

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**PROPOSTA DE UMA
REPRESENTAÇÃO TENSORIAL
PARA MODELOS
MARKOVIANOS OCULTOS**

LUCIANA DA SILVEIRA ESPINDOLA

Dissertação apresentada como requisito parcial
à obtenção do grau de Mestre em Ciência da
Computação na Pontifícia Universidade Católica
do Rio Grande do Sul.

Orientador: Paulo Henrique Lemelle Fernandes

**Porto Alegre
2011**

Dados Internacionais de Catalogação na Publicação (CIP)

E77p Espindola, Luciana da Silveira
Proposta de uma representação tensorial para modelos markovianos ocultos / Luciana da Silveira Espindola. – Porto Alegre, 2011.
113 f.

Diss. (Mestrado) – Fac. de Informática, PUCRS.
Orientador: Prof. Dr. Paulo Henrique Lemelle Fernandes.

1. Informática. 2. Redes de Autômatos Estocásticos.
3. Cadeias de Markov. 4. Simulação e Modelagem em Computadores. I. Fernandes, Paulo Henrique Lemelle.
II. Título.

CDD 003.3

**Ficha Catalográfica elaborada pelo
Setor de Tratamento da Informação da BC-PUCRS**



Pontifícia Universidade Católica do Rio Grande do Sul
FACULDADE DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

TERMO DE APRESENTAÇÃO DE DISSERTAÇÃO DE MESTRADO

Dissertação intitulada "**Proposta de Uma Representação Tensorial para Modelos Markovianos Ocultos**", apresentada por Luciana da Silveira Espindola, como parte dos requisitos para obtenção do grau de Mestre em Ciência da Computação, Processamento Paralelo e Distribuído, aprovada em 17/03/2011 pela Comissão Examinadora:

Prof. Dr. Paulo Henrique Lemelle Fernandes -
Orientador

PPPGCC/PUCRS

Prof. Dr. Fernando Luís Dotti -

PPGCC/PUCRS

Dr. Afonso Henrique Corrêa de Sales -

FACIN/PNPD

Homologada em 24/05/11, conforme Ata No. 008 pela Comissão Coordenadora.

Prof. Dr. Fernando Luís Dotti.
Coordenador.

PUCRS

Campus Central

Av. Ipiranga, 6681 - P32- sala 507 - CEP: 90619-900
Fone: (51) 3320-3611 - Fax (51) 3320-3621
E-mail: ppgcc@pucrs.br
www.pucrs.br/facin/pos

*Dedico aos meus pais, Lucia e Luiz,
e ao meu marido, Mauricio.*

AGRADECIMENTOS

Um conjunto de aleatoriedades me trouxe até aqui, sendo de certa forma irônico que eu tenha recebido o título de mestre por apresentar um método de conversão entre modelos estocásticos: do aleatório ao aleatório.

Iniciei no Programa de Pós-Graduação da PUCRS como aluna especial. Logo tornei-me aluna regular, visto o incentivo explícito dos professores Milene Silveira e Duncan Ruiz, que valorizaram meu empenho e me fizeram crer na possibilidade de uma carreira acadêmica. A vocês, meus agradecimentos, admiração e respeito.

Não me contive em satisfação ao saber que havia sido selecionada para o mestrado. Na primeira conversa com meu futuro orientador veio a responsabilidade: fui escolhida pela minha formação em física, o que criava expectativas sobre um futuro brilhante como modeladora. Não sou delirante a ponto de achar que minha modesta contribuição faça jus a tão exagerada expectativa, contudo, agradeço a confiança e a oportunidade. Meu orientador, Paulo Fernandes, me apoiou, apostou em mim, me tranquilizou em várias situações e, o mais importante de tudo, me deu um norte. Você não deixou de acreditar. Obrigada!

Poderia-se dizer que quatro pessoas fizeram as vezes de coorientadores. Ora um, ora outro, todos foram importantes, cada qual contribuindo à sua maneira. Agradeço a Marcelo Thielo por me incentivar, ajudar a desenvolver minha compreensão sobre meu objeto de estudo e se mostrar solícito em inúmeras situações, quando precisei conversar com alguém que entendesse minhas angústias. Agradeço a Thais Webber por me ajudar a organizar e estruturar meu trabalho, dando sugestões que se mostaram valiosas. Agradeço a Ricardo Czekster, que sempre trouxe questionamentos, dúvidas e desafios, sempre disposto a discutir e ampliar os horizontes. Por fim, agradeço a Afonso Sales, que talvez tenha sido um dos poucos com quem pude conversar de igual para igual; sempre esteve interessado no meu trabalho e contribuiu enormemente com idéias e perguntas que me fizeram refletir. Obrigada a todos!

Tive a felicidade de conviver com colegas maravilhosos, sem os quais não sei o que faria. Em especial, agradeço a Antonio Lima, Paulo Zanin e Rui Ribeiro, sempre presentes, bons amigos. Vocês me ajudaram a superar dificuldades. Obrigada! Apesar de ter por todos apreço em iguais proporções, destaco Paulo Zanin como meu maior contribuinte. Nós, orientandos do mesmo orientador, aprendemos a modelar, tivemos as mesma dúvidas, trocamos idéias e crescemos juntos. Obrigada!

Tive vários revisores e opinantes e todos merecem reconhecimento, a saber: Dione Taschetto, Eduardo Fabris, Élder Bernardi, Felipe Christmann, Gabriel Pozza, Guilherme Fedrizzi, Ingrid Faber, Joaquim Assunção, Leandro Pasetto, Luciano Preissler, Mauricio Porto, dentre outros. Minha irmã, Juliana Espindola, contudo, foi minha principal revisora, lendo e relendo o que eu escrevia e reescrevia. A Juliana foi tão presente que não tenho dúvidas de que palestraria sobre o tema de minha dissertação em qualquer hora, qualquer lugar, para qualquer público. Obrigada por estares sempre disponível, sempre tão prestativa, muitas vezes em detrimento do teu próprio trabalho. Obrigada!

Não poderia deixar de fazer um agradecimento geral a todos que se preocuparam, torceram e acompanharam como espectadores esses dois anos de muitos desafios. São amigos e parentes, pessoas que amo muito. Em especial, agradeço à minha família de berço por sempre apostar, mesmo sem indício algum de futuro sucesso, sempre no escuro e sempre acreditando, porque família que é família é assim mesmo! Não menos importante, agradeço à minha família adquirida: meu marido, Mauricio Schramm, e meus filhos, Gatolino, Gatolina e Gatofredo. Vocês me deram tranquilidade, carinho e o suporte emocional que eu precisava, me ajudando de formas inimagináveis. Sem vocês eu não conseguiria. Obrigada!

Luciana da Silveira Espindola

PROPOSTA DE UMA REPRESENTAÇÃO TENSORIAL PARA MODELOS MARKOVIANOS OCULTOS

RESUMO

O propósito desta dissertação é propor uma representação tensorial para Modelos Markovianos Ocultos (*Hidden Markov Models* – HMM). A forma escolhida para alcançar esse objetivo passa pelo estudo de como converter um modelo HMM em um modelo SAN (*Stochastic Automata Networks*): estruturado e cujo formato tensorial é conhecido. A estratégia de conversão consiste na criação de dois autômatos, um correspondendo à cadeia de Markov oculta e outro para representar as emissões do modelo HMM. Esses autômatos se relacionam por transições sincronizadas e dependências funcionais são definidas. Um passo intermediário é necessário para mostrar a equivalência entre as representações SAN e HMM, sendo este passo a obtenção de uma cadeia de Markov global capaz de representar o modelo HMM. A igualdade entre as cadeias de Markov globais obtidas a partir de ambos os formalismos SAN e HMM constitui a prova de equivalência.

Palavras-chave: Modelagem de sistemas; Formalismos estruturados; Cadeias de Markov; Modelos Markovianos Ocultos; Modelos Ocultos de Markov; Redes de Autômatos Estocásticos; Descritor Markoviano; Álgebra Tensorial.

PROPOSAL OF A TENSOR REPRESENTATION FOR HIDDEN MARKOV MODELS

ABSTRACT

The purpose of this Master Thesis is to propose a tensor representation for Hidden Markov Models (HMM). The chosen way to reach this goal goes through the study of how to convert an HMM into a SAN model (Stochastic Automata Networks – SAN): structured and with a known tensor format. The conversion strategy consists on the the creation of two automata, one corresponding to the hidden Markov chain and another to represent the HMM model emissions. These automata interact with each other by means of synchronized transitions and some defined functional dependencies. An intermediate step is necessary to show the equivalence between the SAN and HMM representations, being this step the obtainment of a global Markov chain capable of representing the HMM model. The equality between the global Markov chains obtained from both the SAN and HMM formalisms constitutes the equivalence proof.

Keywords: System modeling; Structured formalisms; Markov Chains; Hidden Markov Models; Stochastic Automata Networks; Markovian Descriptors; Tensor Algebra.

LISTA DE FIGURAS

Figura 3.1	Uma Cadeia de Markov	42
Figura 3.2	Um Modelo SAN com Dois Autômatos	42
Figura 3.3	Um Modelo SAN com Dois Autômatos	46
Figura 3.4	Uma Cadeia de Markov	47
Figura 3.5	Modelo SAN para o Jogo “Cara ou Coroa”	51
Figura 3.6	Modelo SAN para o Jogo “Dois ou Um” – Abordagem que destaca os vencedores	54
Figura 3.7	Modelo SAN para o Jogo “Dois ou Um” – Abordagem Simplificada	55
Figura 3.8	Modelo SAN para o Jogo “Dois ou Um” – Inclusão de um Autômato de Memória	56
Figura 4.1	Algoritmo <i>Forward-Backward</i> – Parte <i>Forward</i> Ilustrada (Fonte: Rabiner [26])	62
Figura 4.2	Algoritmo <i>Forward-Backward</i> – Parte <i>Backward</i> Ilustrada (Fonte: Rabiner [26])	64
Figura 4.3	Algoritmo <i>Forward-Backward</i> – Solução do <i>Problema 3</i> (Fonte: Rabiner [26])	68
Figura 4.4	Experimento com Lançamento de Moedas – HMM com Uma Moeda	70
Figura 4.5	Experimento com Lançamento de Moedas – HMM com Duas Moedas	70
Figura 4.6	Experimento com Lançamento de Moedas – HMM com Três Moedas	71
Figura 4.7	Bolas em Urnas – HMM com $N = 3$ e $M = 4$	72
Figura 5.1	Modelo SAN para uma HMM Genérica com $ \mathcal{S} = 2$ e $ \mathcal{Y} = 3$	86
Figura 6.1	Máquina de Refrigerantes – Modelagem em MC	89
Figura 6.2	<i>Crazy Soda Machine</i> – Modelagem em HMM	90
Figura 6.3	<i>Crazy Soda Machine</i> – Modelagem em SAN	93
Figura 6.4	<i>Crazy Soda Machine</i> – Cadeia de Markov Equivalente	95
Figura 6.5	<i>Greek Speller</i> – Modelagem em HMM	96
Figura 6.6	<i>Greek Speller</i> – Modelagem em SAN	101
Figura 6.7	<i>Greek Speller</i> – Cadeia de Markov Equivalente (com 20 estados)	103

LISTA DE TABELAS

Tabela 3.1	Eventos Possíveis a Partir de Cada Estado dos Autômatos da Figura 3.2 . . .	43
Tabela 3.2	Conjuntos dos Estados Sucessores dos Autômatos da Figura 3.2	43
Tabela 3.3	Conjunto de Tuplas de Transição do Modelo da Figura 3.2	45
Tabela 3.4	Classificação de Cada Evento do Modelo SAN da Figura 3.2	45
Tabela 3.5	Descritor Markoviano	50
Tabela 4.1	Tabuleiro Numerado	73

LISTA DE SIGLAS

CTA	<i>Classical Tensor Algebra</i>
CTMC	<i>Continuous-Time Markov Chains</i>
DTMC	<i>Discrete-Time Markov Chains</i>
GTA	<i>Generalized Tensor Algebra</i>
HMM	<i>Hidden Markov Models</i>
MC	<i>Markov Chains</i>
SAN	<i>Stochastic Automata Networks</i>

SUMÁRIO

1. INTRODUÇÃO	23
2. CADEIAS DE MARKOV	27
2.1 Processos Markovianos	27
2.2 Fundamentos	28
2.3 Definição Formal	30
2.4 Classificação dos Estados	31
2.5 Irredutibilidade	32
2.6 Distribuição de Probabilidade	33
2.7 Matriz de Probabilidade e o Gerador Infinitesimal	35
2.8 Exemplo: Modelo Markoviano do Tempo	36
3. REDES DE AUTÔMATOS ESTOCÁSTICOS	39
3.1 Fundamentos	39
3.2 Redes de Autômatos Estocásticos	41
3.3 Método de obtenção do Gerador Infinitesimal	48
3.4 Exemplo: Jogo “Cara ou Coroa”	50
3.4.1 Descritor Markoviano e Gerador Infinitesimal	51
3.5 Exemplo: Jogo “Dois ou Um”	53
4. MODELOS MARKOVIANOS OCULTOS	57
4.1 Fundamentos	57
4.2 Definição Formal	58
4.3 Problemas Clássicos	59
4.3.1 Solução do Problema 1	60
4.3.2 Solução do Problema 2	64
4.3.3 Solução do Problema 3	67
4.4 Exemplo: Experimento com Lançamento de Moedas	69
4.5 Exemplo: Bolas em Urnas	71
4.6 Exemplo: Tabuleiro Numerado	72
5. EQUIVALÊNCIA DE REPRESENTAÇÃO	75
5.1 Obtenção do Gerador Infinitesimal para HMM – Primeiro Método	75
5.2 Obtenção do Gerador Infinitesimal para HMM – Segundo Método	77

5.2.1	Método para Obtenção de uma Matriz de Probabilidade de Transição entre Estados	77
5.2.2	Método para Obtenção de uma Matriz de Probabilidade de Transição entre Sinais	78
5.2.3	Método para Obtenção do Gerador Infinitesimal da HMM	80
5.3	Método de Conversão de Modelos HMM em SAN	82
5.3.1	Método para Obtenção do Autômato Correspondente à Cadeia de Markov Oculta	83
5.3.2	Método para Obtenção do Autômato Correspondente à Emissão de Sinais	84
5.3.3	Representação Gráfica do Modelo SAN Obtido a partir de um Modelo HMM	86
5.3.4	Obtenção do Descritor Markoviano	86
6.	EXEMPLOS DE CONVERSÃO DE HMM EM SAN	89
6.1	Exemplo: <i>Crazy Soda Machine</i>	89
6.1.1	Obtenção do Gerador Infinitesimal do Modelo HMM da Figura 6.2	90
6.1.2	Obtenção do Gerador Infinitesimal do Modelo SAN da Figura 6.3	93
6.1.3	<i>Crazy Soda Machine</i> – Representação Gráfica da CTMC Correspondente	95
6.2	Exemplo: <i>Greek Speller</i>	95
6.2.1	Obtenção do Gerador Infinitesimal do Modelo HMM da Figura 6.5	96
6.2.2	Obtenção do Gerador Infinitesimal do Modelo SAN da Figura 6.5	100
6.2.3	<i>Greek Speller</i> – Representação Gráfica da CTMC Correspondente	103
7.	CONSIDERAÇÕES FINAIS E PERSPECTIVAS	105
	Bibliografia	107
A.	ÁLGEBRA TENSORIAL	111
A.1	CTA – Álgebra Tensorial Clássica	111
A.2	GTA – Álgebra Tensorial Generalizada	113

1. INTRODUÇÃO

O estudo formal de sistemas físicos requer a criação de modelos para representá-los. O modelador é, antes de tudo, um atento observador do sistema que deseja modelar; o que ele observa depende de seu conhecimento prévio, de forma que somente aquelas características às quais ele atribui algum significado serão mapeadas ou identificadas. Dentre essas, serão selecionadas as de maior relevância ao fenômeno estudado, do ponto de vista do modelador. Com esse conjunto reduzido de informações, escolhe-se um formalismo para servir de ferramenta de modelagem, sendo este mais um foco de subjetividade, pois a escolha do formalismo também depende do domínio de conhecimento do modelador e inclusive de suas intenções. Ele pode, por exemplo, não estar de fato interessado no sistema em si mas no estudo da ferramenta de modelagem.

Com frequência, pode-se representar o comportamento global de um sistema físico por um conjunto de estados e correspondentes regras de transição. Se a ocupação de estados respeita alguma lei probabilística, então o sistema pode ser formalmente descrito por um *processo estocástico* [30]. Em alguns processos estocásticos a escolha do próximo estado depende exclusivamente do estado atual do sistema, sendo irrelevante a sequência de ocupações prévias. Processos deste tipo são ditos *Markovianos*. Se além de Markoviano o conjunto de estados representativos de um sistema é discreto, então tem-se uma *cadeia de Markov* (*Markov Chain* – MC) [27, 30].

Existem vários formalismos derivados de cadeias de Markov, contudo, dois são de interesse nesta dissertação: Redes de Autômatos Estocásticos (*Stochastic Automata Networks* – SAN) [25] e Modelos Markovianos Ocultos (*Hidden Markov Models* – HMM) [26]. SAN é um formalismo estruturado que se beneficia da álgebra tensorial, enquanto que HMM é um formalismo duplamente estocástico, sem estruturação e formato tensorial definidos. Alguma contextualização faz-se agora necessária.

Considere um sistema cuja dinâmica seria bem representada por uma cadeia de Markov. Esse sistema não precisa ser grande; suponha que ele consista de quarenta estados fortemente conectados, ou seja, cada estado pode ser alcançado a partir de todos os outros, em um número qualquer de transições. Poderia ser usada uma matriz quadrada, com quarenta linhas e colunas, para armazenar as probabilidades¹ de transição associadas a cada par (*linha, coluna*), sendo *linha* o estado de origem e *coluna* o estado destino. Por sua vez, uma representação gráfica para essa cadeia poderia ser obtida ao distribuir quarenta pequenos círculos em uma folha de papel e conectá-los por setas indicando as transições previstas. Teria-se então um *único* autômato de quarenta estados, um pouco confuso, porém compreensível. Mas o que seria dito se, para descrever completamente este sistema, ao invés de quarenta estados fossem necessários apenas onze, distribuídos em *três* autômatos com dois, quatro e cinco estados cada? Sem dúvida a representação gráfica seria mais simples e clara. E se o sistema tivesse seiscentos e vinte e cinco estados? Nesse caso a representação gráfica da cadeia

¹Probabilidades ou taxas, dependendo se a escala de tempo é discreta ou contínua, como será visto no Capítulo 2.

de Markov seria um caos, mas se fosse possível *estruturar* o sistema em mais de um autômato, por exemplo *quatro* autômatos de cinco estados cada, teria-se uma modelagem equivalente porém com representação gráfica não apenas mais clara, mas principalmente viável.

