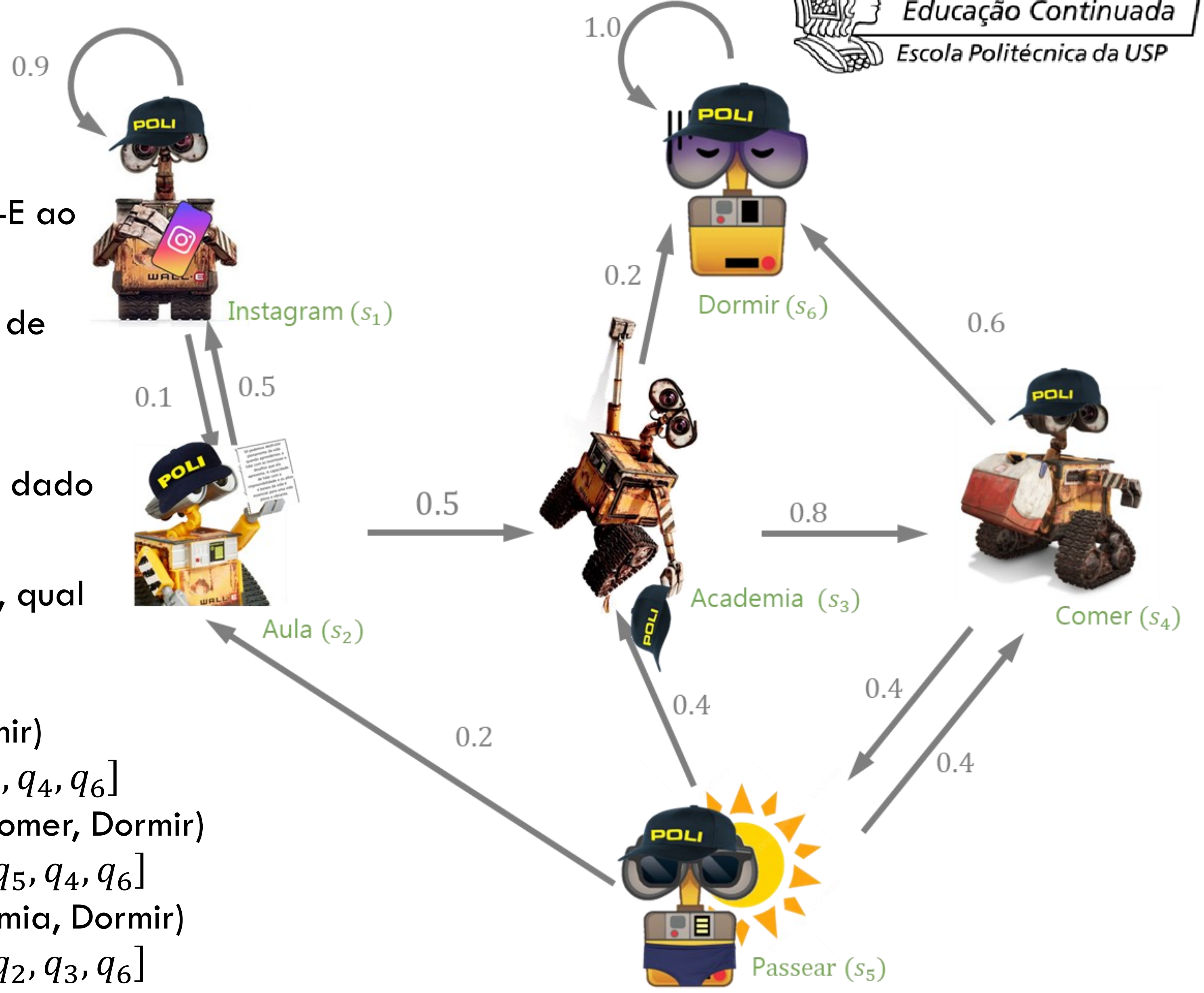
$$\boldsymbol{S}_1 = [q_2, q_3, q_4, q_6]$$

Episódio 2: (Academia, Comer, Passear, Comer, Dormir)

$$\boldsymbol{S}_2 = [q_3, q_4, q_5, q_4, q_6]$$

Episódio 3: (Aula, Instagram, Aula, Academia, Dormir)

$$\boldsymbol{S}_3 = [q_2, q_1, q_2, q_3, q_6]$$



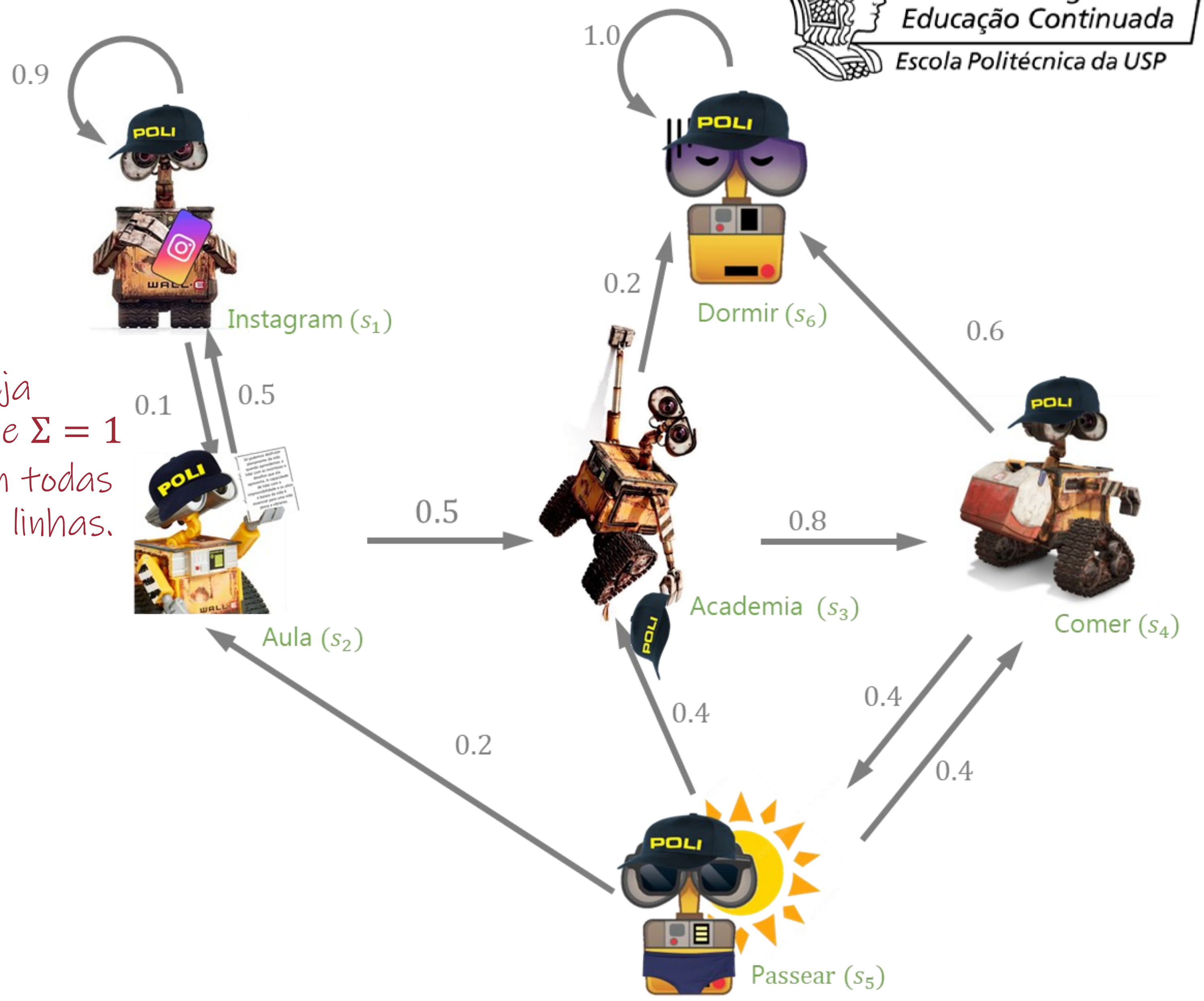
PECE Programa de Educação Continuada
Escola Politécnica da USP

$$
\begin{array}{c|c}
 & \text{DESTINO} \\
 & \begin{array}{cccc} \textit{Instagram} & \textit{Sala} & \ldots & \textit{Dormir} \end{array} \\
\boldsymbol{A} = & \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{16} \\ a_{21} & a_{22} & \cdots & a_{26} \\ \vdots & \vdots & \ddots & \vdots \\ a_{61} & a_{62} & \cdots & a_{66} \end{bmatrix} \begin{array}{c} q_1 \\ q_2 \\ \vdots \\ q_6 \end{array} \ \text{ORIGEM}
\end{array}
$$

$$
\boldsymbol{A} = \begin{bmatrix} 0.9 & 0.1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & \boxed{0.8} & 0 & 0.2 \\ 0 & 0 & 0 & 0 & 0.4 & 0.6 \\ 0 & 0.2 & 0.4 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}
$$

Veja que $\Sigma = 1$ em todas as linhas.

$$
\boldsymbol{\pi} = [0.1 \quad 0.3 \quad 0.3 \quad 0.3 \quad 0 \quad 0]
$$



Qual a probabilidade do Wall-E ir comer, dado que ele está na Academia?

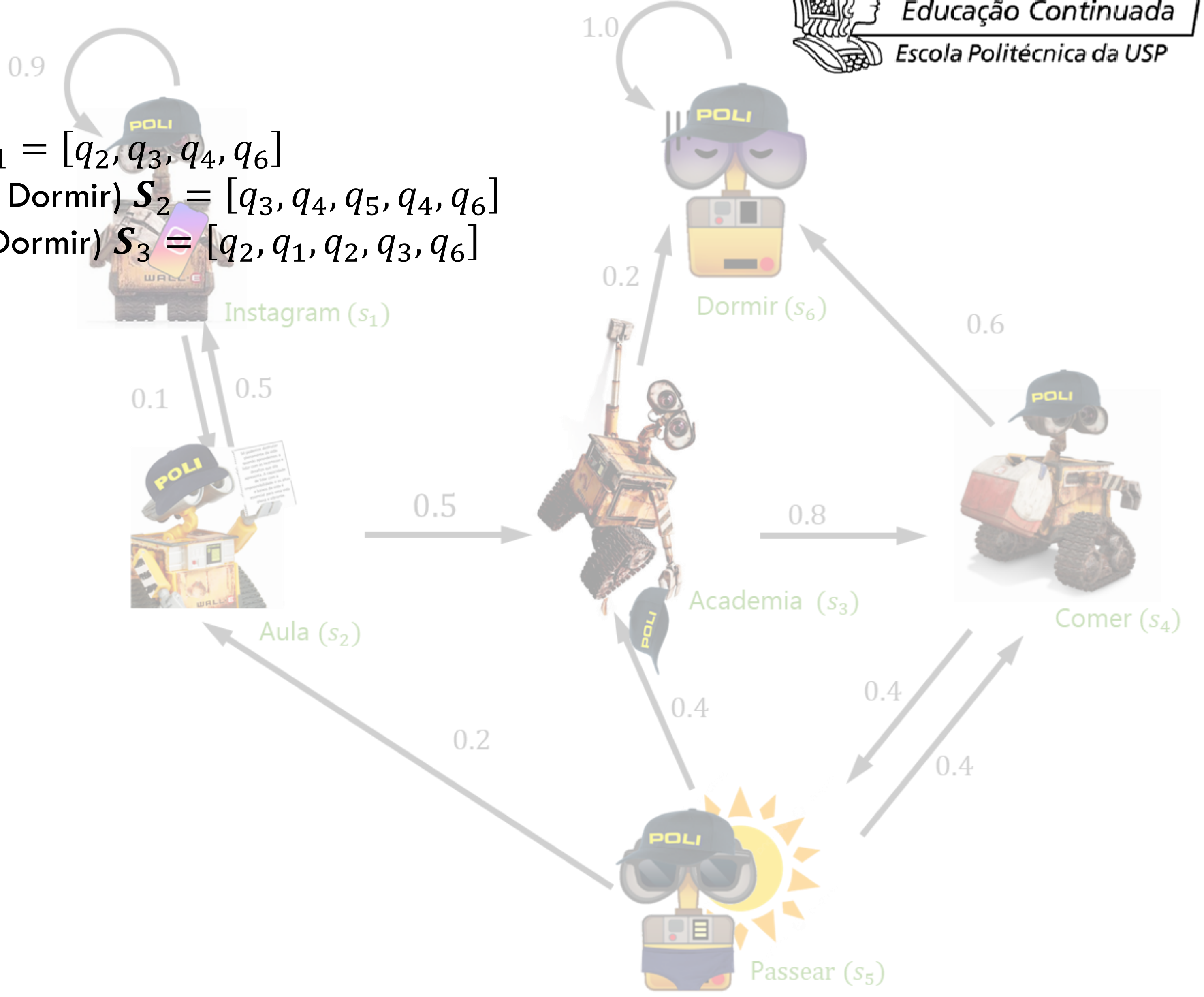
$$
P_{34} = P(s_{t+1} = q_4 | s_t = q_3) = 0{,}8
$$

Episódios:

Episódio 1: (Aula, Academia, Comer, Dormir) $\boldsymbol{S_1} = [q_2, q_3, q_4, q_6]$
Episódio 2: (Academia, Comer, Passear, Comer, Dormir) $\boldsymbol{S}_2 = [q_3, q_4, q_5, q_4, q_6]$
Episódio 3: (Aula, Instagram, Aula, Academia, Dormir) $\boldsymbol{S}_3 = [q_2, q_1, q_2, q_3, q_6]$

$$\boldsymbol{A} = \begin{bmatrix} 0.9 & 0.1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.8 & 0 & 0.2 \\ 0 & 0 & 0 & 0 & 0.4 & 0.6 \\ 0 & 0.2 & 0.4 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\pi = [0.1 \quad 0.3 \quad 0.3 \quad 0.3 \quad 0 \quad 0]$$

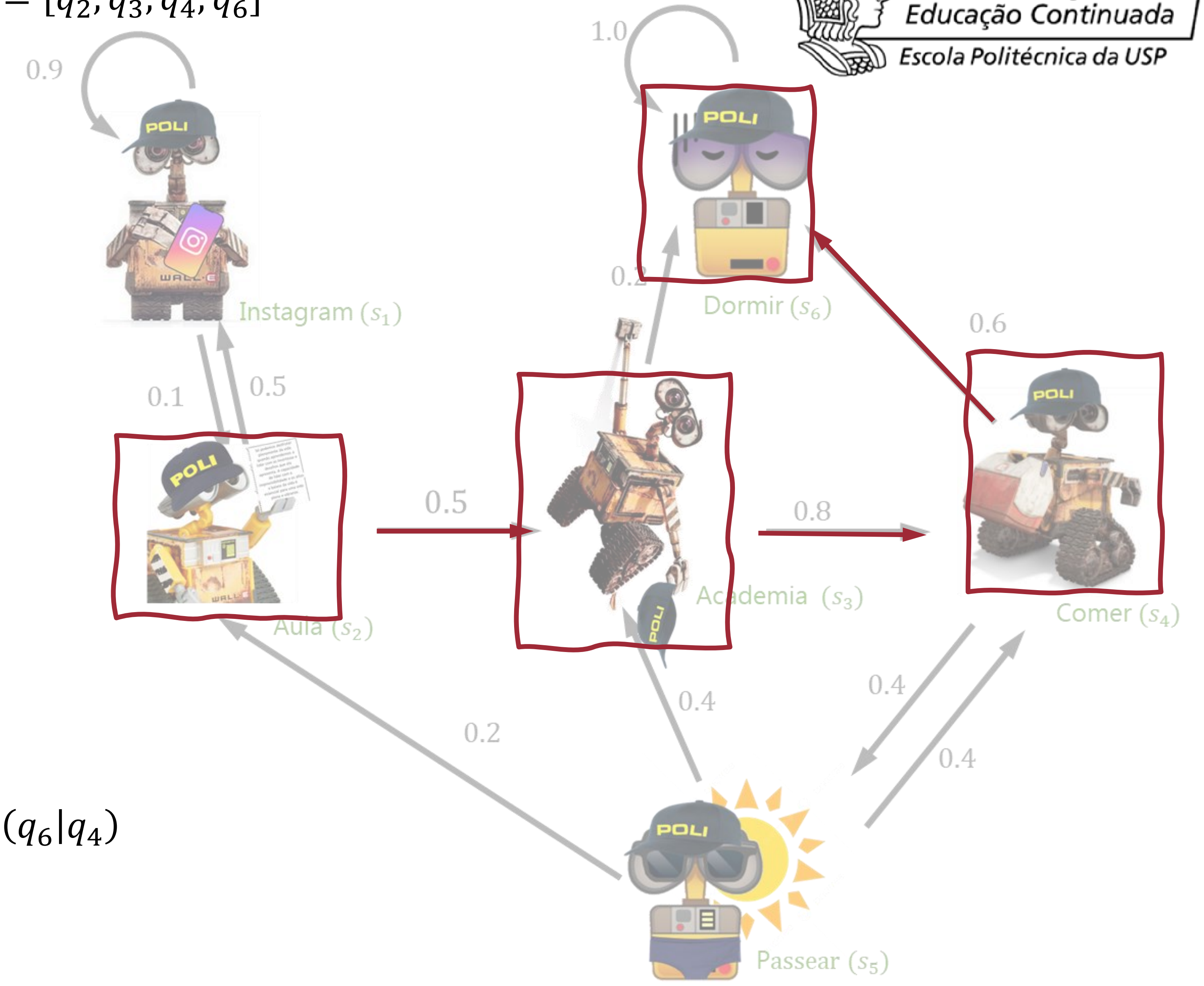


PECE Programa de Educação Continuada
Escola Politécnica da USP

Episódio 1: (Aula, Academia, Comer, Dormir) $\boldsymbol{S_1} = [q_2, q_3, q_4, q_6]$

$$\boldsymbol{A} = \begin{bmatrix} 0.9 & 0.1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.8 & 0 & 0.2 \\ 0 & 0 & 0 & 0 & 0.4 & 0.6 \\ 0 & 0.2 & 0.4 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\pi = [0.1 \quad 0.3 \quad 0.3 \quad 0.3 \quad 0 \quad 0]$$



$$P(\boldsymbol{S}_1) = P(q_2) \times P(q_3|q_2) \times P(q_4|q_3) \times P(q_6|q_4)$$
$$P(\boldsymbol{S}_1) = \pi_2 \times P_{23} \times P_{34} \times P_{46}$$
$$P(\boldsymbol{S}_1) = 0.3 \times 0.5 \times 0.8 \times 0.6 = 0.072$$

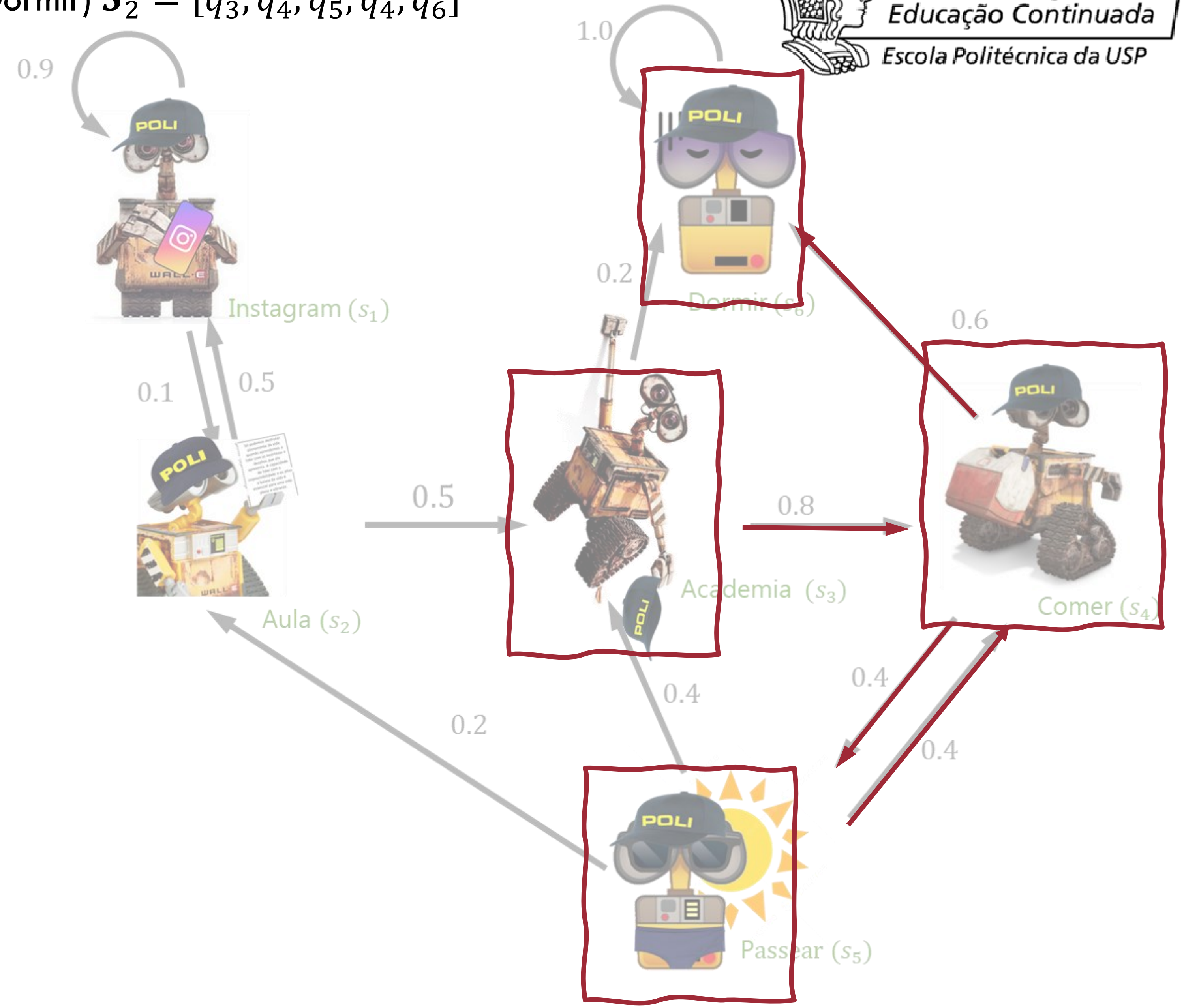
PECE Programa de Educação Continuada
Escola Politécnica da USP

Episódio 2: (Academia, Comer, Passear, Comer, Dormir) $\boldsymbol{S}_2 = [q_3, q_4, q_5, q_4, q_6]$

$$\boldsymbol{A} = \begin{bmatrix} 0.9 & 0.1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.8 & 0 & 0.2 \\ 0 & 0 & 0 & 0 & 0.4 & 0.6 \\ 0 & 0.2 & 0.4 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\pi = [0.1 \quad 0.3 \quad 0.3 \quad 0.3 \quad 0 \quad 0]$$

$$P(\boldsymbol{S}_2) = 0.023$$

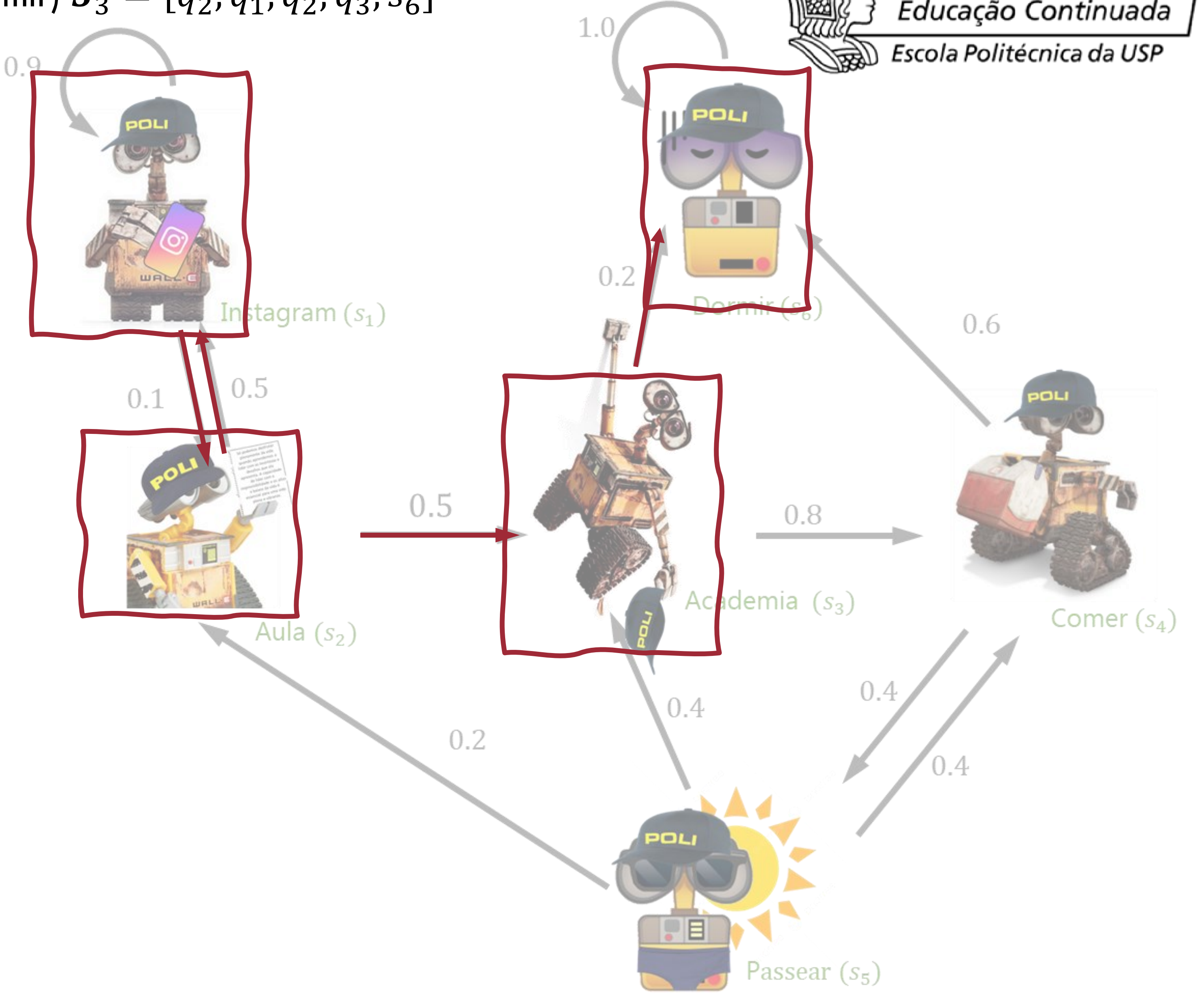


Episódio 3: (Aula, Instagram, Aula, Academia, Dormir) $\boldsymbol{S}_3 = [q_2, q_1, q_2, q_3, s_6]$

$$\boldsymbol{A} = \begin{bmatrix} 0.9 & 0.1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.8 & 0 & 0.2 \\ 0 & 0 & 0 & 0 & 0.4 & 0.6 \\ 0 & 0.2 & 0.4 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\pi = [0.1 \quad 0.3 \quad 0.3 \quad 0.3 \quad 0 \quad 0]$$