Ao usar SAN como ferramenta para modelar um sistema, o modelador reconhece a possibilidade de identificar subsistemas ou estruturas, os autômatos, que interagem por meio de funções ou transições sincronizadas, de forma a retratar exatamente o mesmo processo. Uma das vantagens do formalismo de Redes de Autômatos Estocásticos está, então, na estruturação do processo, permitindo uma modelagem mais compacta, clara e com representação gráfica simplificada. Outra vantagem está na representação matricial de um modelo SAN, onde a matriz global de transição entre os estados do sistema é substituída por matrizes menores, os *tensores*, das ordens dos autômatos da rede. Todavia, se desejado, é possível reconstituir a matriz global a partir de operações tensoriais bem definidas entre essas matrizes menores. Disso segue que SAN é um formalismo estruturado e dotado de formato tensorial [17, 25].

A relação que se estabelece entre Modelos Markovianos Ocultos e Cadeias de Markov é um pouco diferente. HMM é um formalismo adequado à modelagem de processos estocásticos em dois estágios. Um deles constitui a parte *oculta* do modelo, que consiste em uma cadeia de Markov inacessível ou isolada do observador. O caminhar nessa MC produz uma sequência de sinais, cuja emissão pode ser observada, constituindo a parte *observável* do modelo HMM. A partir de uma sequência de emissões, podem ser obtidas informações sobre a MC oculta, produtora do sinal. Para exemplificar, suponha um sinal de rádio cuja qualidade está comprometida pela ação de interferência ou ruído. O ouvinte escutou algo como “*raise*” ou “*race*” mas não sabe ao certo o que realmente foi dito. Contudo, com base no sinal completo e nas possibilidades de combinação fonética, o ouvinte é capaz de atribuir probabilidades às possibilidades e ponderar sobre o real significado da mensagem. Por essas características, HMM é um formalismo bastante aplicado em reconhecimento de padrões, tal como fala, escrita, sequenciamento de DNA e reconhecimento de proteínas [15, 18, 21, 26].

O propósito desta dissertação é propor uma representação tensorial para Modelos Markovianos Ocultos. A forma escolhida para alcançar esse objetivo passa pelo estudo de como converter um modelo HMM em um modelo SAN: estruturado e cujo formato tensorial é conhecido. A estratégia de conversão consiste na criação de dois autômatos, um correspondendo à cadeia de Markov oculta e outro para representar as emissões do modelo HMM. Esses autômatos se relacionam por transições sincronizadas e dependências funcionais são definidas. Um passo intermediário é necessário para mostrar a equivalência entre as representações SAN e HMM, sendo este passo a obtenção de uma cadeia de Markov global (MC global) capaz de representar o modelo HMM. A igualdade entre as MC globais obtidas a partir de ambos os formalismos SAN e HMM constitui a prova de equivalência.

A motivação do trabalho pode ser entendida com um exemplo ilustrativo [1]. Suponha que se queira usar HMM para fazer um modelo fonético de reconhecimento de fala para a língua inglesa. Ao escolher o *fonema* como unidade de modelagem tem-se uma cadeia de Markov oculta composta de quarenta estados, pois são identificados quarenta fonemas na língua inglesa. Nessa situação, basta pensar na representação gráfica da MC oculta para imaginar os possíveis benefícios de uma

representação estruturada. Entretanto, fonemas sofrem efeito de coarticulação fonética, ou seja, podem ser modificados por influência dos fonemas vizinhos. Para diminuir esse efeito, o modelador poderia escolher, por exemplo, o *difone*, *trifone*, *sílaba* ou até mesmo a *palavra* como unidade de modelagem. No caso do trifone, a sequência combinatória dos fonemas tomados três a três formaria uma cadeia de Markov oculta com $40^3 = 64 \text{ mil}$ potenciais estados. A idéia de uma representação estruturada parece agora bastante atrativa.

Trata-se de um trabalho inicial, com foco no estudo da viabilidade de proporcionar ao formalismo de Modelos Markovianos Ocultos uma representação estruturada e com formato tensorial. Portanto, a apresentação de uma estruturação ótima não faz parte do escopo deste trabalho. Também não são feitas tentativas explícitas de estruturação da cadeia de Markov oculta de uma HMM, sendo isso deixado para trabalhos futuros.

Este volume é composto de sete capítulos e um apêndice. No Capítulo 2 são vistos conceitos relacionados a processos Markovianos, com foco em Cadeias de Markov. O Capítulo 3 é dedicado ao estudo e formalização de Redes de Autômatos Estocásticos (SAN), enquanto que no Capítulo 4 são formalizados conceitos sobre o formalismo de Modelos Markovianos Ocultos (HMM). No Capítulo 5 são apresentados dois métodos para obtenção de um gerador infinitesimal para HMM e um método de conversão de modelos HMM em modelos SAN; exemplos de aplicação desses métodos são vistos no Capítulo 6. O Capítulo 7 é dedicado à conclusão do trabalho, considerações finais e perspectivas. Conceitos básicos de Álgebra Tensorial Clássica e Generalizada são vistos no Apêndice A.

2. CADEIAS DE MARKOV

Neste capítulo, inteiramente baseado no trabalho de Stewart [30], são tratados aspectos fundamentais relativos a cadeias de Markov, um formalismo para modelagem de processos estocásticos com espaço de estados discreto e que atendem à propriedade Markoviana. Os primeiros resultados na área datam de 1906, graças aos trabalhos do matemático russo Andrey Markov.

A exposição é organizada tal que na Seção 2.1 são informalmente apresentadas algumas noções sobre processos estocásticos e Markovianos; na Seção 2.2 são encontrados aspectos mais formais relativos a processos estocásticos com espaço de estados discreto; a definição formal de cadeias de Markov é vista na Seção 2.3; definições e propriedades relativas à classificação dos estados e irreduzibilidade de uma cadeia de Markov são vistas nas Seções 2.4 e 2.5; a Seção 2.6 é dedicada à solução de modelos em cadeias de Markov, enquanto que a Seção 2.7 trata das diferenças e peculiaridades da representação de cadeias de Markov em escala de tempo discreta com relação à representação em escala de tempo contínua. A Seção 2.8 mostra um exemplo de modelagem.

2.1 Processos Markovianos

Frequentemente, o comportamento de um sistema físico pode ser representado por um conjunto de estados associado a uma descrição de como o sistema se move de um estado a outro com o passar do tempo [30]. O conjunto de todos os estados ocupados pelo sistema recebe o nome de *espaço de estados*, podendo ser contínuo ou discreto, finito ou infinito. Se a ocupação dos estados respeita alguma lei probabilística, então o sistema pode ser formalmente descrito por um *processo estocástico*.

Um *processo estocástico* é definido pelo seu *espaço de estados* – $S = \{s_i\}, i \in \mathbb{N}$ – e por uma *variável estocástica* – X – indexada por um *parâmetro* – $t \in T \subset \mathbb{R}$.¹ O parâmetro é geralmente imaginado como o intervalo temporal de observação do sistema. Então, se o conjunto T for discreto, tem-se um processo estocástico em escala de *tempo discreto*; se for contínuo, o processo estocástico é de *tempo contínuo*. A variável estocástica, definida em um espaço de probabilidades, assume valores no espaço de estados em diferentes instantes de observação. Assim, $X_t = s_i$ significa que o estado s_i foi observado no instante t como resultado de um processo estocástico em X .

Em um processo estocástico, assume-se que as transições entre estados são instantâneas e que um único estado é visitado por vez. São exemplos de processos estocásticos em espaço de estados contínuo a variação da umidade relativa do ar e o nível de glicose no sangue; um exemplo de espaço de estados discreto é o número de pessoas em uma fila, esperando atendimento.

Por fim, um *processo Markoviano* é um processo estocástico que exhibe a chamada *propriedade Markoviana*, que estabelece que a evolução futura do sistema depende somente do estado em que ele se encontra atualmente, sendo irrelevante a sua história pregressa. Processos Markovianos cujo

¹ \mathbb{N} e \mathbb{R} são os conjuntos dos números *Naturais* e *Reais*, respectivamente.

o espaço de estados é *discreto* recebem o nome de *cadeias de Markov* [30]. A seção seguinte aborda aspectos relativos a cadeias de Markov, tais como classificações, propriedades e tipos de soluções.

2.2 Fundamentos

De acordo com Stewart [30], cadeias de Markov (*Markov Chains* – MC) possuem espaço de estados discreto à escala de tempo discreta ou contínua. Já Reichl [27] restringe a definição de cadeias de Markov como um processo Markoviano com espaço de estados discreto à escala de tempo também discreta. Neste trabalho, adota-se a definição apresentada por Stewart, visto que, dentre estas, é a mais difundida no meio acadêmico da ciência da computação.

Desde já fique claro que o capítulo tem foco no estudo de cadeias de Markov com espaço de estados *finito*. Contudo, por vezes os conceitos são apresentados de forma não tão restritiva, englobando cadeias de Markov com espaço de estados *infinito*. Isso é feito para melhor contextualizar algumas definições e propriedades de interesse, servindo de embasamento e argumentação em busca de uma melhor definição de escopo.

Na Seção 2.1, alguns conceitos fundamentais ao desenvolvimento da compreensão sobre cadeias de Markov foram informalmente apresentados. Recapitulando, tem-se as seguintes definições:

Definição 2.2.1. Propriedade Markoviana: *a evolução futura do sistema depende exclusivamente do seu estado atual, sendo irrelevante a sua história pregressa.*

Definição 2.2.2. Processo Markoviano: *processo estocástico que atende à propriedade Markoviana.*

Definição 2.2.3. Cadeia de Markov: *processo Markoviano com espaço de estados discreto.*

Mais adiante será visto um método para conversão de cadeias de Markov à escala de tempo discreta (*Discrete-Time Markov Chains* – DTMC) em cadeias de Markov à escala de tempo contínua (*Continuous-Time Markov Chains* – CTMC). Em vista desse método, que consiste em um par de operações triviais, optou-se por apresentar os conceitos e definições formais tendo como base uma DTMC. Salienta-se que essa decisão não implica em alguma limitação ou restrição dos conceitos apresentados, sendo apenas uma opção visando uma representação mais simples, direta e compreensível.

Sejam

$\mathcal{S} = \{s_i\}$	conjunto discreto de estados;
$ \mathcal{S} $	cardinalidade de \mathcal{S} ;
X	variável estocástica a assumir valores do espaço de estados em diferentes instantes de tempo;
$P(X_t = s_i)$	probabilidade de observação do estado s_i no instante t .

Considere um processo estocástico e a sequência de observações $\{X_1 X_2 \cdots X_t \cdots X_T\}$, definida para um espaço de estados discreto, onde $1 \leq t \leq T$ representa uma ordenação também discreta no tempo. De acordo com Jelinek [21], se nada mais é dito, a probabilidade conjunta dessas variáveis estocásticas é dada pela fórmula de Bayes:

$$\begin{aligned} P(X_1 X_2 \cdots X_T) &= \prod_{t=1}^T P(X_t | X_1 X_2 \cdots X_{t-1}) \\ &= P(X_1)P(X_2|X_1)P(X_3|X_1 X_2) \cdots P(X_T|X_1 X_2 X_3 \cdots X_{T-1}), \end{aligned} \quad (2.1)$$

onde $P(X_t | X_1 X_2 \cdots X_{t-1})$ é a *probabilidade condicional* de se observar a variável estocástica X no tempo t , dada a sequência de observações anteriores, $\{X_1 X_2 \cdots X_{t-1}\}$. Assim, $P(X_2|X_1)$ é a probabilidade de observação de X_2 visto que X_1 foi observado. Uma notação alternativa, porém mais completa, seria $P(X_2 = s_j | X_1 = s_i)$: probabilidade condicional de se observar o estado s_j no instante $t = 2$, dado que s_i foi observado em $t = 1$.

Um processo estocástico, tal como descrito pela Equação 2.1, é dito Markoviano se satisfaz à propriedade Markoviana (Definição 2.2.1):

$$P(X_t | X_1 X_2 X_3 \cdots X_{t-1}) = P(X_t | X_{t-1}) \quad (2.2)$$

Reescrevendo a Equação 2.1, tem-se:

$$\begin{aligned} P(X_1 X_2 \cdots X_T) &= \prod_{t=1}^T P(X_t | X_{t-1}) \\ &= P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3) \cdots P(X_T|X_{T-1}) \end{aligned} \quad (2.3)$$

Seja a Equação 2.4 para a *probabilidade conjunta* das variáveis estocásticas $\{X_1, X_2, X_3\}$ de um processo Markoviano. Aplicando o somatório sobre X_2 obtém-se a probabilidade conjunta de $\{X_1, X_3\}$, dada pela Equação 2.5. Ao dividir esta equação por $P(X_1)$, obtém-se a probabilidade condicional, $P(X_3|X_1)$, de uma determinada observação no instante $t = 3$ dada a observação ocorrida no instante $t = 1$, sendo indiferente a observação feita no instante $t = 2$. A Equação 2.6 é conhecida como *equação de Chapman-Kolmogorov*.

$$P(X_1 X_2 X_3) = P(X_1)P(X_2|X_1)P(X_3|X_2) \quad (2.4)$$

$$P(X_1 X_3) = P(X_1) \sum_{X_2} P(X_2|X_1)P(X_3|X_2) \quad (2.5)$$

$$P(X_3|X_1) = \sum_{X_2} P(X_2|X_1)P(X_3|X_2). \quad (2.6)$$

A equação de Chapman-Kolmogorov evidencia a propriedade Markoviana ao mostrar que passos sucessivos são estatisticamente independentes, como nota Reichl [27]. Ou seja, na Equação 2.6 a

probabilidade de transição entre X_2 e X_3 não é afetada pela transição anterior, de X_1 para X_2 . Essa equação pode ser estendida para um número arbitrário de passos. A Chapman-Kolmogorov para $(t' - t)$ passos é dada pela Equação 2.7:

$$\begin{aligned} P(X_{t'} = s_j | X_t = s_i) &= \\ &= \sum_{\forall \neq \{X_t, X_{t'}\}} P(X_{t+1} | X_t = s_i) \cdots P(X_{\tau_n} | X_{\tau_{n-1}}) \cdots P(X_{t'} = s_j | X_{t'-1}) \end{aligned} \quad (2.7)$$

2.3 Definição Formal

Definição 2.3.1. *Uma cadeia de Markov é definida por:*

$$\begin{aligned} \mathcal{S} = \{s_i\} & \quad \text{conjunto discreto de estados;} \\ \pi = \{\pi_i\} & \quad \text{distribuição de probabilidade inicial, onde } \pi_i = P(X_1 = s_i) \text{ para } s_i \in \mathcal{S}; \\ A(t) = \{a_{ij}(t)\} & \quad \text{distribuição de probabilidade de transição, onde} \\ & \quad a_{ij}(t) = P(X_{t+1} = s_j | X_t = s_i) \text{ para } t \in T \text{ e } s_i, s_j \in \mathcal{S}. \end{aligned}$$

Por tratar-se de um processo estocástico, os elementos de uma DTMC devem respeitar algumas propriedades probabilísticas.

Propriedade 2.3.1. *O processo inicia invariavelmente em algum estado $s_i \in \mathcal{S}$. Disso decorre resultar na unidade o somatório sobre os elementos de π :*

$$\sum_{i=1}^{|\mathcal{S}|} \pi_i = 1, \quad \pi_i \geq 0, \quad (2.8)$$

Propriedade 2.3.2. *O espaço de estados é fechado, significando que apenas estados $s_i \in \mathcal{S}$ são visitados. Assim, a transição partindo de um estado s_i deve levar a qualquer estado $s_j \in \mathcal{S}$:*

$$\sum_{j=1}^{|\mathcal{S}|} a_{ij}(t) = 1, \quad a_{ij}(t) \geq 0 \quad (2.9)$$

$$\sum_{j=1}^{|\mathcal{S}|} P(X_t = s_j | X_{t-1} = s_i) = 1 \quad (2.10)$$

Definição 2.3.2. *Processo Markoviano Homogêneo: aquela cuja a probabilidade condicional de transição não varia com o tempo:*

$$\begin{aligned} a_{ij}(t) &= P(X_{t+1} = s_j | X_t = s_i) \\ a_{ij}(t=1) &= P(X_2 = s_j | X_1 = s_i) \\ a_{ij}(t) &= a_{ij}(t=1) = a_{ij} = P(s_j | s_i), \quad \forall t \in T, \quad \forall i, j \in \mathcal{S} \end{aligned} \quad (2.11)$$

Se o processo é homogêneo, sua matriz de transição A também é. Neste caso, a Equação 2.7 pode ser reescrita da seguinte forma:

$$P(X_{t'} = s_j | X_t = s_i) = \left(A^{(t'-t)} \right)_{ij} \quad (2.12)$$

onde $A^{(t'-t)}$ é a matriz A elevada na potência $(t' - t)$, que representa o número de passos desejado. Por exemplo, se $(t' - t) = 5$, os elementos da matriz resultante da potenciação A^5 indicam a probabilidade de sair de um estado e chegar a outro em *cinco* passos.

2.4 Classificação dos Estados

Os estados em uma cadeia de Markov podem ser individualmente classificados quanto a sua recorrência. Um estado é dito *recorrente* se é garantido que ele seja revisitado no decorrer do processo. Um estado é *transiente* se há probabilidade não-nula de que ele nunca mais seja visitado. Isso não significa, porém, que um estado transiente não possa ser visitado muitas vezes, significa apenas que existe uma chance do processo não retornar a este estado. Observe o esquema que segue:

$$\text{Estado} \left\{ \begin{array}{l} \text{Recorrente} \left\{ \begin{array}{ll} \text{Positivo-Recorrente:} & \text{recorrência média } \textit{finita} \\ \text{Nulo-Recorrente:} & \text{recorrência média } \textit{infinita} \end{array} \right. \\ \\ \text{Transiente:} & \text{probabilidade não-nula de nunca retornar ao estado} \end{array} \right.$$

Para uma DTMC homogênea, vale a Equação 2.12, de Chapman-Kolmogorov. Seja a seguinte forma alternativa de escrever essa equação:

$$a_{ij}^{(n)} = \sum_{\forall k} a_{ik}^{(n-l)} a_{kj}^{(l)} \quad (2.13)$$

onde $a_{ij}^{(n)}$ é o elemento na posição (i, j) da matriz A elevada à potência n , que representa o número de passos. A Equação 2.13 pode ser usada para calcular a probabilidade de retorno ao estado após n passos, $a_{ii}^{(n)}$, bastando que se troque todas as ocorrências de j por i . Contudo, esse cálculo não exclui passagens por i nos passos intermediários.