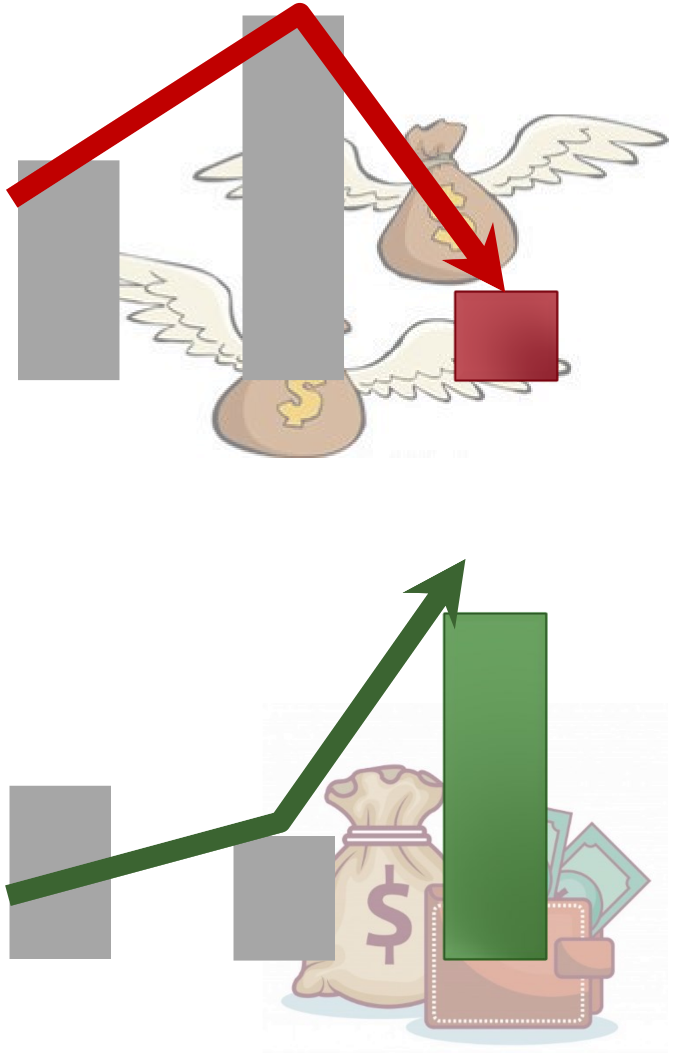
$$P(\boldsymbol{S}_3) = 0.0015$$



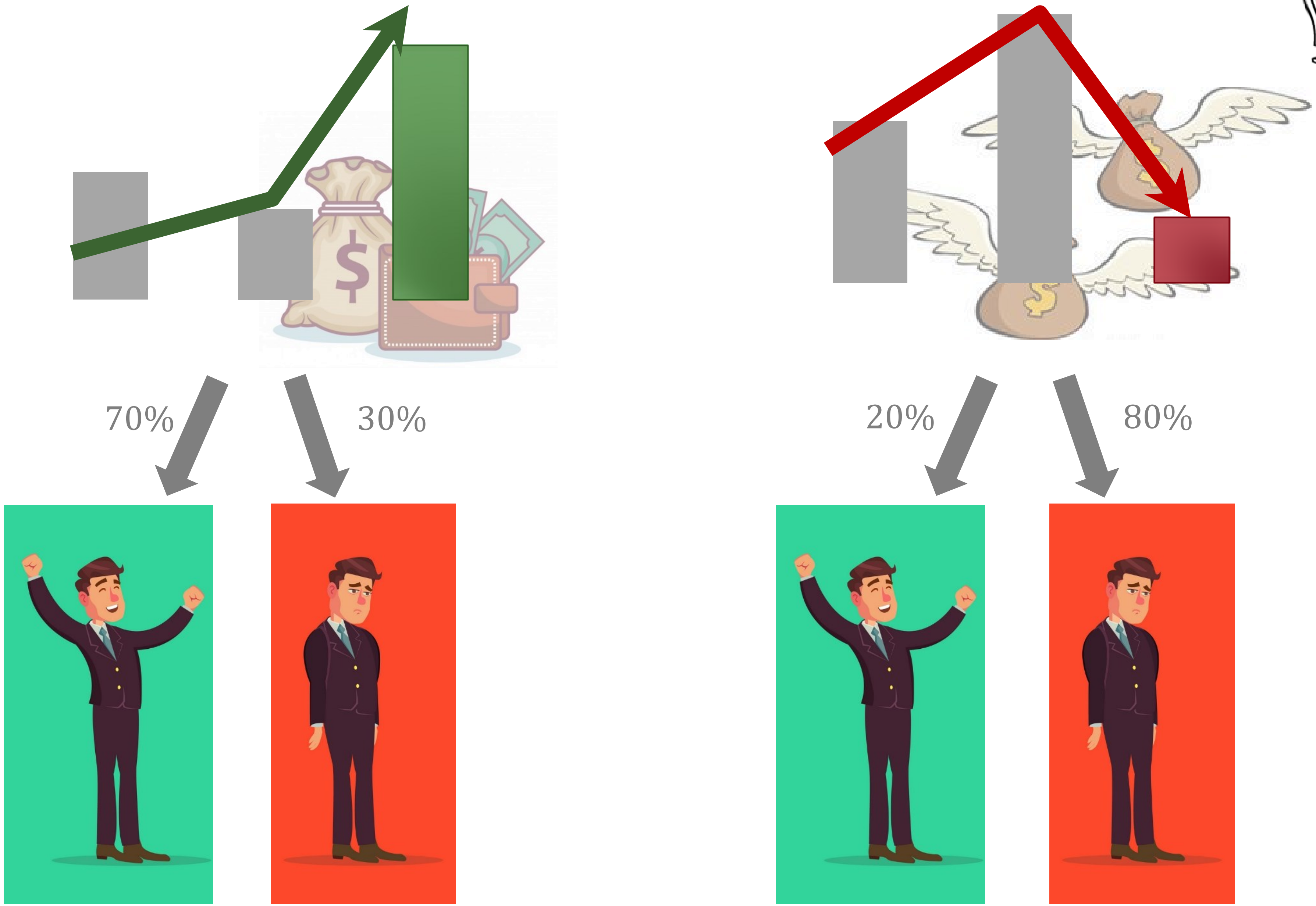
MODELO DE MARKOV OCULTO (HMM)

Não podemos esperar que seja sempre possível observar completamente o real estado do sistema a cada instante.

ANA E MÁRIO

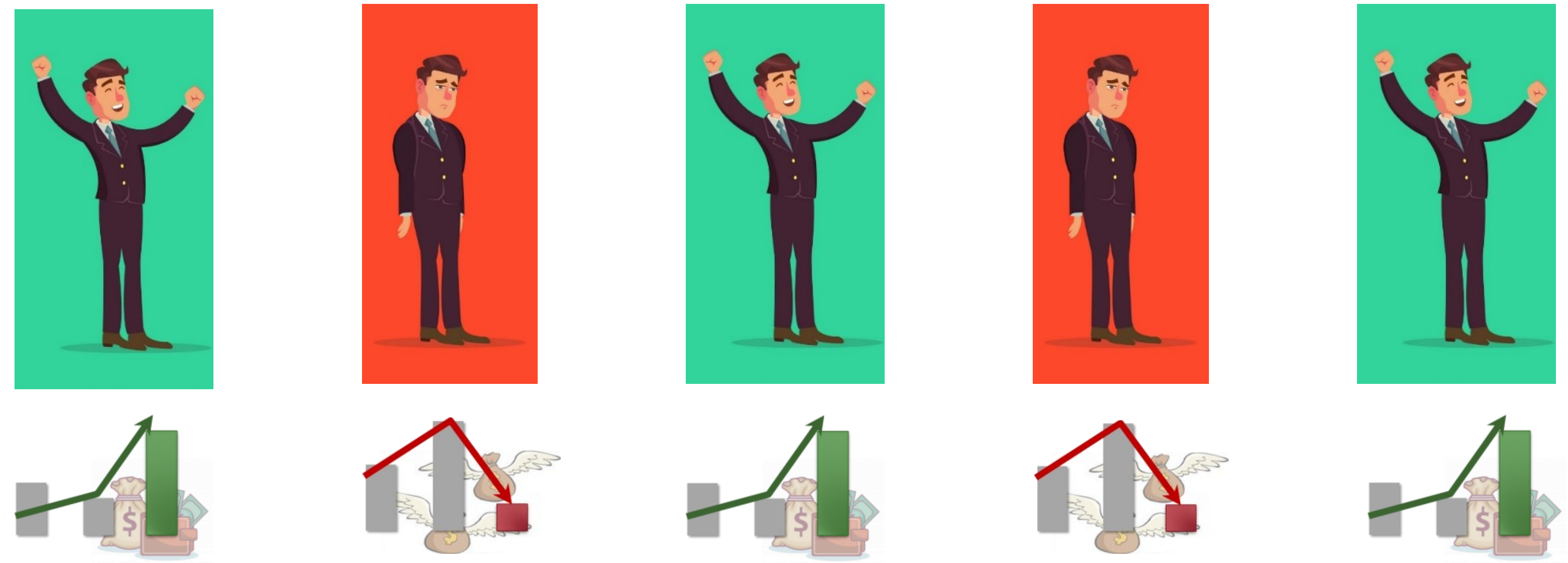


PECE Programa de Educação Continuada
Escola Politécnica da USP



ESSA SEMANA ELA CONVERSOU COM MÁRIO, E ELE ESTAVA BEM INSTÁVEL EM SEU HUMOR...

Segunda Terça Quarta Quinta Sexta



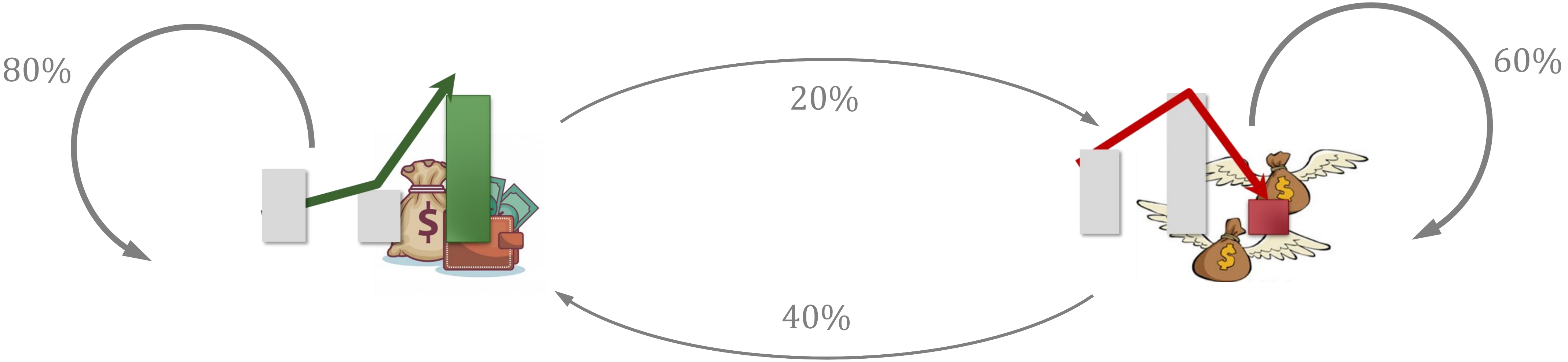
MODELO DE MARKOV OCULTO

PROBABILIDADE DE TRANSIÇÃO OU DINÂMICA:

PROBABILIDADE DE PASSAR DE UM ESTADO OCULTO X PARA OUTRO

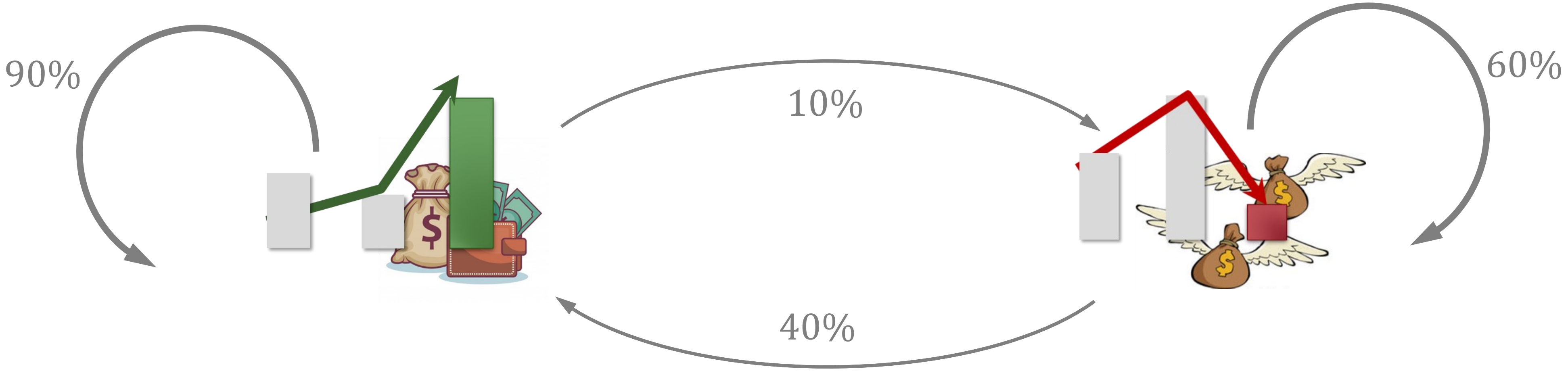
Estados $\boldsymbol{Q} = (q_1 \quad q_2) = (Alta \uparrow \quad Baixa \downarrow)$


$N = 2$ número de estados possíveis



Condições iniciais

$$\boldsymbol{\pi} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

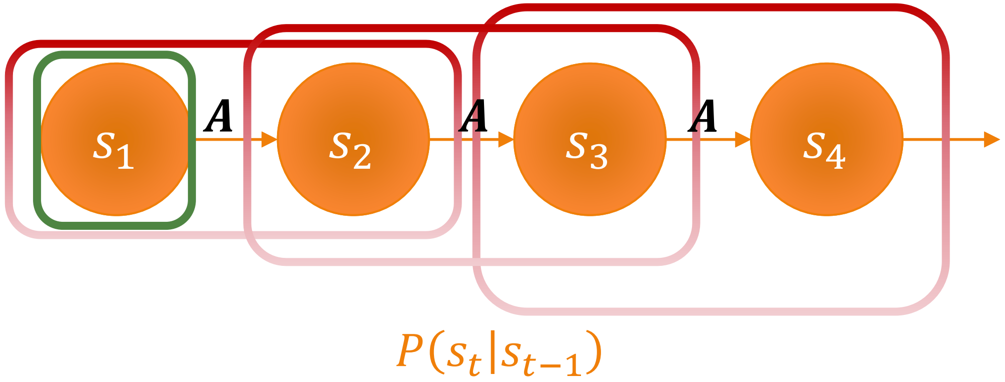


Os estados ocultos $s_i, i = 1, \dots, T$ vêm de um conjunto finito conhecido $\boldsymbol{Q}$, chamado *espaço de estado oculto*.

$$\boldsymbol{S} = \{\, s_1 \quad s_2 \quad \cdots \quad s_t \quad \cdots \quad s_T \}$$

Condições iniciais

$$\boldsymbol{\pi} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$



Matriz de transição $\boldsymbol{A}_{N\times N} \rightarrow \boldsymbol{A}_{2\times 2}$

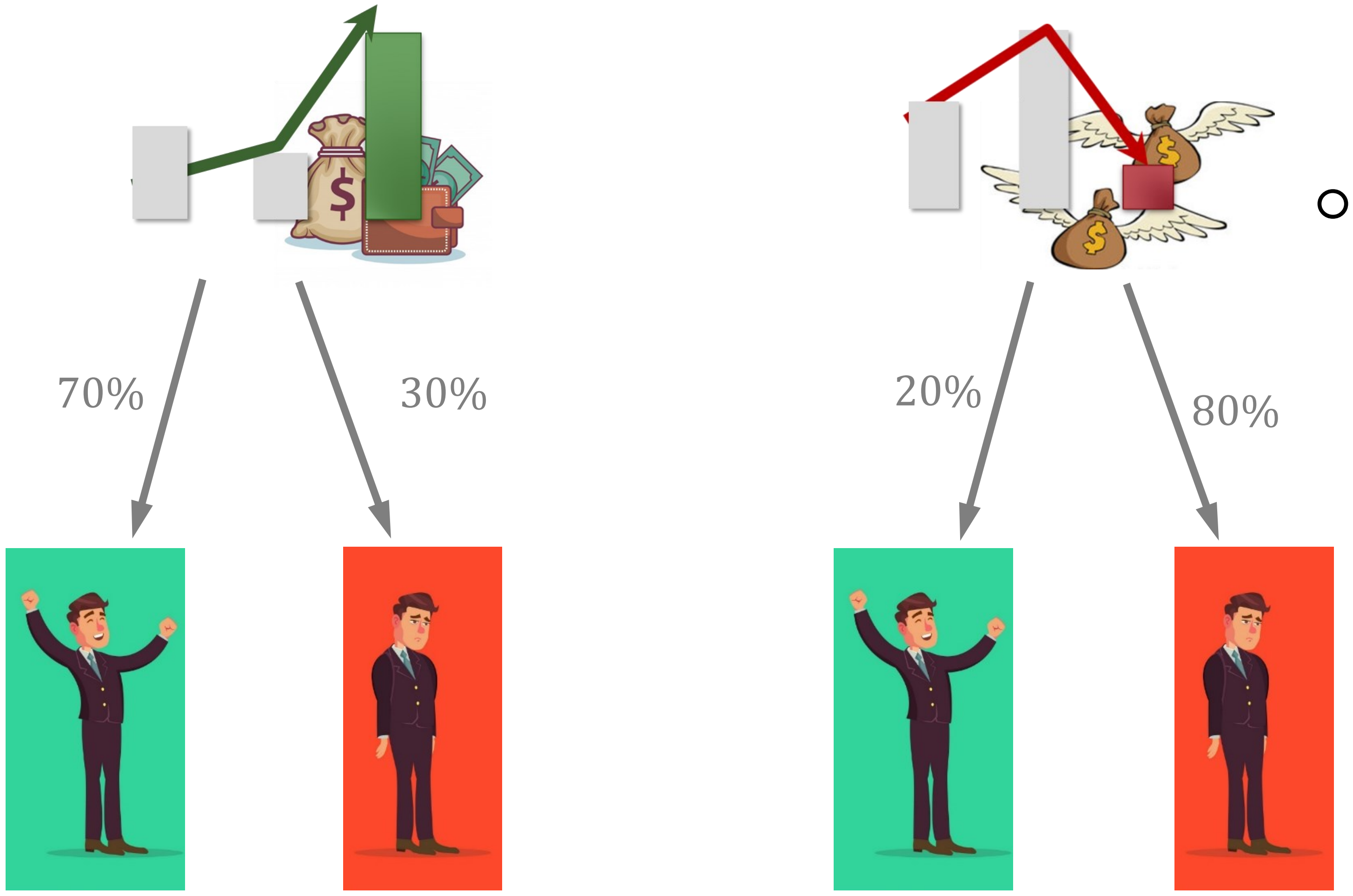
$$a(s_t, s_{t-1}) = P(s_t|s_{t-1})$$

$s_{t-1} \downarrow \quad s_t \rightarrow$	***Alta*** ↑	***Baixa*** ↓
Alta ↑	0.90	0.10
Baixa ↓	0.40	0.60

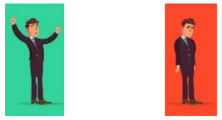
A probabilidade de estar em um estado específico na etapa t é conhecida quando sabemos em que estado estávamos na etapa $t-1$.

PROBABILIDADE DE EMISSÃO:

PROBABILIDADE DE QUE AS OBSERVAÇÕES SEJAM EMITIDAS A PARTIR DOS ESTADOS OCULTOS

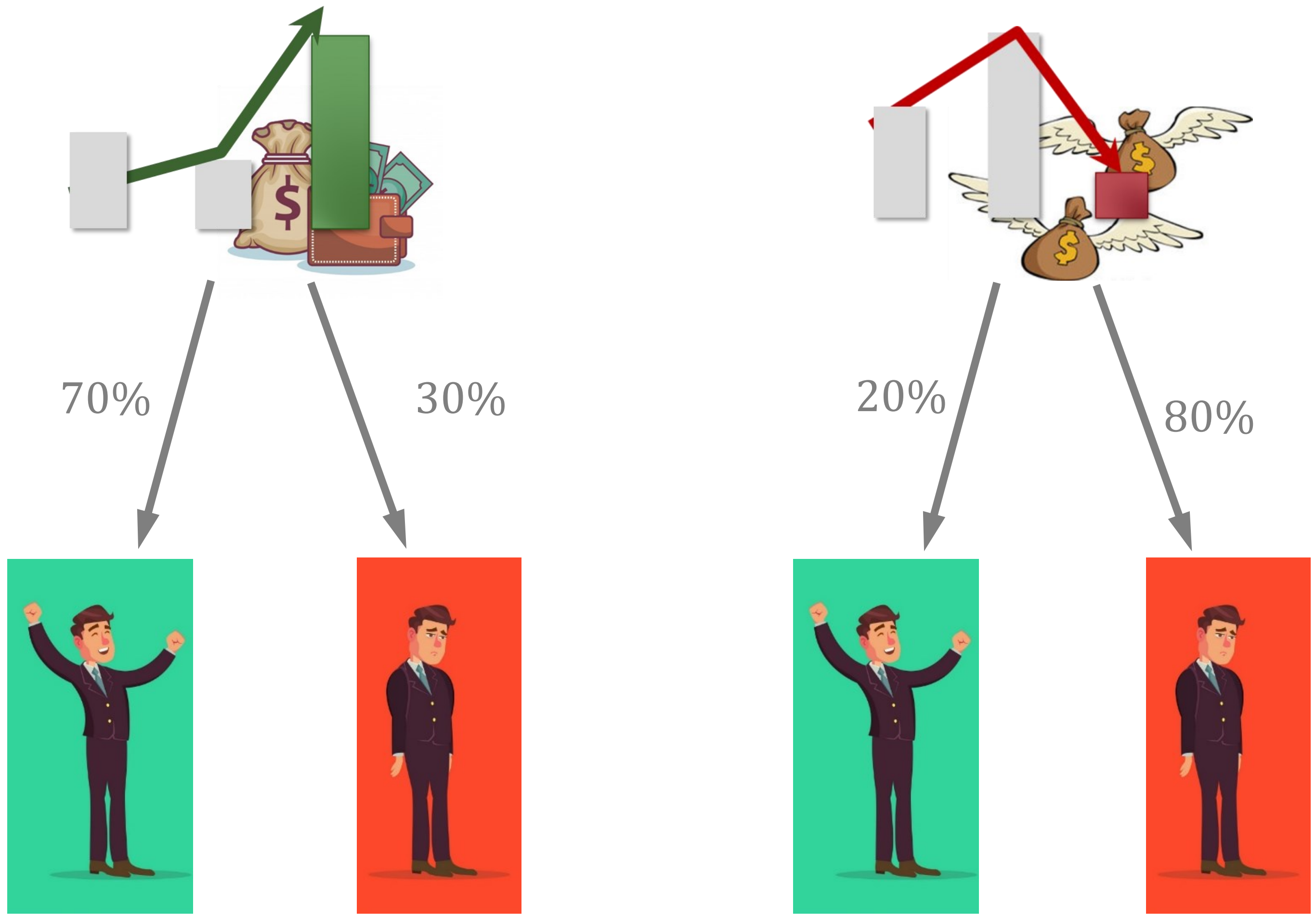


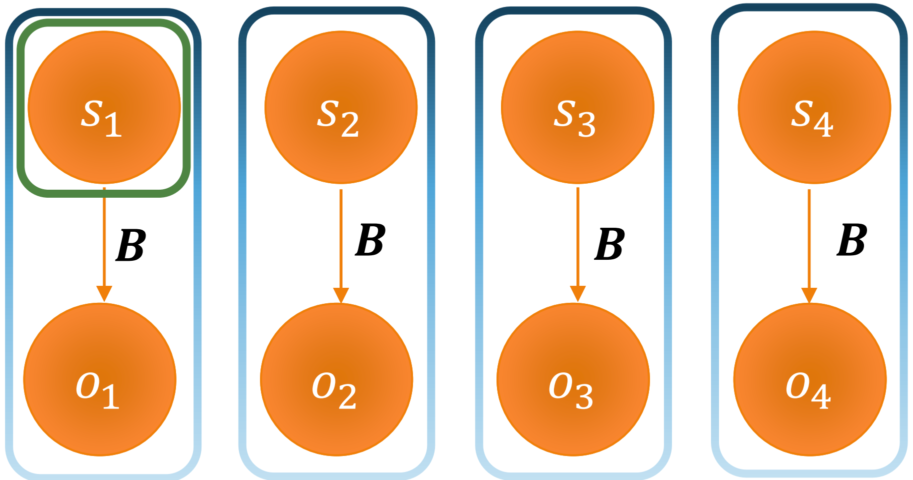
Observações $\boldsymbol{V} = (v_1 \quad v_2) = (Feliz \quad Triste)$



$N = 2$ número de estados possíveis

$V = 2$ número de observações possíveis





Matriz de emissão $\boldsymbol{B}_{N \times V} \rightarrow \boldsymbol{B}_{2\times 2}$

$$b(\, s_t, o_t) = P(\, o_t | s_t)$$

$s_t \downarrow \quad o_t \rightarrow$	***Feliz***	***Triste***
Alta ↑	0.70	0.30
Baixa ↓	0.20	0.80

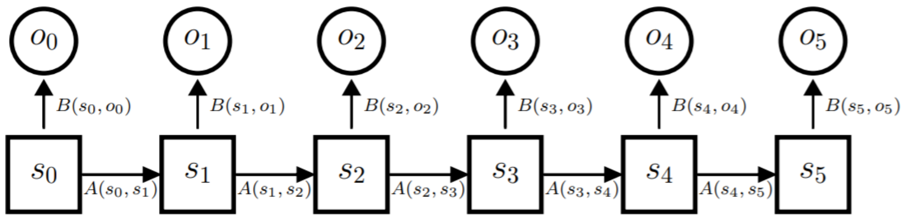
A probabilidade de ver uma emissão específica na etapa t é conhecida quando sabemos em que estado estamos na etapa t.