Seja

$$f_{ii}^{(n)} \quad \text{probabilidade de deixar o estado } s_i \text{ e retornar a ele pela primeira vez em exatamente } n \text{ passos.}$$

A probabilidade do primeiro retorno a s_i em n passos pode ser calculada recursivamente pela Equação 2.14:

$$\begin{aligned} a_{ii}^{(0)} &= 1 \\ a_{ii}^{(1)} &= a_{ii} = f_{ii}^{(1)} \\ a_{ii}^{(n)} &= \sum_{l=1}^n a_{ii}^{(n-l)} f_{ii}^{(l)}, \quad n \geq 1 \end{aligned} \quad (2.14)$$

Enquanto que a probabilidade do processo retornar ao estado s_i em algum momento futuro é dada por:

$$f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)}. \quad (2.15)$$

Se $f_{ii} = 1$, então o estado é *recorrente*, significando que ele é garantidamente revisitado. Se $f_{ii} < 1$ então o estado é dito *transiente*, existindo probabilidade não-nula do estado nunca ser revisitado.

Quando um estado é recorrente, $f_{ii} = 1$, então é possível calcular seu *tempo médio de recorrência*, M_{ii} , dado pela equação:

$$M_{ii} = \sum_{n=1}^{\infty} n a_{ii}^{(n)}. \quad (2.16)$$

Se M_{ii} é finito, então o estado é dito *positivo-recorrente*. Se $M_{ii} = \infty$, então o estado s_i é dito *nulo-recorrente*, significando que seguramente o estado é revisitado, mas que isso acontece no limite do tempo tendendo a infinito. De acordo com Stewart [30], em uma cadeia de Markov com espaço de estados finito nenhum dos seus estados pode ser nulo-recorrente.

Propriedade 2.4.1. *Uma cadeia de Markov finita possui pelo menos um estado positivo-recorrente, sendo cada um dos estados restantes classificado como positivo-recorrente ou transiente.*

Os estados de uma cadeia de Markov também podem ser classificados quanto à periodicidade. Um estado s_i é dito *periódico com período p* se, ao deixá-lo, o retorno a s_i só é possível em um número de transições que seja múltiplo de $p > 1$. Nas palavras de Stewart [30], o período do estado s_i é definido como o maior divisor comum do conjunto de inteiros n para os quais $a_{ii}^{(n)} > 0$. Um estado cujo período é $p = 1$ é dito *aperiódico*.

Definição 2.4.1. Estado Ergódico: *todo estado classificado como positivo-recorrente e aperiódico.*

Definição 2.4.2. Cadeia de Markov Ergódica: *aquela onde todos os estados $s_i \in \mathcal{S}$ são ergódicos.*

2.5 Irredutibilidade

Uma DTMC é dita *irredutível* quando cada estado pode ser alcançado a partir de todos os outros restantes, *i.e.*, para qualquer par de estados $s_i, s_j \in \mathcal{S}$ deve existir uma sequência de transições em

cada sentido, tal que $a_{ij}^{(m)} > 0$ e $a_{ji}^{(n)} > 0$, sendo m e n inteiros.

Para verificar a redutibilidade de uma cadeia de Markov inicia-se particionando o espaço de estados \mathcal{S} em dois subconjuntos disjuntos \mathcal{S}_1 e \mathcal{S}_2 . O subconjunto $\mathcal{S}_1 \neq \{\emptyset\}$ é dito *fechado* se não houver transição que leve de $s_i \in \mathcal{S}_1$ a um estado $s_j \notin \mathcal{S}_1$. Assim, se existir um subconjunto fechado $\mathcal{S}_1 \subset \mathcal{S}$, então a DTMC é *redutível*. Por outro lado, se não for possível encontrar um subconjunto fechado $\mathcal{S}_1 \subset \mathcal{S}$, a DTMC é dita *irredutível*. Para o caso especial de $\mathcal{S}_1 = \{s_i\}$ ser fechado, s_i é identificado como estado *absorvente*.

A seguir são apresentados dois importantes teoremas sobre irredutibilidade, sem suas correspondentes provas formais.

Teorema 2.5.1. *Se uma cadeia de Markov é irredutível, então os seus estados são:*

- todos positivo-recorrentes, ou
- todos nulo-recorrentes, ou
- todos transientes.

Além disso, todos os estados são periódicos com o mesmo período, p , ou então são todos aperiódicos.

O Teorema 2.5.2, a seguir, é consequência direta da associação entre o Teorema 2.5.1 e a Propriedade 2.4.1:

Teorema 2.5.2. *Se uma cadeia de Markov é irredutível, aperiódica e finita, então todos os seus estados são ergódicos.*

Assim, dados os Teoremas 2.5.1 e 2.5.2, relativos à irredutibilidade, aliados à Propriedade 2.4.1, conclui-se que toda a cadeia de Markov finita e irredutível é composta somente de estados positivo-recorrentes. Neste caso, todos os seus estados são periódicos com o mesmo período p ou então são todos aperiódicos. Dessa forma, basta que se mostre que *um* desses estados é aperiódico para provar que a cadeia de Markov é *ergódica*. Em suma, se a DTMC é finita e irredutível, então seus estados são todos positivo-recorrentes. Se além disso forem aperiódicos, então a DTMC é ergódica.

2.6 Distribuição de Probabilidade

Frequentemente tem-se interesse em determinar a probabilidade de uma DTMC ocupar um determinado estado em um dado instante de tempo. Seja $\psi_i(n)$ a notação para indicar a probabilidade que a cadeia de Markov esteja no estado s_i no instante $t = n$:

$$\psi_i(n) = P(X_n = s_i) \quad (2.17)$$

Em notação vetorial, tem-se o vetor linha $\psi(n) = (\psi_1(n), \psi_2(n), \dots, \psi_i(n), \dots, \psi_{|\mathcal{S}|}(n))$. A distribuição de probabilidade sobre os estados de \mathcal{S} em um dado passo n na evolução do processo pode

ser obtida a partir da distribuição de probabilidade inicial, $\pi = \{\pi_i\}$, e da matriz de probabilidade de transição, $A = \{a_{ij}\}$:

$$\psi_j(n) = \sum_{\forall i} \psi_i(1) P(X_n = s_j | X_1 = s_i) \quad (2.18)$$

$$\psi_j(n) = \sum_{\forall i} \pi_i a_{ij}^{(n-1)} \quad (2.19)$$

$$\psi(n) = \psi(1)A^{(n-1)} = \pi A^{n-1} \quad (2.20)$$

Definição 2.6.1. *Seja A a matriz de probabilidade de transição de uma DTMC, e seja $z = \{z_i\}$ um vetor de distribuição de probabilidade, i.e.,*

$$z_i \in \mathbb{R}, 0 \leq z_i \leq 1 \quad \text{e} \quad \sum_{\forall i} z_i = 1. \quad (2.21)$$

O vetor z é chamado de distribuição estacionária se e somente se $zA = z$.

Nas palavras de Stewart [30], se z é escolhido como distribuição de probabilidade inicial, i.e., $\psi_i(1) = z_i$ para todo s_i , então para qualquer n tem-se $\psi_i(n) = z_i$.

Definição 2.6.2. *Seja dada uma distribuição de probabilidade inicial $\psi(1)$. Se existe o limite de $\psi(n)$ para n tendendo a infinito, então este limite é chamado distribuição limite e é escrito*

$$\psi = \lim_{n \rightarrow \infty} \psi(n). \quad (2.22)$$

De acordo com Stewart [30], o fato de uma cadeia de Markov possuir distribuição limite não implica na existência de distribuição estacionária, e vice-versa. Uma cadeia de Markov *irredutível* se comportará da seguinte forma:

- se *aperiódica*, possui distribuição limite;
- se *positiva-recorrente*, possui distribuição estacionária única.

Pelas Definições 2.4.1 e 2.4.2, se os estados de uma cadeia de Markov são todos positivo-recorrentes e aperiódicos, então ela é ergódica. Assim, conclui-se que cadeias de Markov irredutíveis e ergódicas possuem distribuição limite e também distribuição estacionária, sendo uma igual à outra. Isto é o mesmo que dizer que a distribuição de probabilidade $\psi(n)$ sempre converge para uma solução limite, ψ , independentemente da distribuição de probabilidade inicial. Ao atingir a distribuição limite, diz-se que o processo alcançou *equilíbrio*, e essa solução de equilíbrio é facilmente calculada por:

$$\psi = \psi A, \quad (2.23)$$

tal qual visto na Definição 2.6.1. A seguir, são apresentadas algumas propriedades de cadeias de Markov ergódicas.

Propriedade 2.6.1. *Sejam $\psi(n)$ um processo ergódico e ψ o vetor solução para esse processo no equilíbrio, e seja $A^{(n)} = A^n$. No limite para n tendendo a infinito, todas as linhas da matriz A^∞ são idênticas a ψ .*

Propriedade 2.6.2. *Sejam $\psi(n)$ um processo ergódico e ψ o vetor solução para esse processo no equilíbrio. O tempo médio que a cadeia passa em cada estado $s_i \in \mathcal{S}$ é proporcional a ψ_i , a posição correspondente no vetor.*

Propriedade 2.6.3. *Sejam $\psi(n)$ um processo ergódico e ψ o vetor solução para esse processo no equilíbrio. A razão ψ_i/ψ_j resulta no número de vezes que o estado s_i é visitado entre duas visitas ao estado s_j .*

2.7 Matriz de Probabilidade e o Gerador Infinitesimal

Enquanto que uma DTMC é representada por uma matriz de *probabilidade de transição*, uma CTMC é representada por uma matriz de *taxa de transição*, também conhecida pelo nome de *gerador infinitesimal*. De acordo com Stewart [30], probabilidade de transição é uma quantidade que depende do intervalo de tempo da observação. Assim, se este intervalo é muito pequeno, também será muito pequena a probabilidade de transição, ao passo que se o intervalo de observação é muito grande, não se pode ignorar a probabilidade da ocorrência de múltiplas transições no mesmo intervalo. Por outro lado, uma matriz de *taxa de transição* não depende de um intervalo de observação, visto que seus elementos são definidos como o limite da probabilidade de transição para quando o intervalo de observação tende a zero:

$$Q_{ij}(t) = \lim_{\Delta t \rightarrow 0} \left(\frac{a_{ij}(t, t + \Delta t)}{\Delta t} \right), \quad \forall i \neq j; \quad (2.24)$$

$$Q_{ii}(t) = - \sum_{i \neq j} Q_{ij}(t). \quad (2.25)$$

E em notação matricial, tem-se:

$$Q(t) = \lim_{\Delta t \rightarrow 0} \left(\frac{A(t, t + \Delta t) - I}{\Delta t} \right). \quad (2.26)$$

A dedução da expressão para o cálculo do gerador infinitesimal, Equação 2.26, pode ser encontrada no livro de Stewart [30]. Apesar de optar-se por não apresentar esses cálculos, parece importante notar que probabilidades são sempre valores positivos, enquanto que taxas podem ser negativas, como é o caso de $Q_{ii}(t)$. Como explicado por Stewart, uma taxa de transição é definida como uma derivada e, como tal, pode ser positiva ou negativa. O valor negativo de $Q_{ii}(t)$ apenas indica que a probabilidade de permanência no estado s_i diminui com o tempo, enquanto que a probabilidade de ocorrência de uma transição de s_i para s_j aumenta; por isso $Q_{ij}(t)$ é positivo.

Como consequência, a matriz de probabilidade de transição, A , de uma DTMC possui apenas valores positivos, tal que a soma sobre os valores de cada linha resulta em "1". Já a matriz de

taxa de transição, Q , de uma CTMC tem sua diagonal toda preenchida por elementos negativos, de forma que a soma sobre os valores de cada linha resulta em “0”.

A solução de equilíbrio de uma DTMC é dada pela resolução da Equação 2.27, enquanto que a solução de equilíbrio de uma CTMC é obtida pela resolução da Equação 2.28:

$$\psi A = \psi \quad (2.27)$$

$$\psi Q = \underline{0}, \quad (2.28)$$

onde $\underline{0}$ é o vetor de zeros da mesma dimensão que ψ . Como dito anteriormente, um modelo DTMC pode ser convertido em um modelo CTMC e vice-versa. Segue então, que o vetor de solução ψ para o processo no equilíbrio é um único, independentemente de ter usado a Equação 2.27 ou a Equação 2.28 para resolver o modelo.

Para converter uma DTMC em uma CTMC correspondente é suficiente subtrair de A a identidade:

$$Q = A - I \quad (2.29)$$

Por sua vez, para converter uma CTMC em uma DTMC correspondente, primeiro, todos os elementos de Q devem ser divididos por $(k + \epsilon)$, onde k é o maior elemento em módulo e ϵ é qualquer valor não negativo conveniente. Isto é útil para quando $k < 1$, então Q pode ser dividido por $(k + \epsilon) = 1$, i.e., simplesmente não ser dividido. Depois, basta somar a identidade à matriz resultante para obter uma DTMC correspondente:

$$A = \frac{1}{(k + \epsilon)} Q + I \quad (2.30)$$

2.8 Exemplo: Modelo Markoviano do Tempo

Este exemplo é apresentado por Rabiner [26] como forma de ilustrar a aplicação de cadeias de Markov. Trata-se de uma modelagem do tempo para observações diárias. Assim, seja a variável estocástica X , que representa o tempo e tem suas realizações definidas no conjunto discreto $\{s_1 = \text{chuvoso}, s_2 = \text{nublado}, s_3 = \text{ensolarado}\}$. Determina-se que as observações são feitas uma vez ao dia, que o resultado obtido será sempre um desses três estados, não sendo possível a combinação entre eles, e que as probabilidades de transição entre os estados são dadas pela matriz:

$$A = \{a_{ij}\} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}$$

Dado que o tempo no dia 1 é *ensolarado* ($X_1 = s_3$), qual a probabilidade (de acordo com o modelo) de que o tempo nos próximos 7 dias seja “ensolarado-ensolarado-chuvoso-chuvoso-ensolarado-

nublado-ensolarado"? Mais formalmente, define-se a sequência de observação $O = \{X_1 = s_3, X_2 = s_3, X_3 = s_3, X_4 = s_1, X_5 = s_1, X_6 = s_3, X_7 = s_2, X_8 = s_3\}$. Deseja-se obter a probabilidade de O , dado o modelo:

$$\begin{aligned}
 P(O|Modelo) &= P(s_3, s_3, s_3, s_1, s_1, s_3, s_2, s_3|Modelo) \\
 &= P(s_3)P(s_3|s_3)P(s_3|s_3)P(s_1|s_3)P(s_1|s_1)P(s_3|s_1)P(s_2|s_3)P(s_3|s_2) \\
 &= \pi_3 \cdot a_{33} \cdot a_{33} \cdot a_{31} \cdot a_{11} \cdot a_{13} \cdot a_{32} \cdot a_{23} \\
 &= 1 \cdot (0.8)(0.8)(0.1)(0.4)(0.3)(0.1)(0.2) \\
 &= 1.536 \times 10^{-4}
 \end{aligned}$$

onde a notação

$$\pi_i = P(X_1 = s_i), \quad 1 \leq i \leq |\mathcal{S}| \quad (2.31)$$

é usada para indicar a probabilidade inicial de cada estado.

Outra questão de interesse é: dado que o modelo está em um estado conhecido, qual a probabilidade dele permanecer nesse estado por exatamente d dias? Essa probabilidade pode ser avaliada como sendo a probabilidade da sequência de observação

$$O = \{s_i, s_i, s_i, \dots, s_i, s_j \neq s_i\}$$

$\begin{matrix} 1 & 2 & 3 & & d & d+1 \end{matrix}$

dado o modelo, a qual é

$$P(O|Model, X_1 = s_i) = (a_{ii})^{d-1}(1 - a_{ii}) = p_i(d). \quad (2.32)$$

A quantidade $p_i(d)$ é a densidade de probabilidade de duração d no estado s_i . Essa duração exponencial é característica do tempo de permanência em um estado numa cadeia de Markov. A partir de $p_i(d)$ pode-se calcular o número médio de observações subsequentes em um dado estado, condicionado ao fato desse ter sido o estado inicial

$$\bar{d}_i = \sum_{d=1}^{\infty} d \cdot (p_i(d)) \quad (2.33)$$

$$= \sum_{d=1}^{\infty} d \cdot ((a_{ii})^{d-1}(1 - a_{ii})) = \frac{1}{1 - a_{ii}} \quad (2.34)$$

Assim, de acordo com o modelo, o número esperado de dias ensolarados consecutivos é $1/(0.2) = 5$ (o atual, mais 4 dias de sol), de dias nublados é 2.5 e de dias chuvosos é 1.67.

3. REDES DE AUTÔMATOS ESTOCÁSTICOS

Este capítulo é dedicado ao estudo de Redes de Autômatos Estocásticos em escala de tempo contínua. O formalismo foi desenvolvido na década de 80, por Plateau [25], e recebe contribuições até os dias de hoje. É particularmente adequado à modelagem de processos Markovianos que possam ser estruturados em subsistemas que apresentem baixo grau de acoplamento entre si.

O capítulo tem como base os trabalhos de Sales [28, 29] e é organizado como segue. Na Seção 3.1 são informalmente introduzidos conceitos fundamentais; a formalização de Redes de Autômatos Estocásticos é feita na Seção 3.2; a Seção 3.3 apresenta os passos para a obtenção do *gerador infinitesimal* de um modelo; as Seções 3.4 e 3.5 são dedicadas ao estudo de exemplos de modelagem usando o formalismo.

3.1 Fundamentos

Redes de Autômatos Estocásticos (*Stochastic Automata Networks – SAN*) tem na tese de Plateau [25] seu trabalho seminal. Trata-se de uma forma alternativa para modelagem de processos Markovianos com espaço de estados discreto. Qualquer processo caracterizado como uma cadeia de Markov pode ser modelado em SAN, seja a cadeia finita ou infinita, redutível ou irredutível, periódica ou aperiódica, homogênea ou não, ergódica ou não, em tempo contínuo ou discreto. O trabalho de Plateau introduz Redes de Autômatos Estocásticos em *tempo contínuo*, que é o foco deste capítulo. O formalismo de SAN em tempo discreto não será apresentado, sendo o leitor convidado a consultar os trabalhos de Sales [29] e Brenner [10], se assim desejar.

SAN em escala de tempo contínua adequa-se à modelagem de sistemas cujo comportamento Markoviano é descrito nesta mesma escala de tempo. Contudo, é sempre possível modificar a escala de tempo de uma cadeia de Markov, como visto na Seção 2.7 (Página 36). Dessa forma, pode-se converter qualquer DTMC¹ em uma CTMC² correspondente, viabilizando a modelagem em SAN na escala de tempo contínua.

Dado um sistema, suponha então que o modelador escolha representá-lo por uma rede de autômatos estocásticos ao invés de uma CTMC. Sua primeira tarefa consiste em identificar os autômatos e, para isso, o modelador seguirá critérios por vezes não evidentes e até mesmo subjetivos. No contexto de SAN, autômatos são estruturas representando subsistemas que, dado o funcionamento conjunto, substituem o sistema original sem qualquer prejuízo, visto a equivalência de representação entre SAN e CTMC [25]. Usualmente, busca-se definir os autômatos de forma a minimizar a dependência entre eles.