As observações $o_i, i = 1, \dots, T$ vêm de um conjunto finito conhecido $\boldsymbol{V}$, chamado espaço de observações.

$$\boldsymbol{O} = \{\, o_1 \quad o_2 \quad \cdots \quad o_t \quad \cdots \quad o_T\}$$

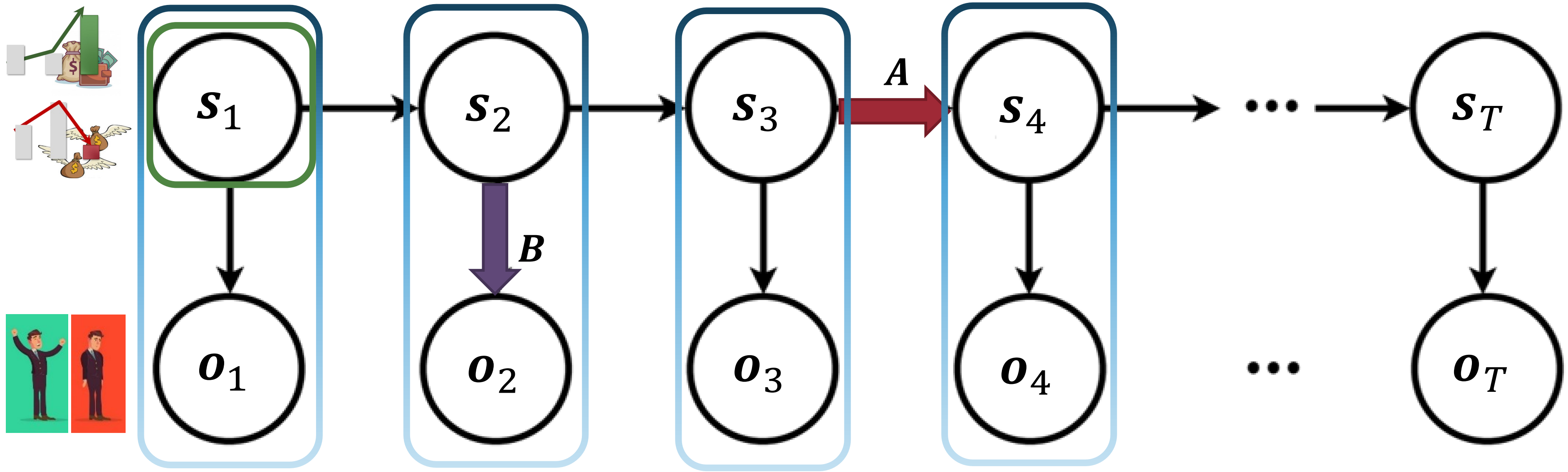
COMPONENTES DO MODELO DE MARKOV

Nome	Notação	Significado, propriedade
Espaço de estados	$\boldsymbol{Q} = (q_1, q_2, \dots, q_N)$	Conjunto de N estados ocultos possíveis
Espaço de observações	$\boldsymbol{V} = (v_1, v_2, \dots, v_V)$	Conjunto de V estados observáveis
Sequência de estados	$\boldsymbol{S} = (s_1, s_2, \dots, s_T)$	Sequência de T estados, $s_i \in \boldsymbol{Q}$
Sequência de observações	$\boldsymbol{O} = (o_1, o_2, \dots, o_T)$	Sequência de T observações, $o_i \in \boldsymbol{V}$
Probabilidade de transmissão	$\boldsymbol{A} \in \mathbb{R}^{N \times N}$	Cada linha é uma distribuição de probabilidade $a(s_t, s_{t-1}) = P(s_t \mid s_{t-1})$
Probabilidade de emissão	$\boldsymbol{B} \in \mathbb{R}^{N \times V}$	Cada linha é uma distribuição de probabilidade $b(s_t, o_t) = P(o_t \mid s_t)$
Condições iniciais	$\boldsymbol{\pi}$	São os estados iniciais



PROBABILIDADE CONJUNTA

$\boldsymbol{S} = \{s_1 \quad s_2 \quad \cdots \quad s_T\}$ é uma sequência de estados. **Nós não observamos $\boldsymbol{S}$.**
$\boldsymbol{O} = \{o_1 \quad o_2 \quad \cdots \quad o_T\}$ é uma sequência de emissões. **Nós observamos $\boldsymbol{O}$.**

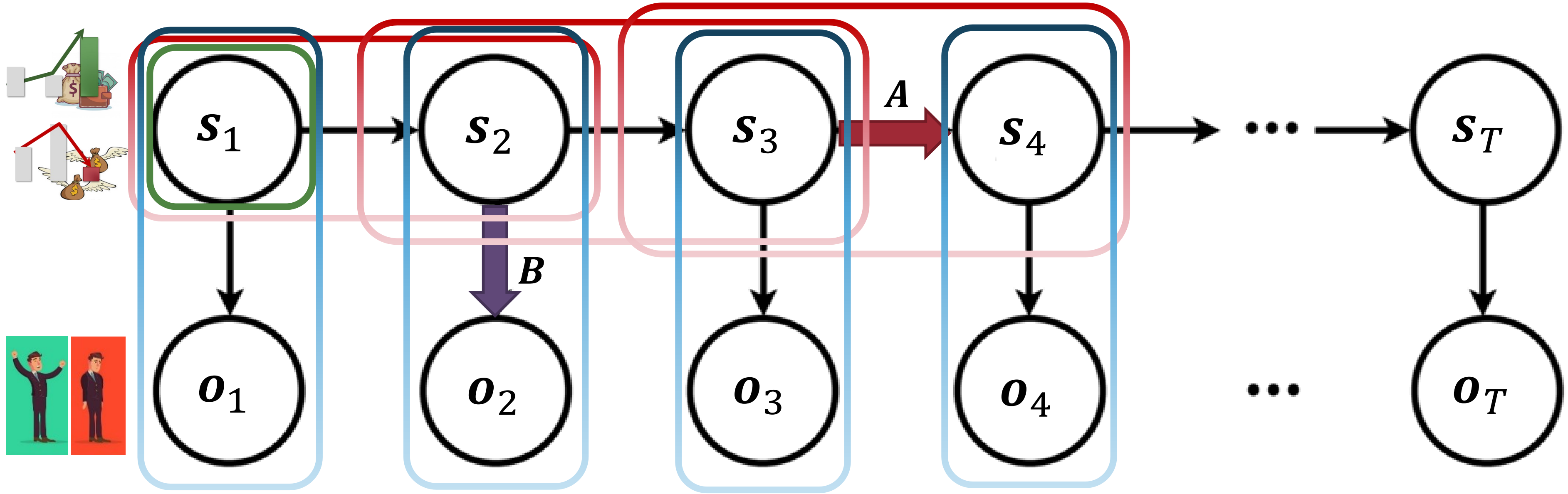


$b(s_t, o_t) = P(o_t|s_t)$

$\boldsymbol{o}_2$ é independente condicionalmente, uma vez que conhecemos $\boldsymbol{s}_2$

PROBABILIDADE CONJUNTA

$\boldsymbol{S} = \{s_1 \quad s_2 \quad \cdots \quad s_T\}$ é uma sequência de estados. **Nós não observamos $\boldsymbol{S}$.**
$\boldsymbol{O} = \{o_1 \quad o_2 \quad \cdots \quad o_T\}$ é uma sequência de emissões. **Nós observamos $\boldsymbol{O}$.**

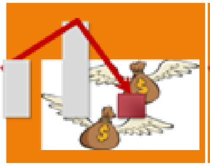
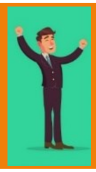
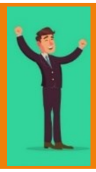
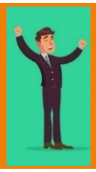
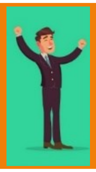


$$a(s_t, s_{t-1}) = P(s_t | s_{t-1})$$

$\boldsymbol{o}_2$ é independente condicionalmente, uma vez que conhecemos $\boldsymbol{s}_2$
$\boldsymbol{s}_4$ é independente condicionalmente, uma vez que conhecemos $\boldsymbol{s}_3$

$$P(s_1, s_2, \dots, s_T, o_1, o_2, \dots, o_T) = \pi_{s_0} \prod_{t=1}^{T} P(o_t|s_t) \prod_{t=2}^{T} P(s_t|s_{t-1}) = \prod_{t=1}^{T} P(o_t|s_t)\, P(s_t|s_{t-1})$$

DADO MEU MODELO $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}$, QUAL A PROBABILIDADE CONJUNTA $\boldsymbol{S}$ E $\boldsymbol{O}$?

$\boldsymbol{S}$										
$\boldsymbol{O}$										
$\boldsymbol{P(o_t \mid s_t)}$	0.8	0.2	0.8	0.7	0.7	0.7	0.8	0.2	0.8	0.3
$\boldsymbol{P(s_t \mid s_{t-1})}$	1/3	0.6	0.6	0.4	0.9	0.9	0.1	0.6	0.6	0.4

$$P(s_1, s_2, \dots, s_T, o_1, o_2, \dots, o_T) = 1/3 \times 0.2^2 \times 0.3 \times 0.4^2 \times 0.6^4 \times 0.7^3 \times 0.8^4 \times 0.9^2 = 9.44 \times 10^{-6}$$

$$\boldsymbol{\pi} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

Matriz de transição $a(s_t, s_{t-1}) = P(s_t \mid s_{t-1})$

$\boldsymbol{s_{t-1} \mid s_t}$	**Alta ↑**	**Baixa ↓**
Alta ↑	0.90	0.10
Baixa ↓	0.40	0.60

Matriz de emissão $b(s_t, o_t) = P(o_t \mid s_t)$

$\boldsymbol{s_t \mid o_t}$	**Feliz**	**Triste**
Alta ↑	0.70	0.30
Baixa ↓	0.20	0.80

ONDE APLICAR O MODELO DE MARKOV OCULTO

PECE Programa de Educação Continuada
Escola Politécnica da USP

1. **Problema de avaliação**

   Dada uma sequência de observações $\boldsymbol{O}$ e o modelo $\{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}\}$, como computamos a probabilidade $P(\boldsymbol{O}|\{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}\})$?

2. **Problema de decodificação**

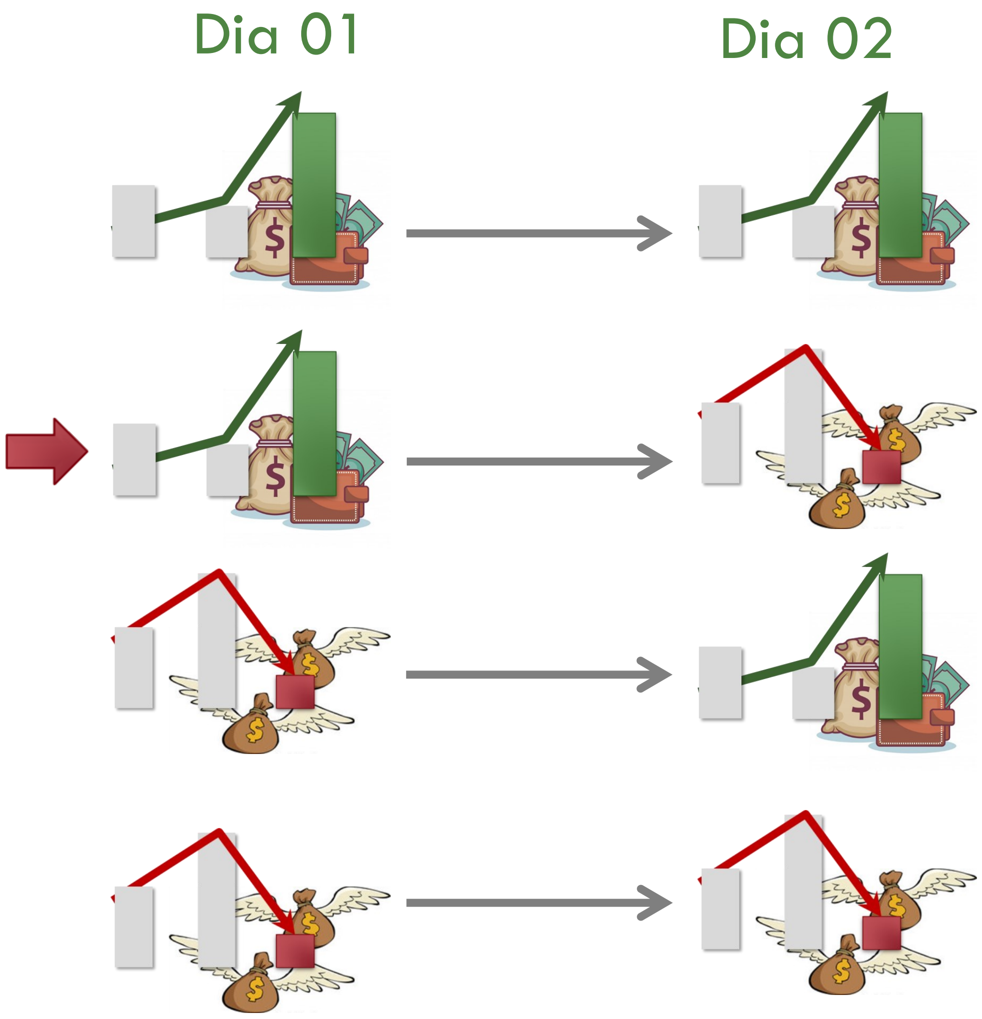
   Dada uma sequência de observações $\boldsymbol{O}$ e o modelo $\{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}\}$, qual a melhor sequência de estados dentro do modelo capaz de gerar essas observações?