Segue uma descrição do formalismo SAN que, embora simplória, pode colaborar na construção dos seus conceitos mais gerais. Cada autômato do modelo SAN consiste em um conjunto de

¹Cadeia de Markov em Tempo Discreto (*Discrete-Time Markov Chain – DTMC*)

²Cadeia de Markov em Tempo Contínuo (*Continuous-Time Markov Chain – CTMC*)

estados com regras de transição entre eles, sendo que essas regras podem ou não depender dos outros autômatos da rede. Além disso, uma transição em um autômato pode acontecer sozinha ou acompanhada de transições simultâneas em outros autômatos da rede. A forma de gerenciar essas transições se dá através de eventos. Por sua vez, cada autômato, composto de estados, transições e correspondentes regras, é representado matematicamente por matrizes, conhecidas como *tensores*³. Por fim, a solução do modelo é obtida ao efetuar operações entre esses tensores. Disso segue que SAN é um formalismo Markoviano estruturado, orientado a eventos e dotado de formato tensorial. Esse conceito será revisitado mais adiante.

Como nota Sales [28], uma cadeia de Markov pode ser vista como uma *máquina de estados*, cujos nodos e arcos representam, respectivamente, os estados e as transições definidas na MC⁴. Em outras palavras, uma cadeia de Markov pode ser representada graficamente por um único autômato. Isto posto, parece natural indagar se existe alguma vantagem na representação de um sistema em SAN. Fernandes *et al.* [17] explicam que são muitas as situações em que SAN é uma escolha apropriada. Cita-se o exemplo de sistemas distribuídos e paralelos, onde os componentes trabalham de forma mais ou menos independente, ocasionalmente sincronizando ações ou operando em diferentes taxas em função do estado de outros componentes do sistema. Ao escolher a representação em SAN, cada componente pode ser modelado como um autômato que interage com os outros da rede. Se a dependência entre os autômatos é mínima, SAN é uma excelente escolha.

No geral, SAN consegue mitigar o problema da *explosão do espaço de estados* que acontece em modelos Markovianos muito grandes, onde a matriz de transição⁵ é tão grande que não há espaço em memória para armazená-la. Em SAN esse problema é abrandado pois o sistema é representado por um *descriptor Markoviano* equivalente à matriz de transição, porém menos oneroso. O descriptor Markoviano consiste em um conjunto de pequenas matrizes, os *tensores*, que guardam informação sobre o comportamento de cada autômato estocástico e a relação entre eles no modelo SAN. Por serem pequenas, no conjunto essas matrizes geralmente requerem menos espaço de armazenamento do que seria necessário para armazenar a matriz de transição correspondente. Como consequência, SAN é capaz de resolver modelos maiores.

Note que, se acaso acontecer de ser criado um modelo muito complexo de um sistema, tal que sejam necessários muitos tensores para representá-lo, a ponto de superar o espaço requerido em memória para o armazenamento da matriz de transição correspondente, ainda assim SAN pode ser uma representação vantajosa, pois existe a possibilidade de distribuir esses tensores entre vários computadores e paralelizar a solução. Em suma, um modelo SAN costuma ocupar menos memória

³ *Tensor* é uma entidade matemática criada para generalizar a idéia de vetor para um espaço n-dimensional. Como consequência dessa generalização, um escalar é classificado como um *tensor de ordem zero*, um vetor é um *tensor de ordem um* e uma matriz bidimensional é um *tensor de ordem dois*. Foi escolhido usar o termo *tensor* desde já para enfatizar que SAN é um formalismo que se beneficia da álgebra tensorial.

⁴ Cadeia de Markov (*Markov Chains* – MC)

⁵ No Capítulo 2 são definidos dois tipos de matrizes de transição: a matriz de *probabilidade* de transição de uma DTMC e a matriz de *taxa* de transição de uma CTMC, também conhecida como *gerador infinitesimal*. Note que SAN modela sistemas classificados como CTMC, portanto, neste capítulo tem-se interesse em matrizes de *taxa* de transição.

que um modelo em MC, mas mesmo que o contrário aconteça, ainda assim é possível paralelizar a solução, dando conta de sistemas com espaço de estados maiores. No que diz respeito à eficiência, isso depende do algoritmo usado para resolução do modelo SAN. Contudo, esse não é o foco deste trabalho, sendo recomendado ao leitor os trabalhos de Webber [31] e Czekster [13], onde esse tópico é extensamente tratado.

Desde o início, o formalismo de Redes de Autômatos Estocásticos conta com a *Álgebra Tensorial Clássica* (*Classical Tensor Algebra – CTA*), a qual possui recursos que possibilitam trabalhar com modelos cujas taxas e probabilidades de roteamento são constantes. Visando tornar a ferramenta de modelagem mais versátil, Fernandes *et al.* [17] estenderam a álgebra tensorial para incorporar o tratamento de funções. Assim, as taxas e probabilidades de roteamento poderiam ter seus valores determinados por funções dependentes dos estados de outros autômatos. Essa nova álgebra recebeu o nome de *Álgebra Tensorial Generalizada* (*Generalized Tensor Algebra – GTA*). Para mais detalhes, aconselha-se ler o Apêndice A ou então referir-se aos trabalhos publicados [17] e [12].

3.2 Redes de Autômatos Estocásticos

Apesar de ser possível utilizar SAN para modelar qualquer sistema que se classifique como uma cadeia de Markov, a ferramenta de solução de modelos SAN, PEPS [11], é capaz de resolver com sucesso⁶ somente modelos homogêneos⁷, ergódicos e finitos. Um modelo Markoviano finito é garantidamente ergódico se também for irredutível e aperiódico (Teorema 2.5.2, p.33). Assim, define-se o foco deste capítulo na modelagem de cadeias de Markov homogêneas, ergódicas, finitas e irredutíveis.

Geralmente opta-se por observar o sistema e, a partir dele, criar um modelo em Redes de Autômatos Estocásticos para representá-lo. Por outro lado, é possível partir de uma representação em cadeias de Markov e obter um modelo SAN correspondente, embora não seja uma tarefa habitual, visto a dificuldade. Contudo, a seguir faz-se uso dessa última abordagem com o objetivo de explicitar a relação entre os dois formalismos.

Considere a cadeia de Markov da Figura 3.1. Trata-se de um modelo ergódico, pois a cadeia é finita, irredutível e aperiódica (Teorema 2.5.2). Ela é irredutível, pois cada estado é alcançado a partir de todos os outros (Seção 2.5, p.32). Sua aperiodicidade pode ser provada tomando-se o estado “10” como exemplo: a sequência {10,11,10} leva a MC de volta a “10” em dois passos e a sequência {10,20,00,10} faz isso em três passos. Portanto, “10” não possui período de retorno definido e, conseqüentemente, a cadeia toda é aperiódica (Seção 2.4, p.32).

A Figura 3.2 apresenta uma SAN que modela o mesmo sistema descrito pela cadeia de Markov da Figura 3.1, possuindo exatamente as mesmas propriedades. Esse modelo é formado por dois autômatos estocásticos, cujas transições entre os estados são reguladas pela ocorrência de eventos.

⁶Entenda-se por “resolver com *sucesso*” a situação em que o modelo converge para a solução estacionária.

⁷Uma cadeia de Markov é dita *homogênea* se sua matriz de transição independe do tempo.

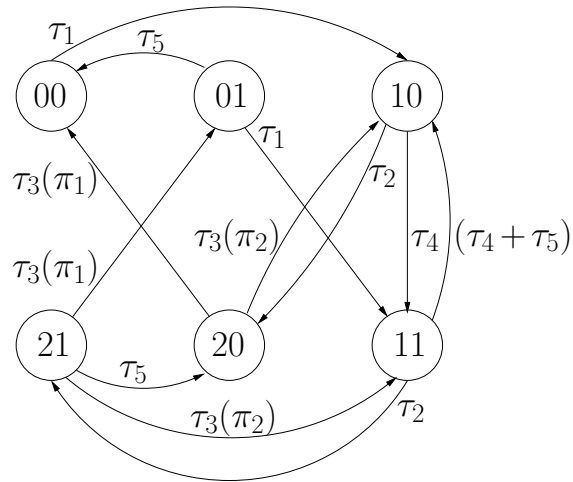


Figura 3.1 – Uma Cadeia de Markov

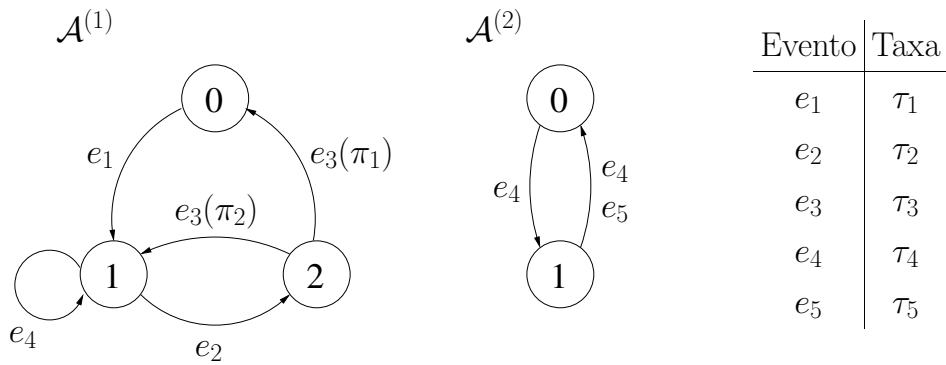


Figura 3.2 – Um Modelo SAN com Dois Autômatos

Notação

- \mathcal{A} conjunto de autômatos do modelo, tal que $|\mathcal{A}| = N$;
- $\mathcal{A}^{(i)}$ i -ésimo autômato, onde $\mathcal{A}^{(i)} \in \mathcal{A}$;
- $\mathcal{S}^{(i)}$ conjunto de estados de $\mathcal{A}^{(i)}$;
- $x^{(i)}$ estado local de $\mathcal{A}^{(i)}$, tal que $x^{(i)} \in \mathcal{S}^{(i)}$.

Notação

- \mathcal{E} conjunto de eventos do modelo, tal que $|\mathcal{E}| = E$;
- e evento do modelo, tal que $e \in \mathcal{E}$.

Definição 3.2.1. *Seja um estado $x^{(i)} \in \mathcal{S}^{(i)}$. Defina-se:*

$pot(x^{(i)})$ conjunto de eventos possíveis a partir de $x^{(i)}$.

Definição 3.2.2. *Seja um estado $x^{(i)} \in \mathcal{S}^{(i)}$ e um evento $e \in \mathcal{E}$. Defina-se:*

$succ_e(x^{(i)})$ conjunto de estados sucessores de $x^{(i)}$ a partir da ocorrência do evento e .

Tome o modelo SAN da Figura 3.2. Dadas as notações e definições recém vistas, seguem o conjunto de estados de cada autômato e o conjunto de eventos do modelo:

$$\mathcal{S}^{(1)} = \{0^{(1)}, 1^{(1)}, 2^{(1)}\}$$

$$\mathcal{S}^{(2)} = \{0^{(2)}, 1^{(2)}\}$$

$$\mathcal{E} = \{e_1, e_2, e_3, e_4, e_5\}$$

Para cada estado em cada autômato do modelo, a Tabela 3.1 apresenta o conjunto de eventos possíveis, enquanto que a Tabela 3.2 mostra o conjunto de estados sucessores em função do evento disparado.

Tabela 3.1 – Eventos Possíveis a Partir de Cada Estado dos Autômatos da Figura 3.2

Estado local	Eventos possíveis
$pot(0^{(1)})$	$\{e_1\}$
$pot(1^{(1)})$	$\{e_2; e_4\}$
$pot(2^{(1)})$	$\{e_3\}$
$pot(0^{(2)})$	$\{e_4\}$
$pot(1^{(2)})$	$\{e_4; e_5\}$

Tabela 3.2 – Conjuntos dos Estados Sucessores dos Autômatos da Figura 3.2

Estado local	Conjunto de estados sucessores				
	$e_n = e_1$	$e_n = e_2$	$e_n = e_3$	$e_n = e_4$	$e_n = e_5$
$succ_{e_n}(0^{(1)})$	$\{1^{(1)}\}$	\emptyset	\emptyset	\emptyset	\emptyset
$succ_{e_n}(1^{(1)})$	\emptyset	$\{2^{(1)}\}$	\emptyset	$\{1^{(1)}\}$	\emptyset
$succ_{e_n}(2^{(1)})$	\emptyset	\emptyset	$\{0^{(1)}; 1^{(1)}\}$	\emptyset	\emptyset
$succ_{e_n}(0^{(2)})$	\emptyset	\emptyset	\emptyset	$\{1^{(2)}\}$	\emptyset
$succ_{e_n}(1^{(2)})$	\emptyset	\emptyset	\emptyset	$\{0^{(2)}\}$	$\{0^{(2)}\}$

Definição 3.2.3. A cada $e \in \mathcal{E}$ associa-se uma tupla de evento, (e, τ_e) , onde:

e *identificador do evento;*
 τ_e *taxa de ocorrência do evento.*

Definição 3.2.4. A $x^{(i)}, y^{(i)} \in \mathcal{S}^{(i)}$ associa-se uma tupla de transição local *partindo de* $x^{(i)}$ *para* $y^{(i)}$, notada por $(e, \pi_e(x^{(i)}, y^{(i)}))$, onde:

e *identificador do evento;*
 $\pi_e(x^{(i)}, y^{(i)})$ *probabilidade de roteamento⁸, assumindo valores entre $[0, 1]$.*

Notação

\mathcal{T} conjunto de tuplas de transição local de um modelo SAN, tal que $\mathcal{T}^* = \mathcal{T} \cup \{\emptyset\}$;
 $2^{\mathcal{T}}$ conjunto de partes (*powerset*) de \mathcal{T} . Uma parte de \mathcal{T} é chamada *elemento de transição*.

Definição 3.2.5. Seja um evento $e \in \mathcal{E}$,

\mathcal{O}_e *conjunto de índices i ($i \in [1..N]$) de identificação dos autômatos nos quais o evento e acontece.*

Definição 3.2.6. Um evento $e \in \mathcal{E}$ é chamado:

evento local, se $|\mathcal{O}_e| = 1$;
 evento sincronizante, se $|\mathcal{O}_e| > 1$.

Definição 3.2.7. O conjunto de eventos locais é definido como $\mathcal{E}_l = \{e \in \mathcal{E} \mid |\mathcal{O}_e| = 1\}$.

Definição 3.2.8. O conjunto de eventos sincronizantes é definido como $\mathcal{E}_s = \{e \in \mathcal{E} \mid |\mathcal{O}_e| > 1\}$.

Faz-se oportuno agora, após todas essas definições envolvendo eventos, retornar ao modelo da Figura 3.2. Seu conjunto de tuplas de transição pode ser visto na Tabela 3.3, enquanto que a classificação de cada evento em *local* ou *sincronizante* pode ser vista na Tabela 3.4.

⁸Em um autômato, é possível que a partir de um estado origem, x , saiam diversas setas de transição, com destinos diferentes, identificadas pelo mesmo evento e . A *probabilidade de roteamento* de um evento indica a chance de uma ou outra transição ocorrer em função do disparo desse evento.

Tabela 3.3 – Conjunto de Tuplas de Transição do Modelo da Figura 3.2

Estado local		Tupla de transição
Origem	Destino	
$0^{(1)}$	$1^{(1)}$	$(e_1, 1)$
$1^{(1)}$	$2^{(1)}$	$(e_2, 1)$
$1^{(1)}$	$1^{(1)}$	$(e_4, 1)$
$2^{(1)}$	$0^{(1)}$	(e_3, π_1)
$2^{(1)}$	$1^{(1)}$	(e_3, π_2)
$0^{(2)}$	$1^{(2)}$	$(e_4, 1)$
$1^{(2)}$	$0^{(2)}$	$(e_4, 1)$
$1^{(2)}$	$0^{(2)}$	$(e_5, 1)$

Tabela 3.4 – Classificação de Cada Evento do Modelo SAN da Figura 3.2

Evento	Tupla de evento	\mathcal{O}_e	Tipo
e_1	(e_1, τ_1)	$\{1\}$	local
e_2	(e_2, τ_2)	$\{1\}$	local
e_3	(e_3, τ_3)	$\{1\}$	local
e_4	(e_4, τ_4)	$\{1; 2\}$	sincronizante
e_5	(e_5, τ_5)	$\{2\}$	local

A classificação dos eventos em locais ou sincronizantes tem uma função prática. Seja um evento $e \in \mathcal{E}$. Se ele é local, então $|\mathcal{O}_e| = 1$, significando que o evento ocorre em apenas um autômato da rede. Por outro lado, se $|\mathcal{O}_e| > 1$, então o evento é dito sincronizante, pois ocorre em mais de um autômato da rede. Este último tipo de evento não recebe este nome por acaso: eventos sincronizantes são aqueles que provocam transições *simultâneas* em autômatos diferentes. Considere novamente o modelo da Figura 3.2. Para que o evento sincronizante e_4 ocorra é necessário que o autômato $\mathcal{A}^{(1)}$ ocupe o estado local $1^{(1)}$. Portanto, se o autômato $\mathcal{A}^{(1)}$ ocupa um estado diferente de $1^{(1)}$ e o autômato $\mathcal{A}^{(2)}$ está no estado local $0^{(2)}$, então este último autômato está impedido de mudar de estado local até que o autômato $\mathcal{A}^{(1)}$ volte a ocupar o estado $1^{(1)}$, habilitando o evento e_4 . Isso não acontece quando o autômato $\mathcal{A}^{(2)}$ está no estado local $1^{(2)}$, pois neste caso, mesmo que $\mathcal{A}^{(1)}$ não esteja em condição de disparar e_4 , o evento local e_5 pode ocorrer, alterando o estado local de $\mathcal{A}^{(2)}$.

Definição 3.2.9. O espaço de estados produto $\hat{\mathcal{S}}$ de uma SAN é definido pelo produto cartesiano dos espaços de estados $\mathcal{S}^{(i)}$ dos autômatos da rede:

$$\hat{\mathcal{S}} = \prod_{i=1}^N \mathcal{S}^{(i)}$$

Definição 3.2.10. O estado global de um modelo SAN é definido pela combinação dos estados locais de todos os autômatos da rede, $\tilde{x} = (x^{(1)}, \dots, x^{(N)})$, onde $\tilde{x} \in \hat{S}$.

Definição 3.2.11. A função de atingibilidade \mathcal{F} é uma função definida em $\hat{S} \rightarrow [0..1]$. A função associa aos estados globais $\tilde{x} \in \hat{S}$ o valor "1" se são atingíveis e o valor "0" quando inatingíveis.

Definição 3.2.12. O espaço de estados atingíveis \mathcal{S} é o subconjunto de \hat{S} ($\mathcal{S} \subset \hat{S}$) composto de todos os estados globais $\tilde{x} \in \hat{S}$ tal que $\mathcal{F}(\tilde{x}) = 1$.