3. **Problema de treinamento**

   Como ajustamos os parâmetros do modelo $\{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}\}$ para maximizar a probabilidade $P(\boldsymbol{O}|\{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}\})$?

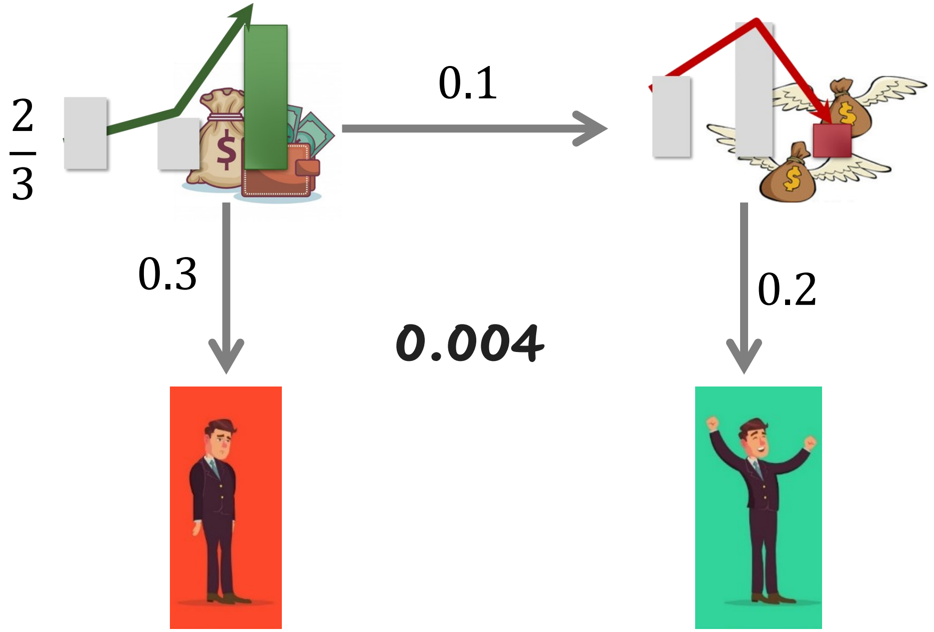
POR EXEMPLO, A PERGUNTA É:

Se Mário está triste, feliz qual o comportamento provável da bolsa?

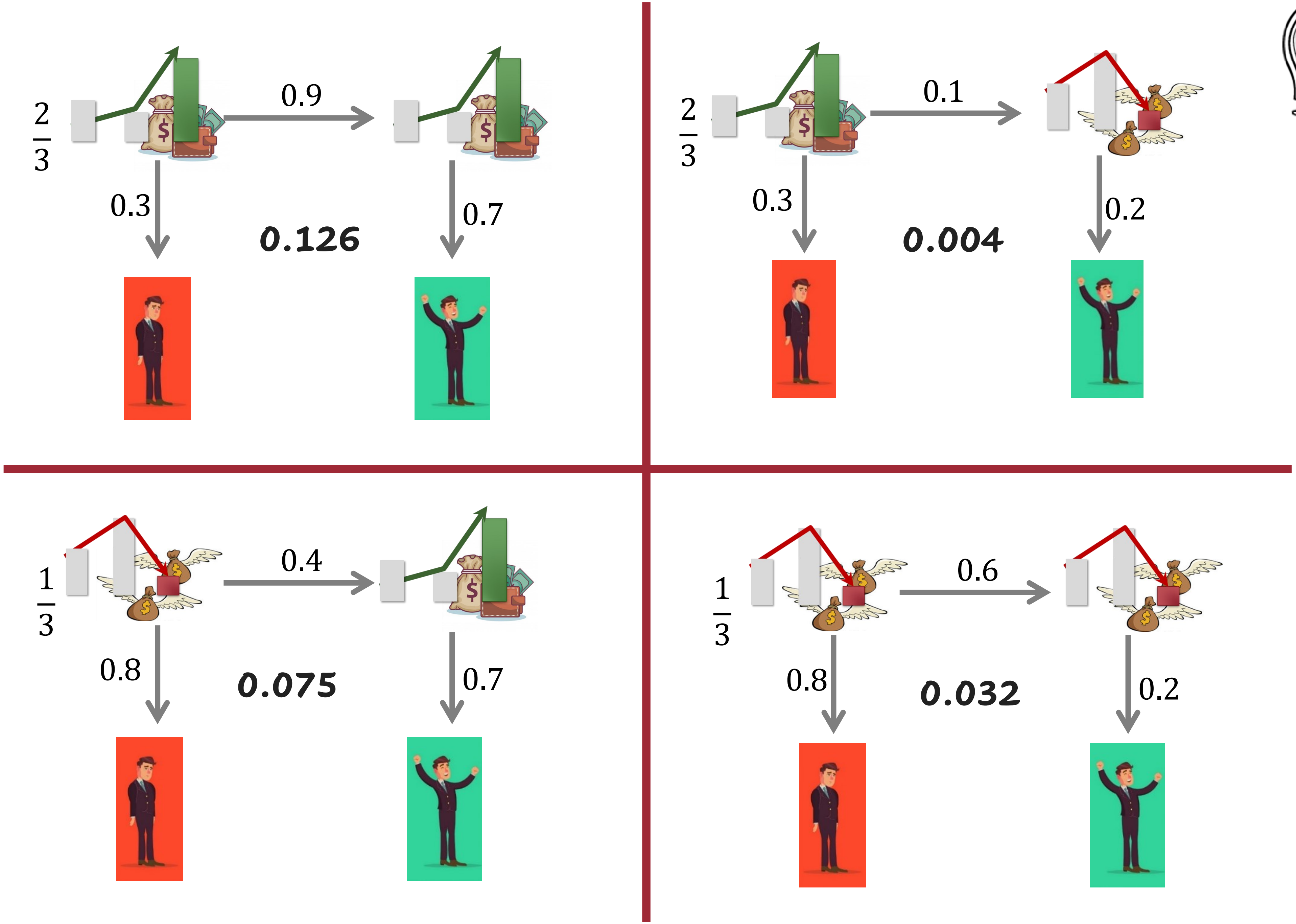


Vamos pegar todas as situações, calcular as probabilidades, e escolher a situação que nos dá a probabilidade mais alta.
Chama-se **estimação de máxima verossimilhança** (maximum likelihood estimation).

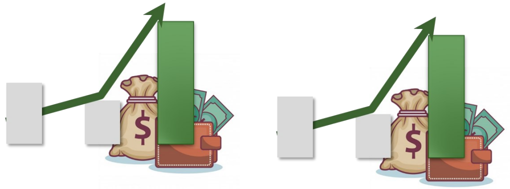
PECE Programa de Educação Continuada
Escola Politécnica da USP



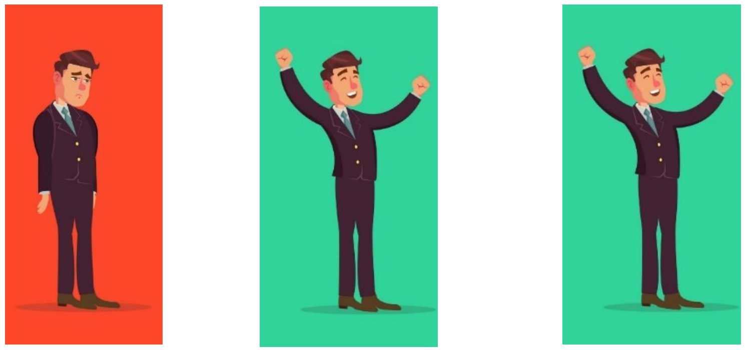
PECE Programa de Educação Continuada
Escola Politécnica da USP

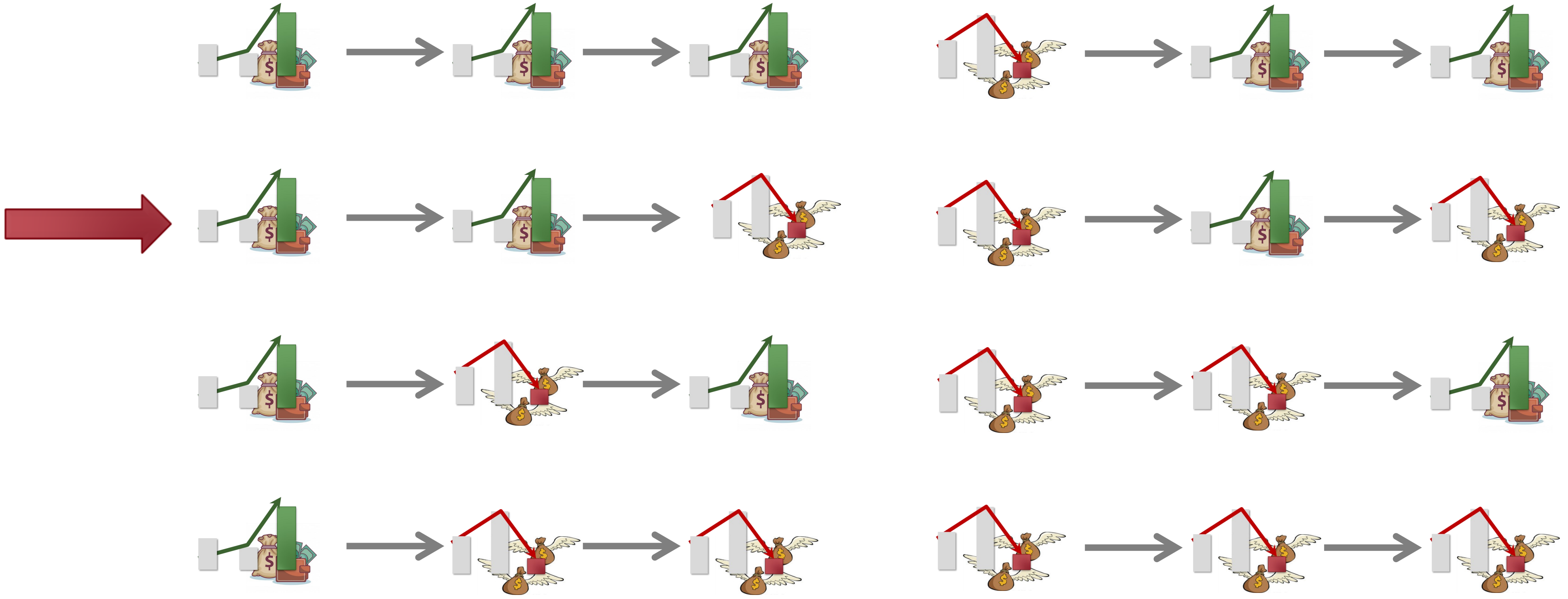


CONCLUSÃO DE ANA...



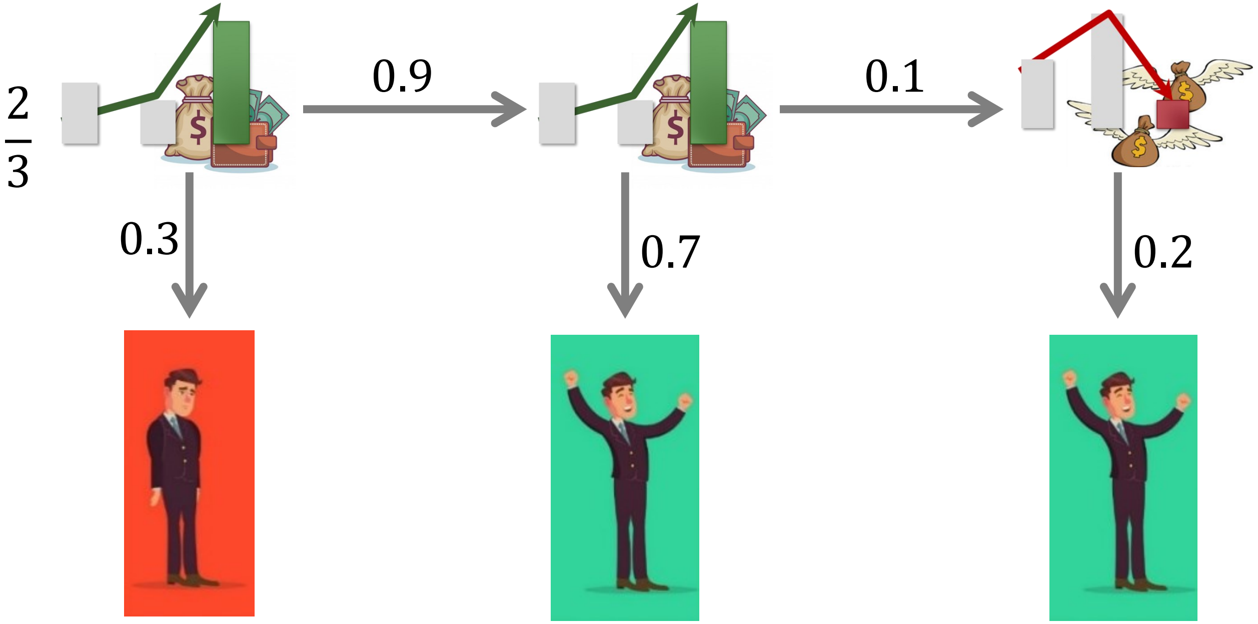
Dia 01 Dia 02 Dia 03



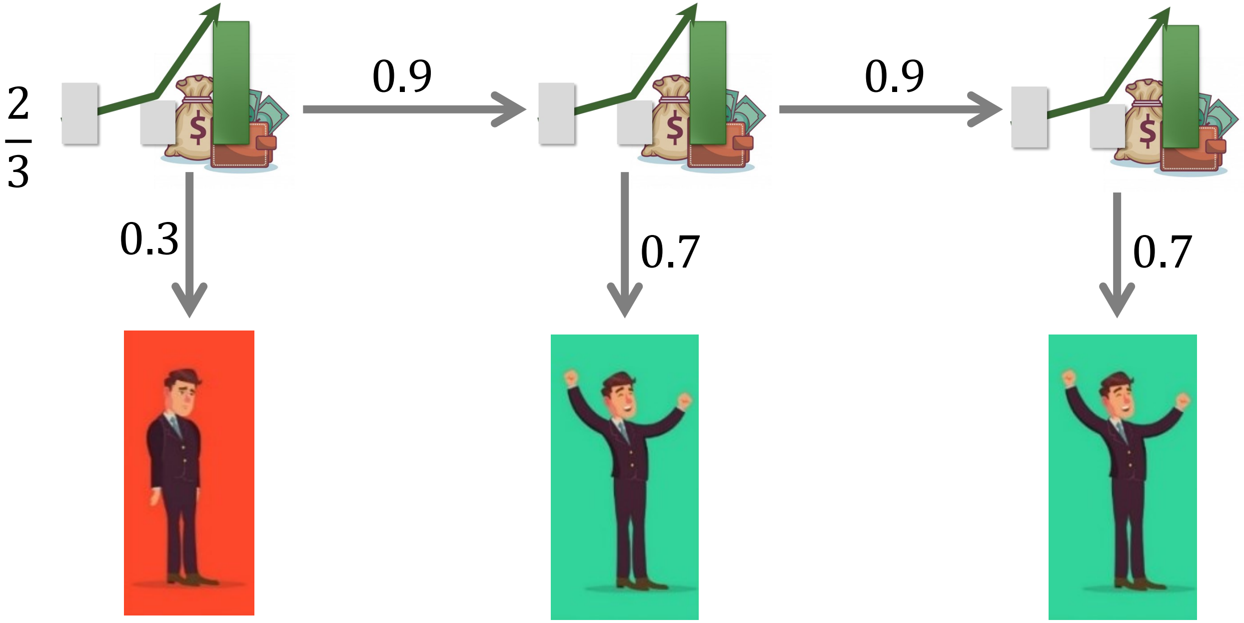


PECE Programa de Educação Continuada
Escola Politécnica da USP

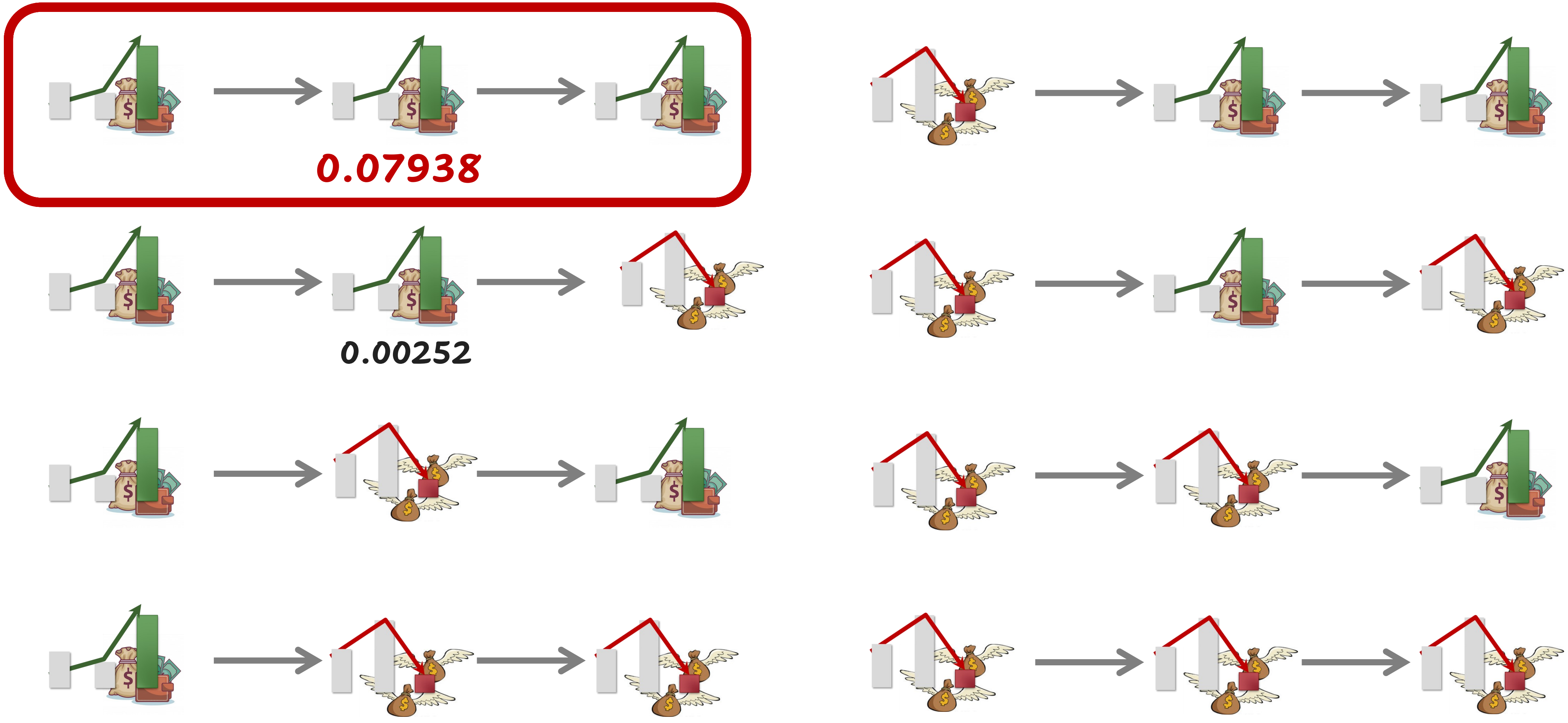
PECE Programa de Educação Continuada
Escola Politécnica da USP