Ainda com relação à Figura 3.2, a Definição 3.2.10 permite verificar que "00"⁹, por exemplo, é um estado global do modelo, e portanto faz parte do seu espaço de estados produto, como visto na Definição 3.2.9. Por sua vez, essas definições, aliadas àquelas relativas à atingibilidade, Definições 3.2.11 e 3.2.12, permitem que se construa a cadeia de Markov correspondente a esse modelo SAN. Essa cadeia é exatamente a mesma vista na Figura 3.1.

Com o objetivo de tornar mais clara a definição de espaço atingível, observe o modelo SAN da Figura 3.3. É praticamente o mesmo modelo apresentado na Figura 3.2, contudo, vê-se diferença no autômato $\mathcal{A}^{(2)}$ no que diz respeito aos eventos. Agora, tem-se dois eventos sincronizantes, e_2 e e_4 (Definições 3.2.5 e 3.2.6). Como nota Sales [28], um evento sincronizante e é realizável no estado global \tilde{x} , se e somente se $\forall i \in \mathcal{O}_e$ o conjunto de estados sucessores $y^{(i)} \in \text{succ}_e(x^{(i)})$ não for vazio. Assim, como visto na Figura 3.4, os eventos e_2 e e_4 restringem o espaço de estados atingíveis desse modelo SAN, pois os estados globais "01" e "21" são inatingíveis a partir de todos os outros. A identificação destes estados inatingíveis pode ser feita como segue: a única forma de alcançar $2^{(1)}$ é através do evento sincronizante e_2 que, por sua vez, sempre leva ao estado $0^{(2)}$ no autômato $\mathcal{A}^{(2)}$, tornando-se impossível alcançar o estado global "21"; analogamente, a única forma de alcançar o estado $1^{(2)}$ é através do evento sincronizante e_4 que, por sua vez, sempre provoca transições simultâneas que levam ao estado local $1^{(1)}$ no autômato $\mathcal{A}^{(1)}$, sendo portanto impossível alcançar o estado "01".

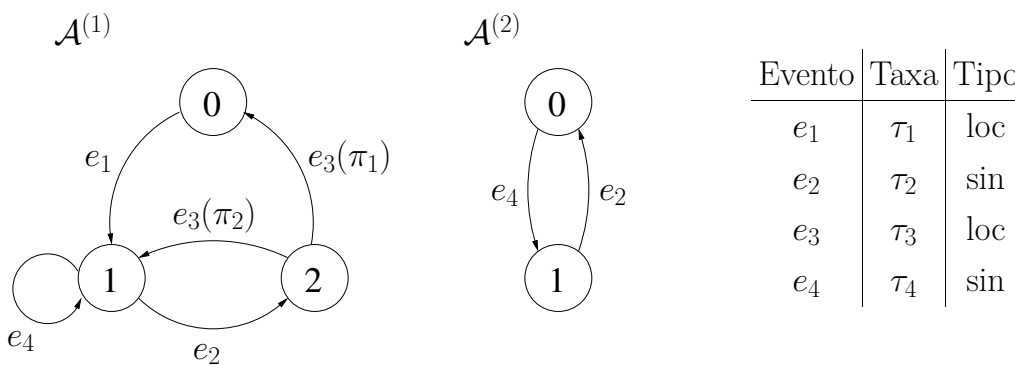


Figura 3.3 – Um Modelo SAN com Dois Autômatos

⁹A notação aqui usada é tal que o primeiro dígito do estado global diz respeito ao autômato $\mathcal{A}^{(1)}$, enquanto que o segundo dígito refere-se ao autômato $\mathcal{A}^{(2)}$.

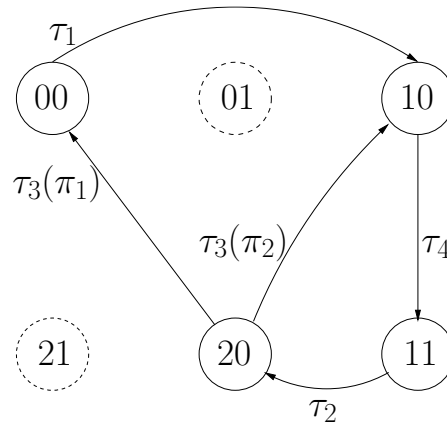


Figura 3.4 – Uma Cadeia de Markov

Definição 3.2.13. A função de transição local $Q^{(i)}$, definida como $\mathcal{S}^{(i)} \times \mathcal{S}^{(i)} \rightarrow \mathcal{T}^*$, contém os rótulos de transição do autômato $\mathcal{A}^{(i)}$.

Definição 3.2.14. A função de transição global \tilde{Q} , definida como $\hat{\mathcal{S}} \times \hat{\mathcal{S}} \rightarrow \mathcal{T}^*$, contém os rótulos de transição do autômato global.

Notação

$Q^{(i)}(x^{(i)}, y^{(i)})$ rótulo de transição do estado local $x^{(i)}$ para $y^{(i)}$ em $Q^{(i)}$, contendo uma lista de tuplas de transição (e, π) em \mathcal{T} ;

$\tilde{Q}^{(i)}(\tilde{x}, \tilde{y})$ rótulo de transição do estado global \tilde{x} para \tilde{y} em \tilde{Q} , contendo uma lista de tuplas de transição (e, π) em \mathcal{T} .

Definição 3.2.15. Um autômato $\mathcal{A}^{(i)}$ é definido por:

um conjunto de estados $\mathcal{S}^{(i)}$;

uma função de transição $Q^{(i)}$.

Definição 3.2.16. Uma rede de autômatos estocásticos é definida por:

\mathcal{A} conjunto de N autômatos. Cada autômato $\mathcal{A}^{(i)} \in \mathcal{A}$ com seu conjunto $\mathcal{S}^{(i)}$ e função de transição $Q^{(i)}$ definidos;

$\hat{\mathcal{S}}$ espaço de estados produto;

\mathcal{E} conjunto de E eventos. Cada evento $e \in \mathcal{E}$ com sua tupla de evento (e, τ_e) ;

\mathcal{F} função de atingibilidade, que define o espaço de estados atingíveis \mathcal{S} do modelo.

3.3 Método de obtenção do Gerador Infinitesimal

O método apresentado nesta seção consiste na construção de uma fórmula algébrica que relaciona um modelo SAN ao seu modelo equivalente em cadeias de Markov. Esta fórmula, composta de operações tensoriais, recebe o nome de *descriptor Markoviano* do modelo SAN e, a partir dela, é possível obter o gerador infinitesimal que corresponde ao modelo. Esta seção é inteiramente baseada no trabalho de Sales [28].

A cada autômato $\mathcal{A}^{(i)}$ são associadas:

- uma matriz de transição $Q_l^{(i)}$ correspondendo aos eventos locais e suas respectivas tuplas de transição;
- $|\mathcal{E}_s|$ matrizes de transição $Q_{e^+}^{(i)}$ correspondendo à *parte positiva* dos eventos sincronizantes que ocorrem nesse autômato e suas respectivas tuplas de transição;
- $|\mathcal{E}_s|$ matrizes de transição $Q_{e^-}^{(i)}$ correspondendo à *parte negativa* dos eventos sincronizantes que ocorrem nesse autômato e suas respectivas tuplas de transição.

Sejam

$Q_k^{(i)}(x^{(i)}, y^{(i)})$ o elemento do tensor $Q_k^{(i)}$ na linha $x^{(i)}$ e coluna $y^{(i)}$, onde $k \in \{l, e^+, e^-\}$;

$I_{|\mathcal{S}^{(i)}|}$ o tensor identidade de ordem $|\mathcal{S}^{(i)}|$.

Definição 3.3.1. Os elementos da matriz de transição local $Q_l^{(i)}$ são definidos por:

1. $\forall e \in \mathcal{E}_l, \forall x^{(i)}, y^{(i)} \in \mathcal{S}^{(i)}$ tal que $y^{(i)} \in \text{succ}_e(x^{(i)})$ e $x^{(i)} \neq y^{(i)}$

$$Q_l^{(i)}(x^{(i)}, y^{(i)}) = \sum_{e \in \mathcal{E}_l} \tau_e \pi_e(x^{(i)}, y^{(i)});$$

2. $\forall x^{(i)} \in \mathcal{S}^{(i)}$

$$Q_l^{(i)}(x^{(i)}, x^{(i)}) = - \sum_{e \in \mathcal{E}_l} \sum_{y^{(i)} \in \text{succ}_e(x^{(i)})} \tau_e \pi_e(x^{(i)}, y^{(i)});$$

3. $\forall x^{(i)}, y^{(i)} \in \mathcal{S}^{(i)}$ tal que $y^{(i)} \notin \text{succ}_e(x^{(i)})$ e $x^{(i)} \neq y^{(i)}$

$$Q_l^{(i)}(x^{(i)}, y^{(i)}) = 0.$$

Definição 3.3.2. Seja $e \in \mathcal{E}_s$ e sua correspondente tupla de evento, (e, τ_e) . Defina-se:

ι_e índice do autômato escolhido para carregar a taxa τ_e desse evento, tal que $\iota_e \in \mathcal{O}_e$.

Definição 3.3.3. Os elementos da matriz de transição $Q_{e^+}^{(i)}$, referente à ocorrência do evento sincronizante $e \in \mathcal{E}_s$ no autômato $\mathcal{A}^{(i)}$, são definidos por:

1. $\forall i \notin \mathcal{O}_e$

$$Q_{e^+}^{(i)} = I_{|\mathcal{S}^{(i)}|};$$
2. $\forall x^{(\iota_e)}, y^{(\iota_e)} \in \mathcal{S}^{(\iota_e)}$ tal que $y^{(\iota_e)} \in \text{succ}_e(x^{(\iota_e)})$

$$Q_{e^+}^{(\iota_e)}(x^{(\iota_e)}, y^{(\iota_e)}) = \tau_e \pi_e(x^{(\iota_e)}, y^{(\iota_e)});$$
3. $\forall i \in \mathcal{O}_e$ tal que $i \neq \iota_e$, $\forall x^{(i)}, y^{(i)} \in \mathcal{S}^{(i)}$ tal que $y^{(i)} \in \text{succ}_e(x^{(i)})$

$$Q_{e^+}^{(i)}(x^{(i)}, y^{(i)}) = \pi_e(x^{(i)}, y^{(i)});$$
4. $\forall i \in \mathcal{O}_e$, $\forall x^{(i)}, y^{(i)} \in \mathcal{S}^{(i)}$ tal que $y^{(i)} \notin \text{succ}_e(x^{(i)})$

$$Q_{e^+}^{(i)}(x^{(i)}, y^{(i)}) = 0.$$

Definição 3.3.4. Os elementos do tensor das transições sincronizadas que representam o ajuste necessário à ocorrência do evento $e \in \mathcal{E}_s$ são definidas por:

1. $\forall i \notin \mathcal{O}_e$

$$Q_{e^-}^{(i)} = I_{|\mathcal{S}^{(i)}|};$$
2. $\forall x^{(\iota_e)} \in \mathcal{S}^{(\iota_e)}$

$$Q_{e^-}^{(\iota_e)}(x^{(\iota_e)}, x^{(\iota_e)}) = - \sum_{y^{(\iota_e)} \in \text{succ}_e(x^{(\iota_e)})} \tau_e \pi_e(x^{(\iota_e)}, y^{(\iota_e)});$$
3. $\forall i \in \mathcal{O}_e$, $i \neq \iota_e$ e $\forall x^{(i)} \in \mathcal{S}^{(i)}$

$$Q_{e^-}^{(i)}(x^{(i)}, x^{(i)}) = \sum_{y^{(i)} \in \text{succ}_e(x^{(i)})} \pi_e(x^{(i)}, y^{(i)});$$
4. $\forall i \in \mathcal{O}_e$, $\forall x^{(i)}, y^{(i)} \in \mathcal{S}^{(i)}$ e $x^{(i)} \neq y^{(i)}$

$$Q_{e^-}^{(i)}(x^{(i)}, y^{(i)}) = 0.$$

Definição 3.3.5. Seja uma SAN bem definida¹⁰. O gerador infinitesimal Q da cadeia de Markov correspondente é representado pelo descritor Markoviano associado à SAN [12, 17, 25], dado pela fórmula tensorial da Equação 3.1:

$$Q = \bigoplus_{i=1}^N Q_l^{(i)} + \sum_{e \in \mathcal{E}_s} \left(\bigotimes_{i=1}^N Q_{e^+}^{(i)} + \bigotimes_{i=1}^N Q_{e^-}^{(i)} \right) \quad (3.1)$$

Como nota Sales [28], uma vez que toda soma tensorial é equivalente a uma soma de produtos tensoriais (veja o Apêndice A), o descritor Markoviano pode ser apresentado como:

$$Q = \sum_{j=1}^{(N+2|\mathcal{E}_s|)} \bigotimes_{i=1}^N Q_j^{(i)}, \quad (3.2)$$

¹⁰O conceito de SAN bem definida está detalhado em outros trabalhos como o de Sales [29] e Brenner [10]. Contudo, cabe dizer que, seguindo à risca as definições aqui apresentadas, os modelos SAN criados tendem a ser naturalmente bem definidos.

$$\text{onde } Q_j^{(i)} = \begin{cases} I_{|S^{(i)}|} & \text{para } j \leq N \text{ e } j \neq i \\ Q_l^{(i)} & \text{para } j \leq N \text{ e } j = i \\ Q_{e_{(j-N)}^+}^{(i)} & \text{para } N < j \leq (N + |\mathcal{E}_s|) \\ Q_{e_{(j-(N+|\mathcal{E}_s|))}^-}^{(i)} & \text{para } j > (N + |\mathcal{E}_s|) \end{cases}$$

Tabela 3.5 – Descritor Markoviano

Σ	N		$Q_l^{(1)} \otimes_g I_{ S^{(2)} } \otimes_g \dots \otimes_g I_{ S^{(N-1)} } \otimes_g I_{ S^{(N)} }$	
			$I_{ S^{(1)} } \otimes_g Q_l^{(2)} \otimes_g \dots \otimes_g I_{ S^{(N-1)} } \otimes_g I_{ S^{(N)} }$	
			\vdots	
			$I_{ S^{(1)} } \otimes_g I_{ S^{(2)} } \otimes_g \dots \otimes_g Q_l^{(N-1)} \otimes_g I_{ S^{(N)} }$	
	$2 \mathcal{E}_s $		e^+	$Q_{e_1^+}^{(1)} \otimes_g Q_{e_1^+}^{(2)} \otimes_g \dots \otimes_g Q_{e_1^+}^{(N-1)} \otimes_g Q_{e_1^+}^{(N)}$
			$Q_{e_{ \mathcal{E}_s }^+}^{(1)} \otimes_g Q_{e_{ \mathcal{E}_s }^+}^{(2)} \otimes_g \dots \otimes_g Q_{e_{ \mathcal{E}_s }^+}^{(N-1)} \otimes_g Q_{e_{ \mathcal{E}_s }^+}^{(N)}$	
e^-		$Q_{e_1^-}^{(1)} \otimes_g Q_{e_1^-}^{(2)} \otimes_g \dots \otimes_g Q_{e_1^-}^{(N-1)} \otimes_g Q_{e_1^-}^{(N)}$		
		$Q_{e_{ \mathcal{E}_s }^-}^{(1)} \otimes_g Q_{e_{ \mathcal{E}_s }^-}^{(2)} \otimes_g \dots \otimes_g Q_{e_{ \mathcal{E}_s }^-}^{(N-1)} \otimes_g Q_{e_{ \mathcal{E}_s }^-}^{(N)}$		

3.4 Exemplo: Jogo “Cara ou Coroa”

Este é um jogo popular de adivinhação com uma série de variações conhecidas. A versão modelada nesta seção conta com somente um jogador e uma moeda. Sempre que a moeda é lançada, o jogador tenta adivinhar qual face cairá virada para cima, ou seja, deve escolher *cara* ou *coroa* como resultado do lançamento. O jogador vence sempre que sua expectativa é confirmada, caso contrário ele perde.

A Figura 3.5 apresenta uma possível modelagem em SAN, composta por dois autômatos, J e M . O primeiro representa a aposta do jogador, dentre as possibilidades *cara* ou *coroa*, identificadas pelos estados Ca e Co . O segundo autômato representa uma espécie de máquina que lança a moeda e verifica o resultado da adivinhação, sendo os estados identificados como Ca , Co , J_V e J_P , os últimos dois significando *jogador vence* e *jogador perde*.

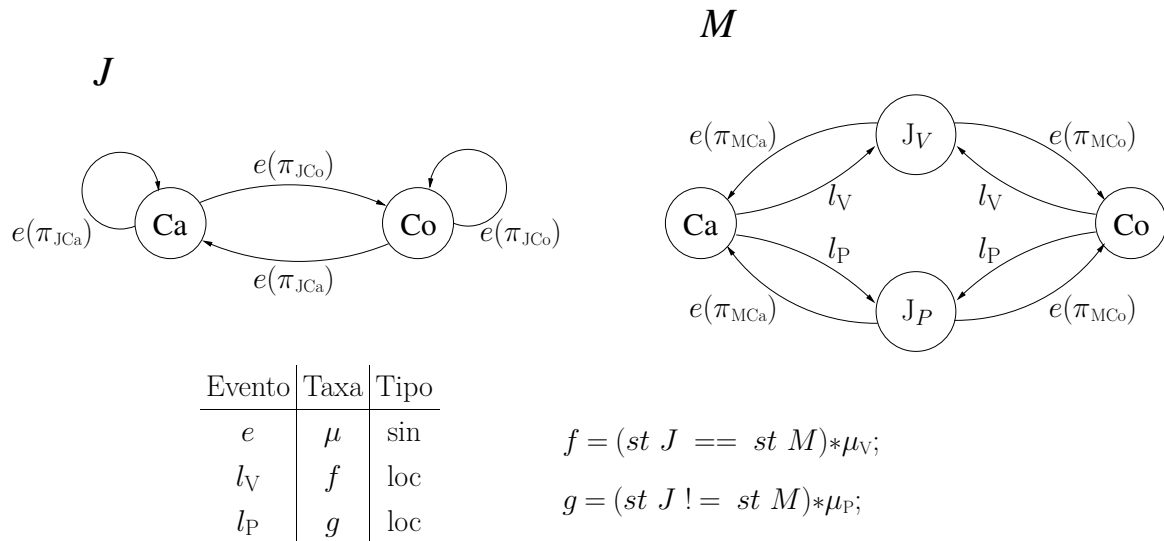


Figura 3.5 – Modelo SAN para o Jogo “Cara ou Coroa”

Neste modelo há dois eventos locais e um evento sincronizante. O lançamento da moeda e a aposta do jogador são ações simultâneas, representadas pelo evento sincronizante e , o qual dispara transições levando a Ca ou Co em ambos os autômatos. Os valores π_{JCa} e π_{JCo} são as probabilidades de escolha, ou probabilidades de roteamento, representando o comportamento do jogador ao escolher *cara* ou *coroa*, enquanto que os valores π_{MCa} e π_{MCo} são as probabilidades de escolha da própria moeda. Se $\pi_{MCa} \neq \pi_{MCo}$, então a moeda é dita *viciada*. O evento e só é realizável se o estado local do autômato M for J_V e J_P , não importando o estado local do autômato J , visto que neste último autômato e é realizável a partir de qualquer estado¹¹.

O evento local l_V acontece quando o jogador vence. Ele altera somente o estado do autômato M , possuindo uma taxa funcional que é diferente de zero (e igual a μ_V) apenas quando ambos os autômatos estão no mesmo estado, ou seja, confirmou-se o palpite do jogador. Analogamente, o evento l_P acontece quando o jogador perde, sendo a taxa funcional igual a μ_P quando os autômatos ocupam estados diferentes, ou seja, o jogador errou no palpite.

3.4.1 Descritor Markoviano e Gerador Infinitesimal

O descritor Markoviano desse modelo é composto dos seguintes termos:

$$Q = Q_l + Q_s$$

$$Q_l = Q_l^J \oplus_g Q_l^M = Q_l^J \otimes_g I_{n_M} + I_{n_J} \otimes_g Q_l^M$$

$$Q_s = Q_{e^+}^J \otimes_g Q_{e^+}^M + Q_{e^-}^J \otimes_g Q_{e^-}^M$$

¹¹Condição para realização de um evento sincronizante: $\forall i \in \mathcal{O}_e, \text{succ}_{e^i}(x^{(i)}) \neq \{\emptyset\}$.

Seja a ordenação dos estados locais de cada autômato: $J = \{Ca, Co\}$ e $M = \{J_V, J_P, Ca, Co\}$. Dada a ordenação, as matrizes que compõem o descritor Markoviano são:

$$Q_l^J = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad Q_l^M = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ f & g & -(f+g) & 0 \\ f & g & 0 & -(f+g) \end{pmatrix}$$

$$Q_{e^+}^J = \begin{pmatrix} \mu \pi_{JCa} & \mu \pi_{JCo} \\ \mu \pi_{JCa} & \mu \pi_{JCo} \end{pmatrix} \quad Q_{e^-}^J = \begin{pmatrix} -\mu (\pi_{JCa} + \pi_{JCo}) & 0 \\ 0 & -\mu (\pi_{JCa} + \pi_{JCo}) \end{pmatrix}$$

$$Q_{e^+}^M = \begin{pmatrix} 0 & 0 & \pi_{MCa} & \pi_{MCo} \\ 0 & 0 & \pi_{MCa} & \pi_{MCo} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad Q_{e^-}^M = \begin{pmatrix} \pi_{MCa} + \pi_{MCo} & 0 & 0 & 0 \\ 0 & \pi_{MCa} + \pi_{MCo} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

O gerador infinitesimal do modelo é obtido ao resolver explicitamente as operações tensoriais definidas no descritor Markoviano. A matriz que segue corresponde ao gerador infinitesimal antes de ter seus elementos funcionais avaliados:

$$Q = \left(\begin{array}{cccc|cccc} -\mu & 0 & \pi_{MCa} & \pi_{JCa} \mu & \pi_{MCo} & \pi_{JCa} \mu & 0 & 0 & \pi_{MCa} & \pi_{JCo} \mu & \pi_{MCo} & \pi_{JCo} \mu \\ 0 & -\mu & \pi_{MCa} & \pi_{JCa} \mu & \pi_{MCo} & \pi_{JCa} \mu & 0 & 0 & \pi_{MCa} & \pi_{JCo} \mu & \pi_{MCo} & \pi_{JCo} \mu \\ f & g & -(f+g) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ f & g & 0 & -(f+g) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & \pi_{MCa} & \pi_{JCa} \mu & \pi_{MCo} & \pi_{JCa} \mu & -\mu & 0 & \pi_{MCa} & \pi_{JCo} \mu & \pi_{MCo} & \pi_{JCo} \mu \\ 0 & 0 & \pi_{MCa} & \pi_{JCa} \mu & \pi_{MCo} & \pi_{JCa} \mu & 0 & -\mu & \pi_{MCa} & \pi_{JCo} \mu & \pi_{MCo} & \pi_{JCo} \mu \\ 0 & 0 & 0 & 0 & 0 & 0 & f & f & -(f+g) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & g & g & 0 & -(f+g) & 0 & 0 \end{array} \right)$$

Após atribuir valor aos parâmetros, $\mu = \mu_V = \mu_P = 1$, $\pi_{JCa} = 0.9$, $\pi_{JCo} = 0.1$, $\pi_{MCa} = 0.2$ e $\pi_{MCo} = 0.8$, e avaliar os elementos funcionais, obtém-se o gerador infinitesimal:

$$Q = \left(\begin{array}{cccc|cccc} -1 & 0 & 0.18 & 0.72 & 0 & 0 & 0.02 & 0.08 \\ 0 & -1 & 0.18 & 0.72 & 0 & 0 & 0.02 & 0.08 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0.18 & 0.72 & -1 & 0 & 0.02 & 0.08 \\ 0 & 0 & 0.18 & 0.72 & 0 & -1 & 0.02 & 0.08 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{array} \right)$$

Atribuindo novos valores aos parâmetros, $\mu = 2$, $\mu_V = 5$, $\mu_P = 10$, $\pi_{JCa} = 0.9$, $\pi_{JCo} = 0.1$, $\pi_{MCa} = 0.2$ e $\pi_{MCo} = 0.8$, e avaliando os elementos funcionais, obtém-se um novo gerador infinitesimal:

$$Q = \left(\begin{array}{cccc|cccc} -2 & 0 & 0.36 & 1.44 & 0 & 0 & 0.04 & 0.16 \\ 0 & -2 & 0.36 & 1.44 & 0 & 0 & 0.04 & 0.16 \\ 5 & 0 & -5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & -10 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0.36 & 1.44 & -2 & 0 & 0.04 & 0.16 \\ 0 & 0 & 0.36 & 1.44 & 0 & -2 & 0.04 & 0.16 \\ 0 & 0 & 0 & 0 & 0 & 10 & -10 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 & 0 & -5 \end{array} \right)$$

3.5 Exemplo: Jogo “Dois ou Um”

Esta seção tem por objetivo mostrar que, dado um sistema, não existe um modelo absoluto que o represente. Além da escolha do formalismo, dentre os muitos existentes, cabe ao modelador identificar elementos relevantes e incluí-los na modelagem. A forma de representar esses elementos é também uma decisão subjetiva, o que leva a concluir que um único sistema pode ser modelado de diversas formas.

Considere então o jogo “Dois ou Um”. Esta é uma das variações de um jogo de mão mundialmente popular chamado *Morra*. A partida começa com muitos jogadores e o objetivo é eliminá-los, rodada após rodada, até que sobrem apenas dois, sendo estes declarados os vencedores da partida. Quando a rodada inicia, todos estão com seus punhos fechados e gritam em coro o nome do jogo; ao final do coro e de forma simultânea, cada jogador estende “1” ou “2” dedos expondo sua escolha aos demais. Passam para a próxima rodada aqueles que representam a maioria, enquanto que a minoria é eliminada da partida. O jogo termina quando a maioria constar de dois jogadores. Assim, se a partida inicia com dez jogadores e na primeira rodada sete deles jogam “1” enquanto que os outros três jogam “2”, então esses últimos são eliminados e a partida continua com os sete que constituíram a maioria.

A Figura 3.6 apresenta uma modelagem SAN para o jogo “Dois ou Um” com N jogadores representados por N autômatos. Cada autômato possui quatro estados: “ J ” (*Joga*) indica que o jogador está pronto para iniciar nova rodada, ele permanece nesse estado até que todos estejam em condições de jogar; “ I ” e “ II ” são as opções de jogo; “ P ” (*Perde*) informa que o jogador foi excluído da partida, permanecendo nesse estado nas próximas rodadas, até que uma nova partida comece; por fim, o estado “ V ” (*Vence*) de um autômato é ocupado somente quando ele fica entre os dois finalistas da partida, vencendo o jogo. Note que os autômatos diferem entre si somente na tendência em jogar “1” ou “2”, indicada pela probabilidade de roteamento.

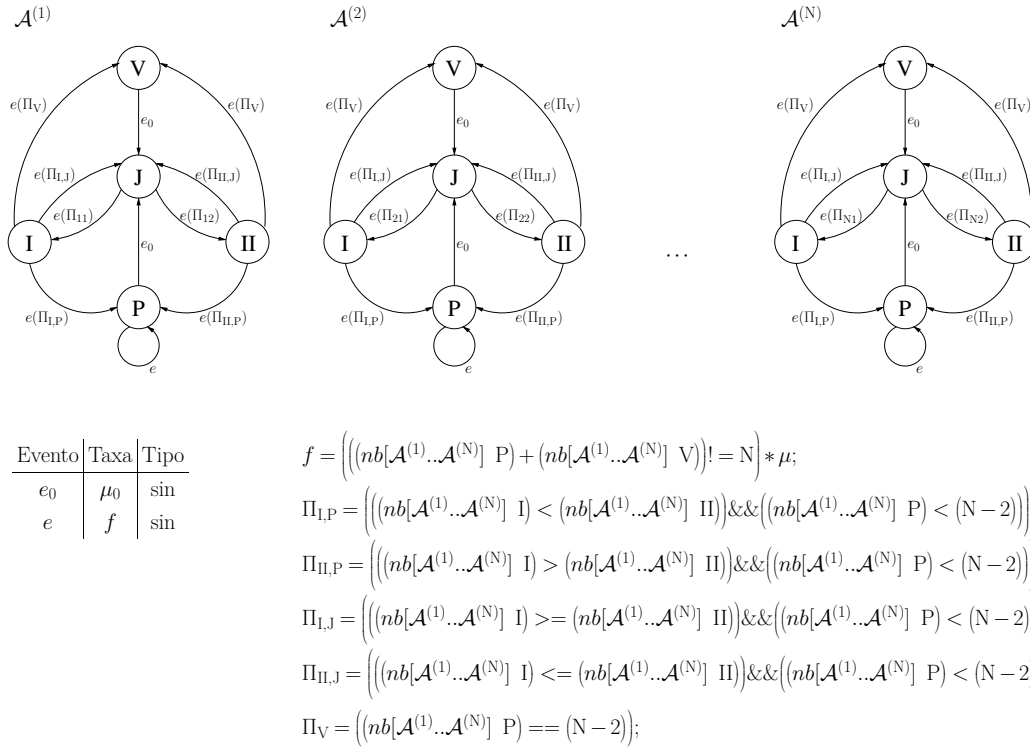


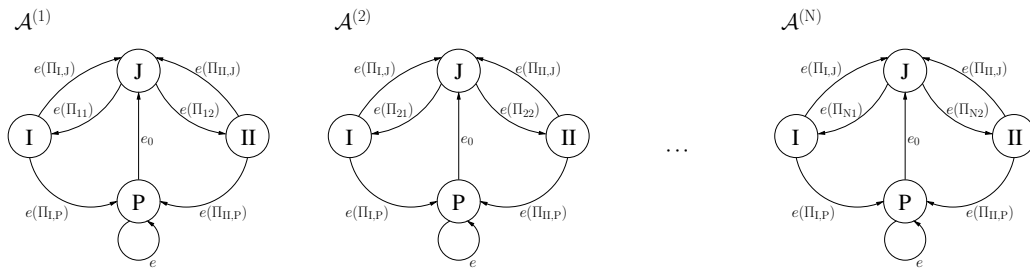
Figura 3.6 – Modelo SAN para o Jogo “Dois ou Um” – Abordagem que destaca os vencedores

Esse modelo possui dois eventos, ambos sincronizantes. O evento e garante a simultaneidade nas ações dos jogadores. Já o evento e_0 é utilizado somente para iniciar uma nova partida, forçando a transição de todos os autômatos para o estado “ J ”. Quando e_0 está habilitado, e não pode ser disparado pois sua taxa funcional é anulada. A dinâmica do modelo é regulada pelas probabilidades de roteamento, também funcionais. As probabilidades de roteamento $\Pi_{I,P}$, $\Pi_{II,P}$, $\Pi_{I,J}$ e $\Pi_{II,J}$ são avaliadas com base na definição da maioria: se a maioria dos jogadores botou “1”, então as funções $\Pi_{I,J}$ e $\Pi_{II,P}$ estão habilitadas; se a maioria jogou “2”, então as funções $\Pi_{II,J}$ e $\Pi_{I,P}$ estão habilitadas. A probabilidade de roteamento Π_V torna-se não-nula somente quando restaram dois jogadores, que são declarados vencedores, sendo portanto encaminhados ao estado “ V ”.

É possível tirar métricas desse modelo sobre o desempenho de cada jogador tendo em vista o comportamento coletivo. Para exemplificar, suponha um modelo com quatro jogadores. Ao resolvê-lo, percebe-se que se apenas um jogador apresenta comportamento não equiprovável na escolha de “1” ou “2”, então todos ganham igualmente, ou seja, cada um vence 25% das partidas jogadas. Contudo, as coisas mudam se dois jogadores não possuírem comportamento equiprovável. Suponha que um deles tende a jogar mais “1” enquanto que o outro tende a jogar mais “2”, como consequência, eles perdem mais do que os outros dois que possuem comportamento equiprovável. Agora se ambos apresentam maior tendência em jogar “1”, por exemplo, então eles tendem a ganhar mais partidas do que aqueles que possuem comportamento equiprovável.

Outro modelo é apresentado na Figura 3.7. Trata-se de uma versão simplificada da modelagem anterior. Nessa nova abordagem, o estado “ V ” foi eliminado, sendo declarados vitoriosos aqueles que se mantiverem por mais tempo jogando, ou seja, os autômatos com maior permanência no

estado “ J ”. Nada mais foi alterado, portanto, a dinâmica é a mesma.



Evento	Taxa	Tipo
e_0	μ_0	sin
e	f	sin

$$f = ((nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] P) = N) * \mu;$$

$$\Pi_{I,P} = \left(\left((nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] I) < (nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] II) \right) \parallel \left((nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] P) \geq (N-2) \right) \right);$$

$$\Pi_{II,P} = \left(\left((nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] I) > (nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] II) \right) \parallel \left((nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] P) \geq (N-2) \right) \right);$$

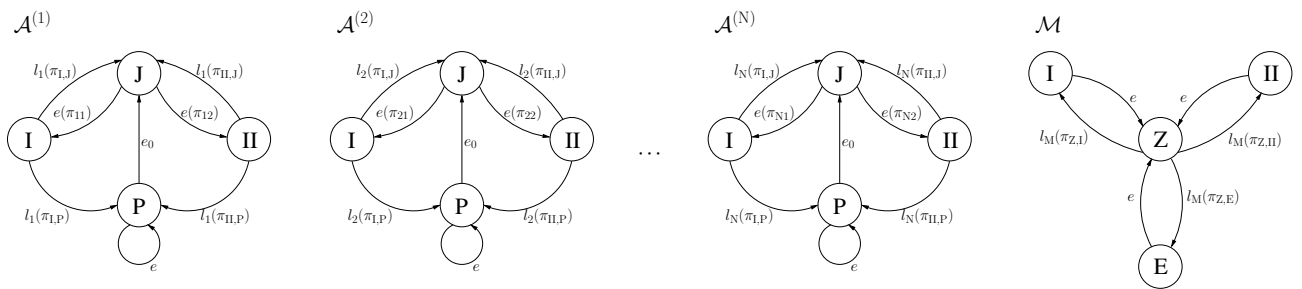
$$\Pi_{I,J} = \left(\left((nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] I) \geq (nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] II) \right) \&\& \left((nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] P) < (N-2) \right) \right);$$

$$\Pi_{II,J} = \left(\left((nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] I) \leq (nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] II) \right) \&\& \left((nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] P) < (N-2) \right) \right);$$

Figura 3.7 – Modelo SAN para o Jogo “Dois ou Um” – Abordagem Simplificada

Na Figura 3.8 é apresentada uma abordagem mais complexa, com a adição de outros N eventos locais, um para cada autômato. Esses novos eventos são responsáveis por alterar o estado local de cada autômato conforme o resultado da rodada, provocando transições que levam aos estados “ J ” e “ P ”. Por se tratarem de eventos locais, essas transições não acontecem de forma simultânea, criando a necessidade de introduzir no modelo um autômato para servir de memória sobre o resultado da rodada¹²: o autômato \mathcal{M} . Após cada rodada, os autômatos que representam os jogadores aguardam a contagem da maioria, computada pelo autômato \mathcal{M} que muda seu estado para “ I ”, “ II ” ou “ E ” (*Empate*). Somente após ocorrida a transição em \mathcal{M} é que as probabilidades funcionais de roteamento $\Pi_{I,P}$, $\Pi_{II,P}$, $\Pi_{I,J}$ e $\Pi_{II,J}$ tornam-se não nulas, permitindo as transições locais que levam cada jogador ao estado devido, “ J ” ou “ P ”.

¹²Idéia de Paulo Zanin.



Evento	Taxa	Tipo
e	f	sin
e_0	μ_0	sin
l_M	μ_M	loc
l_1	g	loc
l_2	g	loc
..	..	loc
l_N	g	loc

$$f = \left((nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] P) = N \right) * \mu_f;$$

$$g = \left(st M = Z \right) * \mu_g;$$

$$\pi_{1,P} = \left((st M = II) \parallel \left((nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] P) \geq (N-2) \right) \right);$$

$$\pi_{II,P} = \left((st M = I) \parallel \left((nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] P) \geq (N-2) \right) \right);$$

$$\pi_{1,J} = \left(\left((st M = I) \parallel (st M = E) \right) \&\& \left((nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] P) < (N-2) \right) \right);$$

$$\pi_{II,J} = \left(\left((st M = II) \parallel (st M = E) \right) \&\& \left((nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] P) < (N-2) \right) \right);$$

$$\pi_{Z,I} = \left((nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] I) > (nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] II) \right);$$

$$\pi_{Z,II} = \left((nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] I) < (nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] II) \right);$$

$$\pi_{Z,E} = \left((nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] I) = (nb[\mathcal{A}^{(1)} \dots \mathcal{A}^{(N)}] II) \right);$$

Figura 3.8 – Modelo SAN para o Jogo “Dois ou Um” – Inclusão de um Autômato de Memória

4. MODELOS MARKOVIANOS OCULTOS

No final da década de 1960, Leonard E. Baum e colaboradores publicaram uma série de artigos lançando as bases para o formalismo de Modelos Markovianos Ocultos [5–9]. As primeiras aplicações estavam voltadas para o reconhecimento de fala, sendo os trabalhos de F. Jelinek (*International Business Machines – IBM*) [3, 20, 22] e J. K. Baker (*Carnegie Mellon University – CMU*) [4], no começo dos anos 70, pioneiros no uso do formalismo. Na segunda metade da década de 80, Modelos Markovianos Ocultos foram aplicados no sequenciamento de DNA, alcançando posteriormente grande importância em todo o campo da bioinformática [15].