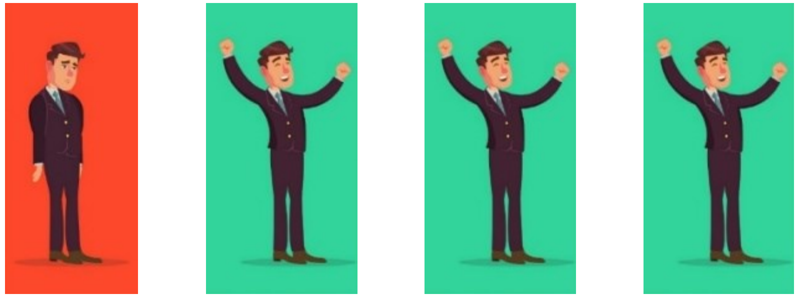


0.00252



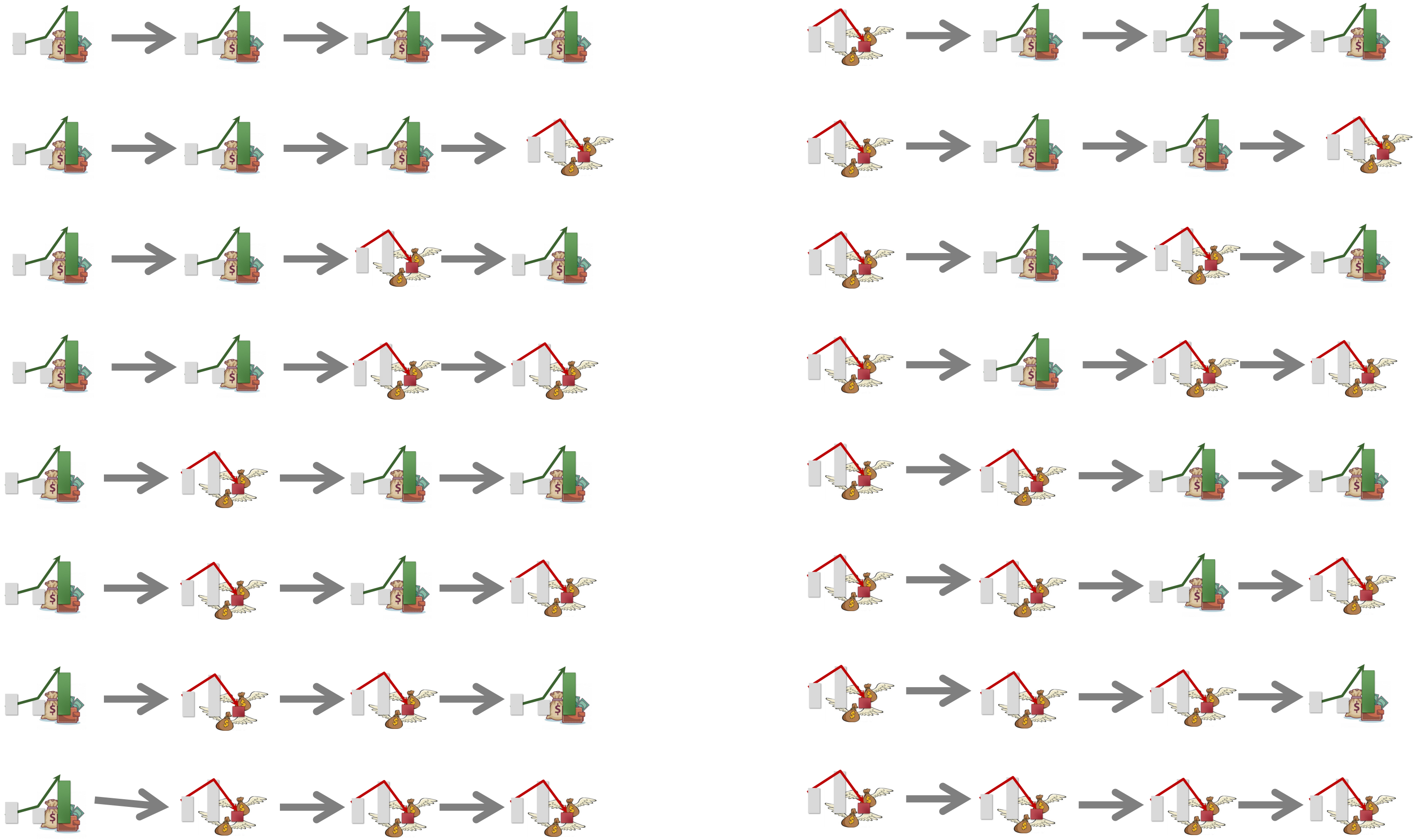
0.07938


0.07938
0.00252



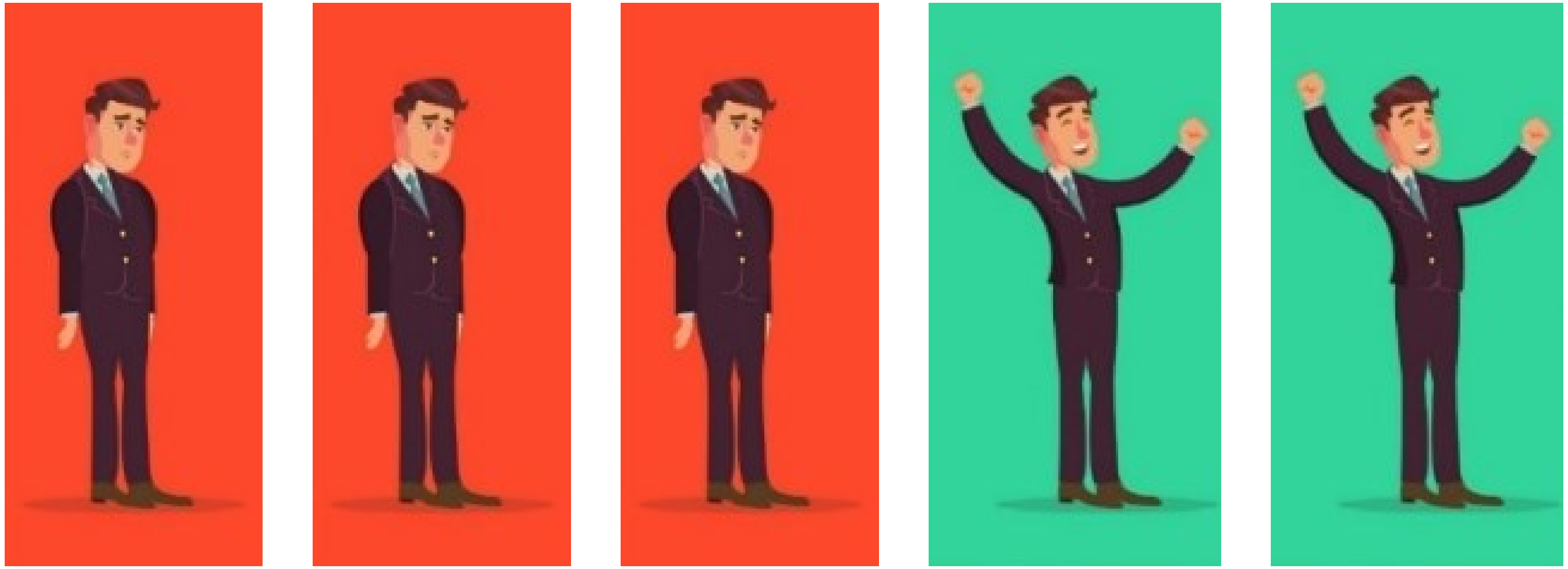
16 casos...

Cresce exponencialmente. Temos que achar um caminho mais inteligente do que testar todos os casos.



UMA SEMANA DE OBSERVAÇÃO...

Algoritmo Viterbi



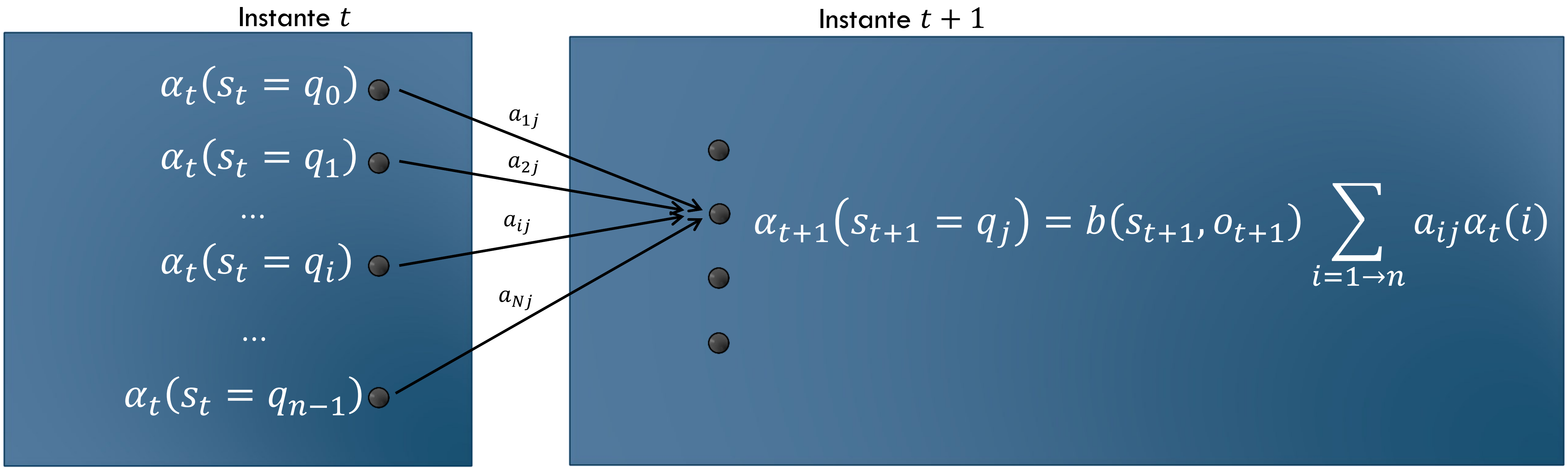
ALGORITMO FORWARD OU $\alpha - Pass$

Define-se $\alpha_t(s_t = q')$ como a probabilidade de estar no estado $s_t \in \boldsymbol{Q} = \{q_0, q_1, \dots, q_{T-1}\}$ depois das t primeiras observações $o_{1:t}$, dado HMM definido por $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}$. Dessa forma,

$$\alpha_t(s_t = q') = P(o_{1:t}, s_t = q')$$

$$\alpha_t(s_t) = P(o_t|s_t) \sum_{s_{t-1}} P(s_t|s_{t-1}) P(o_{1:t-1}, s_{t-1}) = b(s_t, o_t) \sum_{s_{t-1}} a(s_t, s_{t-1}) \alpha_{t-1}(s_{t-1})$$

$$\alpha_t(s_t) = b(\, s_t, o_t) \sum_{j=0}^{n-1} \alpha_{t-1}(q_j) a(s_t, q_j)$$



ALGORITMO VITERBI

Ao invés de verificar do início até o final de todas as opções possíveis, começaremos a verificar passo a passo (para $1, 2, \dots, t, t+1, \dots, n$) e armazenamos o valor da probabilidade máxima de cada passo.

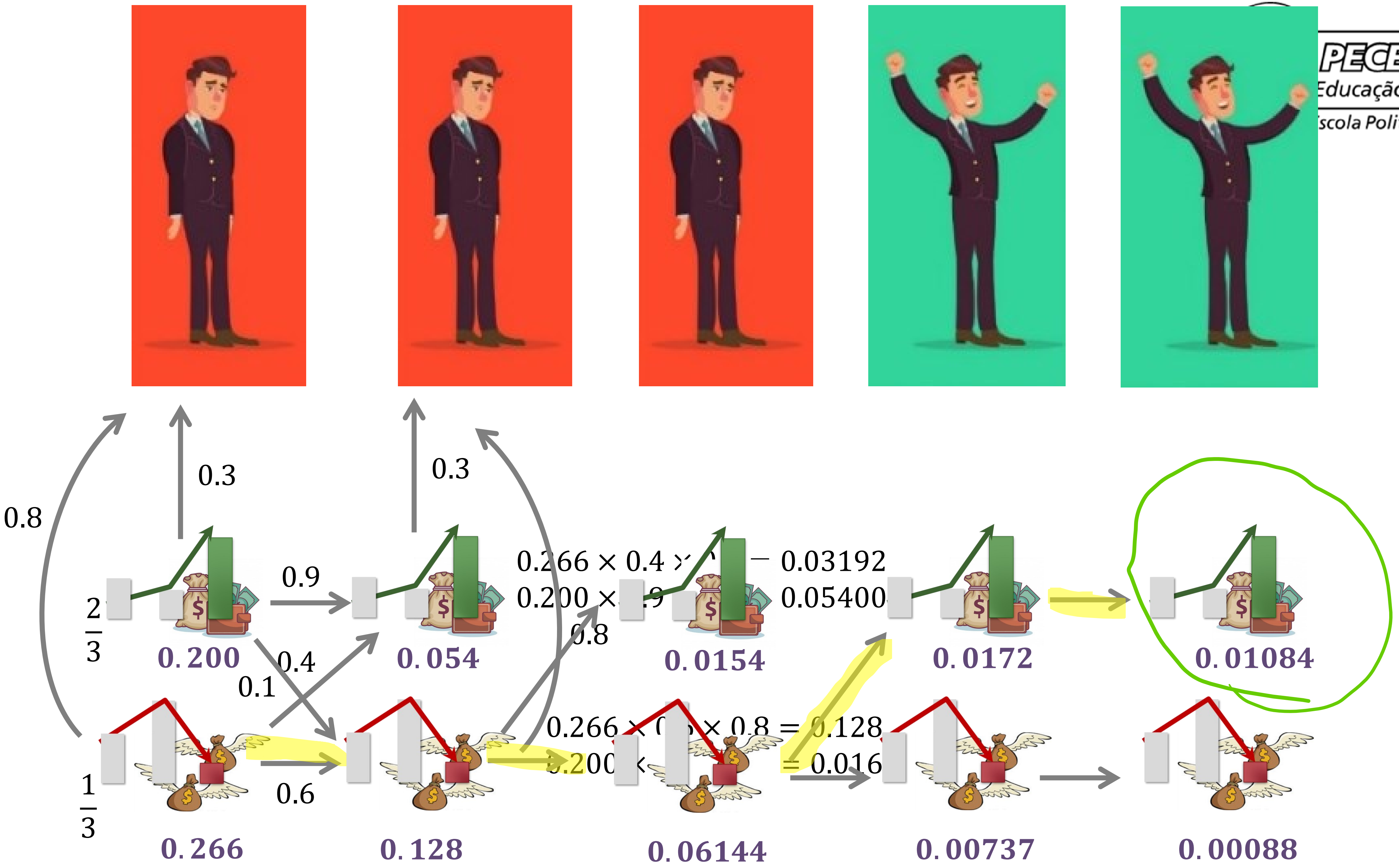
Ao final, comparamos todas as opções.

$$P(s_1, s_2, \dots, s_T, o_1, o_2, \dots, o_T) = \prod_{t=1}^{T} P(o_t|s_t)\, P(s_t|s_{t-1})$$

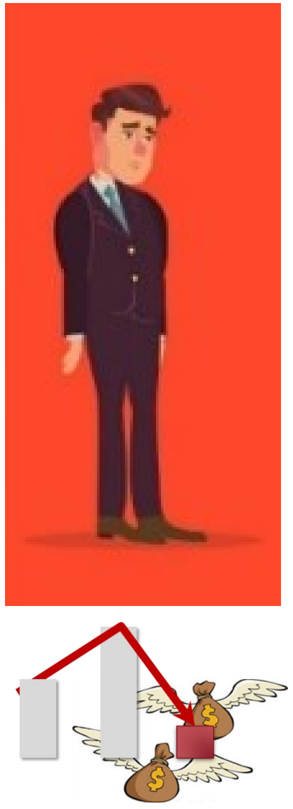
Na verdade, o algoritmo é o mesmo que o *forward*, mas, ao invés da somatória, usamos o valor máximo,

$$\alpha_t(s_t) = \max_{s_{t-1}} \alpha_{t-1}(s_{t-1}) a(s_t, s_{t-1}) b(s_t, o_t)$$

PECE Programa de
Educação Continuada
scola Politécnica da USP

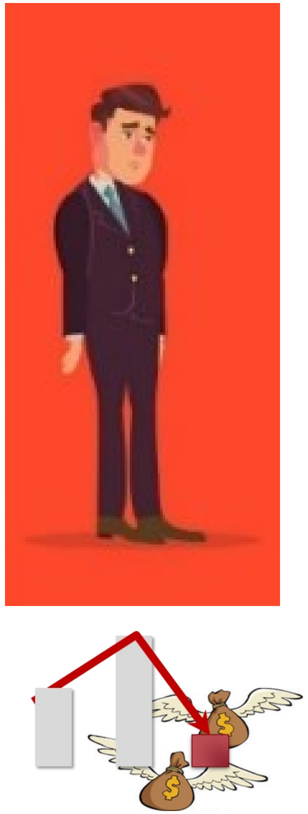


PECE Programa de Educação Continuada
Escola Politécnica da USP



0.266

0.128



0.06144

0.0172

0.01084

LIÇÃO FINAL

PECE Programa de Educação Continuada
Escola Politécnica da USP

A aula de hoje, apesar de complexa, foi apenas uma introdução aos modelos de Markov, que serão muito úteis em nossas próximas etapas. Na lição de casa, você resolverá um problema mais elaborado, com até 3 colegas. **A nota desta lição, substituirá sua menor nota.**

Suponha que alguém lance uma moeda dez vezes e forneça o resultado:. Porém, o lançador tem duas moedas e muda aleatoriamente de uma moeda para outra. Uma moeda é equilibrada E e a outra é tendenciosa NE, que resulta em cara 80% das vezes que é lançada. Após cada lançamento, o jogador pode manter a mesma moeda (probabilidade $0{,}8$), mudar para a outra moeda ($0{,}20$). Além disso, a probabilidade de começar com a moeda equilibrada é $0{,}8$. Com essas informações,

1. Monte o problema e as matrizes;
2. Qual a probabilidade da sequência $CKKCCCCCKC$, onde C é cara e K é coroa, ocorrer?
3. Qual a sequência mais provável das moedas utilizadas?
4. Qual a probabilidade dessa sequência ocorrer cem vezes? Estude o fenômeno de *underflow* e proponha a solução.

ACABOU...

Reveja a aula antes de resolver os exercícios.