Este capítulo está estruturado como segue. Na Seção 4.1 são informalmente apresentados alguns conceitos fundamentais; a definição formal de Modelos Markovianos Ocultos é vista na Seção 4.2; a Seção 4.3 é dedicada aos problemas clássicos do formalismo, sendo que nas Seções 4.3.1, 4.3.2 e 4.3.3 são vistos os algoritmos mais usados na solução desses problemas, *Forward-Backward*, *Viterbi* e *Baum-Welch*, respectivamente; por fim, nas Seções 4.4, 4.5 e 4.6 são vistos exemplos de processos que podem ser modelados usando o formalismo de Modelos Markovianos Ocultos.

4.1 Fundamentos

Na maioria dos processos Markovianos cada estado integrante do espaço de estados corresponde a um observável¹ do sistema, tal que essa correspondência poderia ser descrita por uma função bijetiva. Este é o caso em uma cadeia de Markov, como visto no Capítulo 2: o processo é estocástico, ou seja, associam-se probabilidades na escolha do estado seguinte em função do estado atual, contudo, uma vez escolhido o próximo estado, não há dúvidas sobre qual ele seja, pois o processo é completamente observável.

Existem processos que não podem ser diretamente observados, mas que produzem sinais observáveis. Um exemplo seria um sinal de rádio: interferência e ruído podem prejudicar a recepção, mas a partir do sinal, mesmo que distorcido, pode-se inferir sobre o real conteúdo transmitido pela fonte. Modelos Markovianos Ocultos (*Hidden Markov Models – HMM*)² é um formalismo dedicado à modelagem desse tipo de problema: processos estocásticos que não podem ser diretamente observados, mas que produzem uma sequência estocástica de sinais. Ou seja, HMM é um modelo probabilístico em dois níveis ou estágios, duplamente estocástico [18, 21, 26].

O primeiro nível da HMM diz respeito ao processo oculto, que é caracterizado por uma DTMC³ finita e estacionária. É comum em modelos acústicos descrever a parte oculta como uma cadeia de Markov não ergódica, como é o caso no modelo de Bakis (*left-right model*) [26]. Contudo este

¹*Observável*, no contexto desse trabalho, é algo que pode ser observado. Para ilustrar, ao jogar uma moeda para cima, obtém-se como resultado da observação um dos dois possíveis observáveis: *cara* ou *coroa*.

²*Hidden Markov Models* aparece na literatura sob diversos nomes, tais como *Hidden Markov Processes*, *Markov Sources*, *Hidden Markov Chains*, *Probabilistic Functions of Markov Chains*, *Statistical Methods of Markov Source*.

³Cadeia de Markov em Tempo Discreto (*Discrete-Time Markov Chain – DTMC*)

trabalho dedica-se exclusivamente ao estudo de HMMs cuja cadeia de Markov oculta é ergódica.

O segundo nível trata da parte observável da HMM: a sequência de sinais. Uma HMM é dita *discreta* se o conjunto de sinais for discreto e *contínua* no caso contrário. Encontram-se exemplos de HMM discretas nos problemas aplicados ao sequenciamento de DNA [18]. Em reconhecimento de padrões, fala e escrita à mão é comum optar pela modelagem em HMM contínua [18]. Todavia, apenas HMM discretas, com conjunto finito de sinais, são tratadas nesta dissertação. Uma restrição adicional diz respeito ao comprimento da sequência de sinais, que neste trabalho é sempre finito de tamanho T .

A abordagem para descrição de HMM adotada em trabalhos conduzidos pela IBM [21] define as emissões de sinais observáveis como função das *transições* entre os estados da cadeia de Markov oculta. Assim, dada a MC⁴ oculta, com sua matriz de probabilidade definida, uma vez ocorrida a transição do estado s_i para o estado s_j segue o processo de escolha estocástica do sinal a ser emitido, dentre todos os sinais que essa transição na parte oculta é capaz de emitir. A abordagem adotada neste trabalho é um pouco diferente, porém equivalente [18,21,26]: as emissões acontecem em função dos *estados* ocupados pela MC oculta. Assim, cada sinal que compõe a sequência foi emitido respeitando regras probabilísticas em função do estado da Markov no momento daquela emissão.

Sumarizando, a diferença fundamental entre HMM e o resto dos formalismos Markovianos está na forma de se observar o sistema. Enquanto que na maioria dos processos Markovianos a observação é direta, pois os observáveis são os próprios estados do modelo, em HMM a observação é indireta, feita por inferência, pois os observáveis são funções probabilísticas dos estados da cadeia de Markov oculta.

4.2 Definição Formal

Sejam

$\mathcal{S} = \{s_i\}$	conjunto discreto e finito de estados da cadeia de Markov, tal que $ \mathcal{S} = N$;
$\mathcal{Y} = \{y_k\}$	conjunto de sinais, tal que $ \mathcal{Y} = M$;
q_t	estado da cadeia de Markov no instante t ;
O_t	sinal emitido no instante t .

Definição 4.2.1. Um modelo Markoviano oculto é definido pelos seus parâmetros $\lambda = (A, B, \pi)$, onde:

$$\pi = \{\pi_i\} \quad \text{distribuição de probabilidade inicial, tal que } \pi_i = P(q_1 = s_i) \text{ para } s_i \in \mathcal{S};$$

⁴Cadeia de Markov (*Markov Chain* – MC)

$A = \{a_{ij}\}$ *distribuição de probabilidade de transição, tal que $a_{ij} = P(q_t = s_j | q_{t-1} = s_i)$ para $s_i, s_j \in \mathcal{S}$;*

$B = \{b_i[k]\}$ *distribuição de probabilidade de emissão, tal que $b_i[k] = P(O_t = y_k | q_t = s_i)$ para $s_i \in \mathcal{S}$ e $y_k \in \mathcal{Y}$.*

Notação

$Q = q_1 q_2 q_3 \cdots q_{T-1} q_T$ sequência de estados na cadeia de Markov;

$O = O_1 O_2 O_3 \cdots O_{T-1} O_T$ sequência de emissão de sinais.

4.3 Problemas Clássicos

Rabiner [26] apresenta três problemas fundamentais a serem resolvidos antes de fazer uso de um modelo HMM⁵. Esses problemas, responsáveis pelo *ajuste fino* de um modelo, são os seguintes:

Problema 1: Probabilidade de uma Sequência de Sinais

Sejam o modelo $\lambda = (A, B, \pi)$ e a sequência de sinais $O = O_1 O_2 \cdots O_T$. Como calcular de forma *eficiente* a probabilidade dessa sequência ser gerada pelo modelo, $P(O|\lambda)$?

Problema 2: Sequência Ótima de Estados

Sejam o modelo $\lambda = (A, B, \pi)$ e a sequência de sinais $O = O_1 O_2 \cdots O_T$. Dentre as diversas sequências de estados que poderiam ter gerado essa sequência de sinais, qual é a mais provável?

Problema 3: Maximização da Probabilidade de uma Sequência de Sinais

Como os parâmetros $\lambda = (A, B, \pi)$ do modelo são ajustados para maximizar $P(O|\lambda)$?

A probabilidade $P(O|\lambda)$ pode ser calculada diretamente, pela própria definição dos parâmetros do modelo e probabilidades condicionais associadas, como é visto a seguir, no início da Seção 4.3.1. Contudo, o Problema 1 é específico quanto ao critério de eficiência, satisfeito pela introdução do algoritmo *Forward-Backward*, também na mesma seção. Uma interessante aplicação é a que segue: supondo uma sequência conhecida de emissões de um sistema físico real e um conjunto de modelos HMM criados para descrevê-lo, pode-se selecionar o melhor dentre os modelos pela solução do Problema 1. O melhor modelo seria aquele com maior $P(O|\lambda)$.

O Problema 2 é aquele onde se tenta revelar a parte oculta do modelo, buscando indícios sobre o funcionamento do sistema. Para tanto, é importante selecionar um critério de otimização das

⁵Rabiner [26] fundamenta essa abordagem nas idéias apresentadas por Jack Ferguson, do IDA, em apresentações nos *Laboratórios Bell*.

probabilidades condicionais calculadas, e esta é justamente a parte complicada, dependente inclusive dos objetivos que se deseja alcançar. Na Seção 4.3.2 são apresentados dois critérios de otimização, dentre os muitos que poderia-se pensar, com foco no estudo do algoritmo de *Viterbi*. Tipicamente este problema é solucionado com o objetivo de aprender sobre a estrutura do modelo estudado.

Um modelo é bom quando explica bem o comportamento do sistema modelado. O estudo inicial de um fenômeno fornece ao modelador uma compreensão parcial de seu funcionamento, possibilitando a construção de um modelo rudimentar que o descreva. Para melhor corresponder ao fenômeno observado é necessário que o modelo seja calibrado. Em se tratando de modelos HMM, a solução apresentada para o Problema 3 consiste em um método para realizar o ajuste dos parâmetros de forma iterativa, buscando máximos locais. Na Seção 4.3.3 é feito um estudo do método para solução desse problema: o algoritmo de *Baum-Welch*.

4.3.1 Solução do Problema 1

Sejam

$\lambda = (A, B, \pi)$ parâmetros do modelo HMM;

$O = O_1 O_2 \cdots O_T$ sequência de emissão de sinais.

Deseja-se calcular $P(O|\lambda)$: a probabilidade da sequência O ser gerada pelo modelo λ . A maneira mais direta de realizar esse cálculo parte da identificação de todas as sequências de estados com tamanho T capazes de gerar O . Cada sequência de estados contribui com uma parcela da probabilidade total de emissão de O .

Seja

$Q = q_1 q_2 \cdots q_T$ uma sequência de estados na cadeia de Markov.

A probabilidade de que O seja gerado a partir da sequência de estados Q é dada pela Equação 4.1, onde $P(O_t|q_t, \lambda)$ é a probabilidade condicional de emitir um observável no instante t em função do estado ocupado pela MC neste mesmo instante. Na Equação 4.2 o produtório é expandido considerando que as emissões de sinais são estatisticamente independentes⁶ entre si.

$$P(O|Q, \lambda) = \prod_{t=1}^T P(O_t|q_t, \lambda) \quad (4.1)$$

$$P(O|Q, \lambda) = b_{q_1}[O_1] \cdot b_{q_2}[O_2] \cdots b_{q_T}[O_T] \quad (4.2)$$

A Equação 4.2 corresponde à probabilidade de que O seja gerada por uma determinada sequência Q de estados, ou seja, corresponde à parcela de contribuição da sequência Q na emissão de O . Cabe

⁶A independência entre emissões de sinais é assumida por todo este trabalho.

agora determinar a probabilidade da sequência Q nesse modelo:

$$P(Q|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \cdots a_{q_{T-1} q_T} \quad (4.3)$$

Ao multiplicar a Equação 4.2 pela Equação 4.3 obtém-se a Equação 4.4, que indica a probabilidade conjunta da sequência Q e da emissão simultânea de O .

$$P(O, Q|\lambda) = P(O|Q, \lambda) P(Q|\lambda) \quad (4.4)$$

Na Equação 4.4, ao fazer o somatório sobre todas as sequências Q capazes de emitir O , obtém-se a Equação 4.5, ou seja, probabilidade total de que o modelo λ emita a sequência de sinais O :

$$\begin{aligned} P(O|\lambda) &= \sum_{\forall Q} P(O|Q, \lambda) P(Q|\lambda) \\ &= \sum_{q_1, q_2, \dots, q_T} \pi_{q_1} b_{q_1}[O_1] a_{q_1 q_2} b_{q_2}[O_2] \cdots a_{q_{T-1} q_T} b_{q_T}[O_T] \end{aligned} \quad (4.5)$$

A Equação 4.5 nada mais é do que a definição matemática direta do Problema 1, logo, a solução para este problema pode ser obtida simplesmente resolvendo essa equação. Sua resolução, contudo, envolve aproximadamente $2T \cdot N^T$ cálculos, sendo bastante onerosa e muitas vezes até impraticável, mesmo para modelos pequenos, na época da publicação de Rabiner [26]. Para exemplificar, um modelo com $N = 5$ (estados) e $T = 100$ (observações) envolveria aproximadamente 10^{72} cálculos. Um procedimento mais eficiente era necessário.

Algoritmo *Forward-Backward*

Esse algoritmo é composto de duas partes:

- *forward*: procedimento indutivo baseado no cálculo da variável $\alpha_t(i)$;
- *backward*: procedimento indutivo baseado no cálculo da variável $\beta_t(i)$.

Ambas as partes são apresentadas nesta seção, contudo, apenas a parte *forward* é necessária à solução do Problema 1. A parte *backward* é importante para a solução dos Problema 2 e 3 e por esse motivo é aqui apresentada.

Definição 4.3.1. A variável *forward* é definida pela expressão:

$$\alpha_t(i) = P(O_1 O_2 \cdots O_t, q_t = s_i | \lambda), \quad (4.6)$$

ou seja, $\alpha_t(i)$ armazena a probabilidade de uma sequência parcial de emissões, desde o instante inicial 1 até o instante t , juntamente com a probabilidade de que a cadeia de Markov ocupe o estado s_i no instante t .

A solução do Problema 1 pode ser obtida indutivamente através do cálculo de $\alpha_t(i)$, tal como no procedimento que segue:

1. Inicialização:

$$\alpha_1(i) = \pi_i b_i [O_1], \quad 1 \leq i \leq N \quad (4.7)$$

2. Indução:

$$\alpha_t(j) = \left[\sum_{i=1}^N \alpha_{t-1}(i) a_{ij} \right] b_j [O_t], \quad 2 \leq t \leq T \quad (4.8)$$

$$1 \leq j \leq N$$

3. Finalização:

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (4.9)$$

O passo relativo à inicialização do procedimento atribui a $\alpha_1(i)$ a probabilidade da MC ocupar o estado s_i no instante inicial $t = 1$ e simultaneamente emitir o sinal observável O_1 . Este cálculo é feito para todo $s_i \in \mathcal{S}$.

O passo indutivo é o cerne do procedimento (veja a Figura 4.1). A parte entre colchetes considera a probabilidade conjunta, $\alpha_{t-1}(i)$, da emissão da sequência parcial de sinais até O_{t-1} com a ocupação do estado s_i no instante $t - 1$ e executa sua multiplicação pela probabilidade de transição, a_{ij} , do estado $q_{t-1} = s_i$ ao estado $q_t = s_j$; ao somar sobre todos os possíveis $s_i \in \mathcal{S}$ que levam ao s_j em questão, obtém-se a probabilidade conjunta da sequência parcial de sinais até O_{t-1} com a ocupação do estado s_j no instante t . Isto feito, basta multiplicar o resultado por $b_j [O_t]$ para obter $\alpha_t(j)$, ou seja, a probabilidade conjunta da sequência parcial de sinais até O_t com a ocupação do estado s_j no instante t . Este cálculo também é feito para todo $s_j \in \mathcal{S}$, para todos os instantes no intervalo $2 \leq t \leq T$.

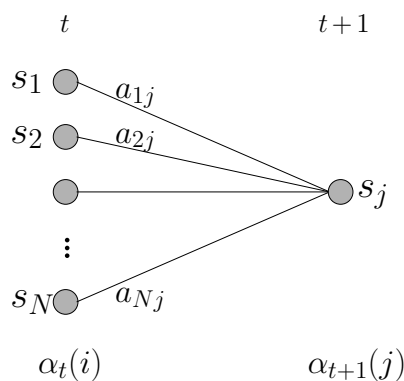


Figura 4.1 – Algoritmo *Forward-Backward* – Parte *Forward* Ilustrada (Fonte: Rabiner [26])

A solução do Problema 1 é obtida ao fazer o somatório de $\alpha_T(i)$ sobre todos os estados $s_i \in \mathcal{S}$, o que fornece o valor de $P(O|\lambda)$: a probabilidade de que a sequência de sinais O seja emitida pelo

modelo λ . Uma forma de verificar a veracidade dessa afirmação pode ser através da própria definição da variável *forward* no instante T :

$$\alpha_T(i) = P(O_1 O_2 \cdots O_T, q_T = s_i | \lambda), \quad (4.10)$$

ao somar sobre todos os estados $s_i \in \mathcal{S}$ obtém-se $P(O|\lambda)$.

A solução para o Problema 1 pode então ser calculada usando tanto a Equação 4.5 como o procedimento para o cálculo da parte *forward* do algoritmo *Forward-Backward*. Enquanto que a Equação 4.5 requer aproximadamente $2T \cdot N^T$ cálculos, o procedimento para o cálculo de $P(O|\lambda)$ através da parte *forward* do algoritmo requer aproximadamente $T \cdot N^2$ cálculos. Para um modelo pequeno, com $N = 5$ (estados) e $T = 100$ (observações), a primeira opção envolveria na ordem de 10^{72} cálculos enquanto que pelo uso da variável *forward* seriam necessários aproximadamente 3000 cálculos, sendo claro o ganho em eficiência que esse algoritmo proporciona.

A seguir é descrita a parte *backward* do algoritmo *Forward-Backward*. Cabe lembrar que o procedimento que segue não é necessário à solução do Problema 1, estando aqui apenas para fins de completude, visto que a variável *backward* será importante mais adiante, quando forem tratadas as soluções para os Problemas 2 e 3.

Definição 4.3.2. A variável *backward* é definida pela expressão:

$$\beta_t(i) = P(O_{t+1} O_{t+2} \cdots O_T | q_t = s_i, \lambda), \quad (4.11)$$

ou seja, $\beta_t(i)$ armazena a probabilidade de uma sequência parcial de emissões, desde o instante $t + 1$ até o instante final T , dado que a cadeia de Markov ocupou o estado s_i no instante t .

Os valores de $\beta_t(i)$ podem ser obtidos indutivamente pela execução do procedimento que segue:

1. Inicialização:

$$\beta_T(j) = 1, \quad 1 \leq j \leq N \quad (4.12)$$

2. Indução:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j [O_{t+1}] \beta_{t+1}(j), \quad (4.13)$$

$$t = T - 1, T - 2, \dots, 1$$

$$1 \leq i \leq N$$

De acordo com Rabiner [26] e Jelinek [21], a inicialização desse procedimento é arbitrária. O passo indutivo é semelhante àquele visto no procedimento para o cálculo da parte *forward*, porém a sequência de observação é criada de trás para frente (veja a Figura 4.2). Para ilustrar, observe os

primeiros termos calculados através desse procedimento:

$$\begin{aligned}\beta_T(j) &= 1, \quad 1 \leq j \leq N \\ \beta_{T-1}(i) &= \sum_{j=1}^N a_{ij} b_j [O_T] \underbrace{\beta_T(j)}_1 = \sum_{j=1}^N a_{ij} b_j [O_T] = P(O_T | q_{T-1} = s_i, \lambda) \\ \beta_{T-2}(h) &= \sum_{i=1}^N a_{hi} b_i [O_{T-1}] \beta_{T-1}(i) = \left[\sum_{i=1}^N a_{hi} b_i [O_{T-1}] \right] \cdot \left[\sum_{i=1}^N \sum_{j=1}^N a_{ij} b_j [O_T] \right] \\ &= P(O_{T-1} | q_{T-2} = s_h, \lambda) \cdot P(O_T | \lambda) = P(O_{T-1} O_T | q_{T-2} = s_h, \lambda)\end{aligned}$$

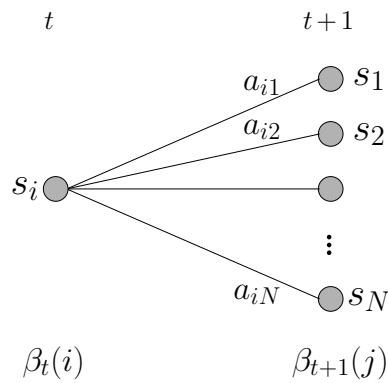


Figura 4.2 – Algoritmo *Forward-Backward* – Parte *Backward* Ilustrada (Fonte: Rabiner [26])

Assim como para a parte *forward*, o procedimento para o cálculo da parte *backward* do algoritmo requer aproximadamente $T \cdot N^2$ cálculos. Essa variável será vista novamente mais adiante no texto.

4.3.2 Solução do Problema 2

Sejam

- $\lambda = (A, B, \pi)$ parâmetros do modelo HMM;
- $O = O_1 O_2 \cdots O_T$ sequência de emissão de sinais;
- $Q = q_1 q_2 \cdots q_T$ sequência de estados na cadeia de Markov.

O Problema 2 procura determinar dentre as diversas sequências de estados Q aquela com maior probabilidade de ter gerado a sequência de sinais O . Rabiner [26] desenvolve dois possíveis critérios de otimização:

- a busca pelo estado individualmente mais provável a cada instante t de observação;

– a busca da melhor sequência Q de T estados.

O primeiro critério é resolvido com o auxílio do algoritmo *Forward-Backward*, detalhado na Seção 4.3.1, enquanto que o segundo critério é trabalhado usando o algoritmo de *Viterbi*, visto mais adiante.

Para auxiliar na implementação da solução do Problema 2 para o critério de omitização que busca os estados individualmente mais prováveis, considere a variável definida pela Equação 4.14.

Definição 4.3.3. A variável $\gamma_t(i)$ é definida pela expressão:

$$\gamma_t(i) = P(q_t = s_i | O, \lambda) \quad (4.14)$$

ou seja, $\gamma_t(i)$ armazena a probabilidade da cadeia de Markov ter ocupado o estado s_i no instante t , dada a sequência completa de emissões O e o modelo λ .

Essa variável pode ser expressa em termos das variáveis *forward* (Equação 4.6) e *backward* (Equação 4.11), como visto na Equação 4.15. Nesta mesma equação vê-se que a variável $\gamma_t(i)$ se constitui em uma nova medida de probabilidade, levando à Equação 4.16, onde o somatório sobre todos os estados $s_i \in \mathcal{S}$ resulta na unidade. Para obter o estado mais provável no instante t , basta que se descubra qual dentre eles que retorna o maior valor de $\gamma_t(i)$, como visto na Equação 4.17.

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^N \alpha_t(i)\beta_t(i)} \quad (4.15)$$

$$\sum_{i=1}^N \gamma_t(i) = 1 \quad (4.16)$$

$$q_t = \underset{1 \leq i \leq N}{\operatorname{argmax}} [\gamma_t(i)], \quad 1 \leq t \leq T \quad (4.17)$$

A abordagem só funciona garantidamente se a cadeia de Markov além de ergódica tiver suas probabilidades condicionais definidas tais que $a_{ij} > 0$, $\forall s_i, s_j \in \mathcal{S}$. Em outras palavras, esse método, que busca o estado individualmente mais provável a cada instante, pode indicar a sequência $Q = \{q_1 = s_h, q_2 = s_i, q_3 = s_j\} = \{s_h s_i s_j\}$ como a mais provável, mesmo que $a_{hi} = 0$ ou $a_{ij} = 0$, situação em que a sequência não existe. Esse é o momento para buscar um novo critério de otimização, como segue.

Algoritmo de Viterbi

Relembrando, o Problema 2 procura identificar dentre as diversas sequências de estados Q aquela com maior probabilidade de ter gerado a sequência de sinais O . De acordo com Rabiner [26], o critério mais usado para resolver esse problema consiste em maximizar $P(Q|O, \lambda)$ pela escolha de uma sequência Q apropriada, que será então identificada como Q^* : a sequência de estados mais provável. Alternativamente, maximizar $P(Q|O, \lambda)$ tem o mesmo efeito que maximizar $P(Q, O|\lambda)$,

visto que em ambos os casos o argumento Q^* é o mesmo. Como forma de auxiliar na implementação dessa solução, considere a variável definida pela Equação 4.18.

Definição 4.3.4. A variável $\delta_t(i)$ é definida pela expressão:

$$\delta_t(i) = \max_{q_1 q_2 \dots q_{t-1}} P(q_1 q_2 \dots q_t = s_i, O_1 O_2 \dots O_t | \lambda) \quad (4.18)$$

ou seja, $\delta_t(i)$ armazena a máxima probabilidade conjunta das sequências parciais de estados e observáveis, desde o instante inicial 1 até o instante $t - 1$, incluindo a probabilidade de que O_t seja emitido em função de $q_t = s_i$.

Nas palavras de Rabiner [26], $\delta_t(i)$ é a máxima probabilidade no instante t ao longo de um único caminho, considerando as primeiras t observações e terminando no estado $q_t = s_i$. Por indução, tem-se:

$$\delta_{t+1}(j) = \max_{1 \leq i \leq N} [\delta_t(i) a_{ij}] b_j [O_{t+1}] \quad (4.19)$$

Rabiner [26] também explica que, a cada instante $t + 1$ e estado s_j , é necessário guardar o argumento $q_t = s_i$ que maximiza Equação 4.19, pois esta é a forma de obter a sequência de estados Q^* . Este argumento é guardado no vetor ψ . O procedimento completo é descrito como segue:

1. Inicialização:

$$\delta_1(i) = \pi_i b_i [O_1], \quad 1 \leq i \leq N \quad (4.20)$$

$$\psi_1(i) = 0 \quad (4.21)$$

2. Recursão:

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j [O_t], \quad 2 \leq t \leq T \quad (4.22)$$

$$1 \leq j \leq N$$

$$\psi_t(j) = \operatorname{argmax}_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad 2 \leq t \leq T \quad (4.23)$$

$$1 \leq j \leq N$$

3. Finalização:

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)] \quad (4.24)$$

$$q_T^* = \operatorname{argmax}_{1 \leq i \leq N} [\delta_T(i)] \quad (4.25)$$

4. Reconstituição da sequência de estados (*Path backtracking*):

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \dots, 1 \quad (4.26)$$

Como nota Rabiner [26], o algoritmo de *Viterbi* é semelhante (a não ser pelo último passo) ao procedimento para a solução do Problema 1 pelo uso da variável *forward*. Uma diferença é que no algoritmo de *Viterbi* busca-se o máximo a cada momento ao invés de somar sobre todas as possibilidades. Além disso, à medida que os máximos são encontrados, os argumentos responsáveis pela maximização são guardados, e isso é tarefa desnecessária ao procedimento apresentado para a solução do Problema 1.

4.3.3 Solução do Problema 3

O Problema 3 é dedicado à calibração do modelo HMM para que corresponda melhor às observações físicas reais do sistema. Em outras palavras, a solução para este problema consiste em ajustar os parâmetros $\lambda = (A, B, \pi)$ de forma a maximizar $P(O|\lambda)$. De acordo com Rabiner [26], não há solução analítica para esse problema.

Algoritmo de Baum-Welch

Definição 4.3.5. A variável $\xi_t(i, j)$ é definida pela expressão:

$$\xi_t(i, j) = P(q_t = s_i, q_{t+1} = s_j | O, \lambda) \quad (4.27)$$

ou seja, $\xi_t(i, j)$ armazena probabilidade conjunta da cadeia de Markov ocupar o estado s_i no instante t e o estado s_j no instante $t+1$, dada a sequência observada O e o modelo λ .

Essa variável pode ser expressa em termos das variáveis *forward* (Equação 4.6) e *backward* (Equação 4.11), como visto na Equação 4.28 e ilustrado na Figura 4.3, evidenciando tratar-se de uma nova medida de probabilidade:

$$\begin{aligned} \xi_t(i, j) &= \frac{\alpha_t(i) a_{ij} b_j[O_{t+1}] \beta_{t+1}(j)}{P(O|\lambda)} \\ &= \frac{\alpha_t(i) a_{ij} b_j[O_{t+1}] \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j[O_{t+1}] \beta_{t+1}(j)} \end{aligned} \quad (4.28)$$

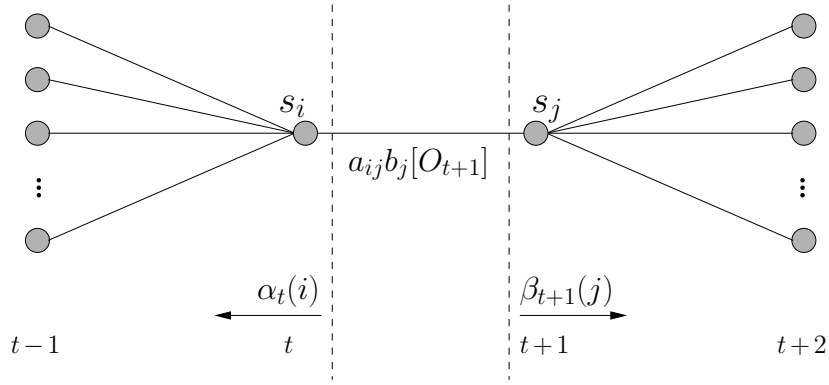


Figura 4.3 – Algoritmo *Forward-Backward* – Solução do *Problema 3* (Fonte: Rabiner [26])

Note que as Equações 4.14 e 4.27 se relacionam como mostra a Equação 4.29. Ao fazer o somatório de $\gamma_t(i)$ e $\xi_t(i, j)$ sobre o tempo de observação, obtém-se o valor esperado dessas quantidades, como visto nas Equações 4.30 e 4.31:

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (4.29)$$

$$\sum_{t=1}^{T-1} \gamma_t(i) = \text{número esperado de transições a partir de } s_i \quad (4.30)$$

$$\sum_{t=1}^{T-1} \xi_t(i, j) = \text{número esperado de transições de } s_i \text{ para } s_j \quad (4.31)$$

A solução do Problema 3 pode ser obtida através do método iterativo de *Baum-Welch*, que reestima os parâmetros do modelo através das Equações 4.32, 4.33 e 4.34, a seguir. À direita da igualdade são fornecidos os parâmetros atuais, $\lambda = (A, B, \pi)$, a partir dos quais são calculados os novos parâmetros, $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\pi})$. A iteração continua, usando $\bar{\lambda}$ no lugar de λ , até que o critério de parada seja satisfeito: $\bar{\lambda} = \lambda$.

$$\begin{aligned} \bar{\pi}_i &= \text{número esperado de vezes no estado } s_i \text{ no instante } (t = 1) \\ &= \gamma_1(i) \end{aligned} \quad (4.32)$$

$$\begin{aligned} \bar{a}_{ij} &= \frac{\text{número esperado de transições do estado } s_i \text{ para o estado } s_j}{\text{número esperado de transições a partir do estado } s_i} \\ &= \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \end{aligned} \quad (4.33)$$

$$\begin{aligned}\bar{b}_j[k] &= \frac{\text{número esperado de vezes no estado } s_j \text{ e observando o sinal } y_k}{\text{número esperado de vezes no estado } s_j} \\ &= \frac{\sum_{t=1}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} \quad (4.34)\end{aligned}$$

Foi provado por Baum *et al.* que $\bar{\lambda}$ é sempre um modelo melhor que λ , visto que com o método de *Baum-Welch* a inequação $P(O|\bar{\lambda}) > P(O|\lambda)$ é sempre verdade [26]. O resultado final desse procedimento de reestimativa é chamado *maximum likelihood estimate*, ou em tradução livre, a estimativa de probabilidade máxima da HMM. Rabiner [26] destaca que esse algoritmo leva em conta apenas máximos locais, sendo uma solução inadequada para o caso de sistemas mais complexos, com vários máximos locais.

4.4 Exemplo: Experimento com Lançamento de Moedas

Suponha uma sala dividida ao meio por uma cortina. À direita da cortina está um observador, curioso sobre o que se passa do outro lado. À esquerda da cortina, são realizados experimentos com moedas. Esta cortina impede observações diretas mas permite que o observador, à direita, receba sinais sobre a realização desses experimentos. Esses sinais podem ser *cara* ou *coroa* e eles são emitidos pelo realizador do experimento. Então, cada vez que o observador ouve “cara!”, entende que este foi o resultado do último experimento.

Contudo, o observador não sabe o que se passa do outro lado da cortina, ou seja, ele não conhece o processo. Ele pode imaginar que o processo seja composto por uma moeda e um lançador, mas é possível que existam N moedas e M lançadores. Pode ser que alguma moeda esteja viciada, pode ser que simplesmente não existam moedas e que as respostas sejam escolhidas randomicamente, ou que um *software* sorteie uma resposta aleatória. Enfim, o processo é desconhecido.

Isso não impede, porém, que o observador crie modelos, baseados nos sinais que ele escuta, para tentar explicar o processo. Esta seção apresenta possíveis modelagens para este problema usando o formalismo HMM, tal como sugerido por Rabiner [26].

Primeiro modelo: O processo é composto por uma única moeda. Suas duas faces correspondem aos dois estados do processo: *cara* e *coroa*, que na Figura 4.4 são representados pelos estados “1” e “2”. Ao ouvir o realizador do experimento gritar “cara!”, o observador sabe exatamente em que estado o processo se encontra naquele instante: *cara*; analogamente, ao ouvir “coroa!”, sabe que o processo ocupa o estado *coroa*. Dada a descrição, é irrelevante o fato de existir uma cortina separando o experimento do observador, pois o sinal sonoro recebido corresponde exatamente ao estado do processo, ou seja, tanto o experimentador quanto o observador observam os mesmos resultados. Assim, o processo poderia ser modelado como uma simples cadeia de Markov, ao invés de uma HMM, visto que o experimento não está de fato oculto.

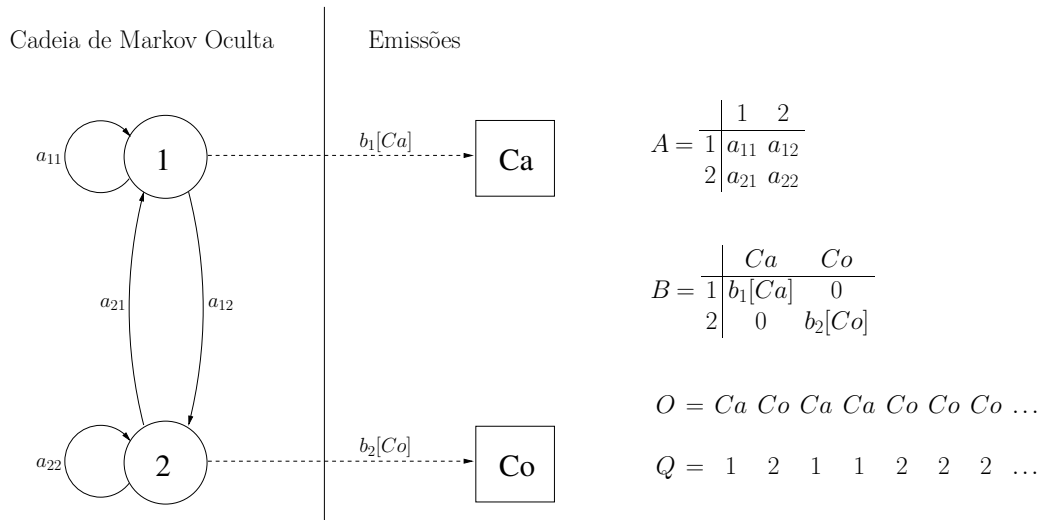


Figura 4.4 – Experimento com Lançamento de Moedas – HMM com Uma Moeda

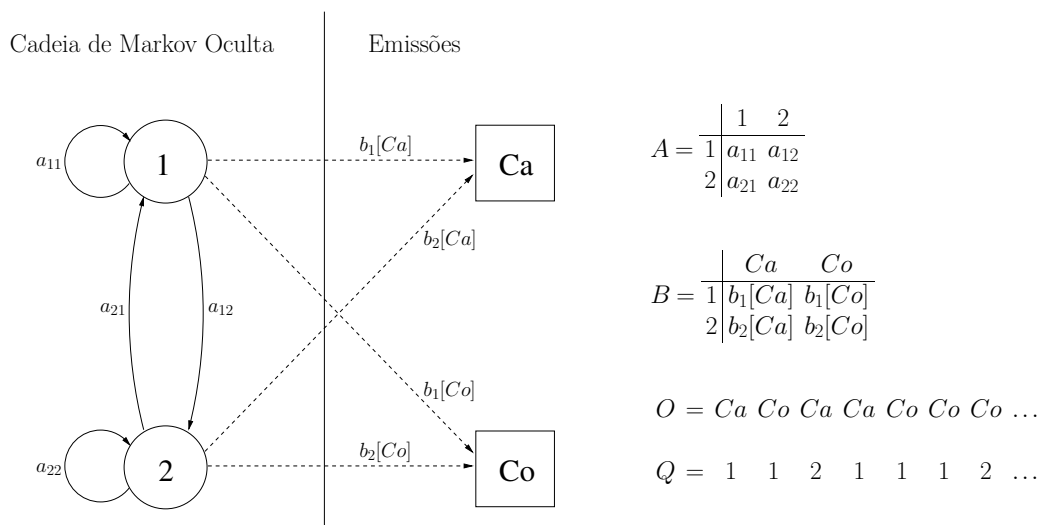


Figura 4.5 – Experimento com Lançamento de Moedas – HMM com Duas Moedas

Segundo modelo: O processo é composto por N moedas, constituindo os N estados do modelo, sendo que cada estado pode emitir dois sinais: *cara* e *coroa*. Ao ouvir o realizador do experimento gritar “cara!”, o observador sabe que o lançamento resultou em *cara*, mas não sabe qual foi a moeda lançada; analogamente, ao ouvir “coroa!”, o observador sabe que alguma moeda foi lançada e resultou em *coroa*, mas não sabe qual. Esse é um caso em que a modelagem em HMM é adequada, pois existe de fato um processo oculto do observador. As Figuras 4.5 e 4.6 mostram esse modelo para $N = 2$ e $N = 3$ moedas, respectivamente.

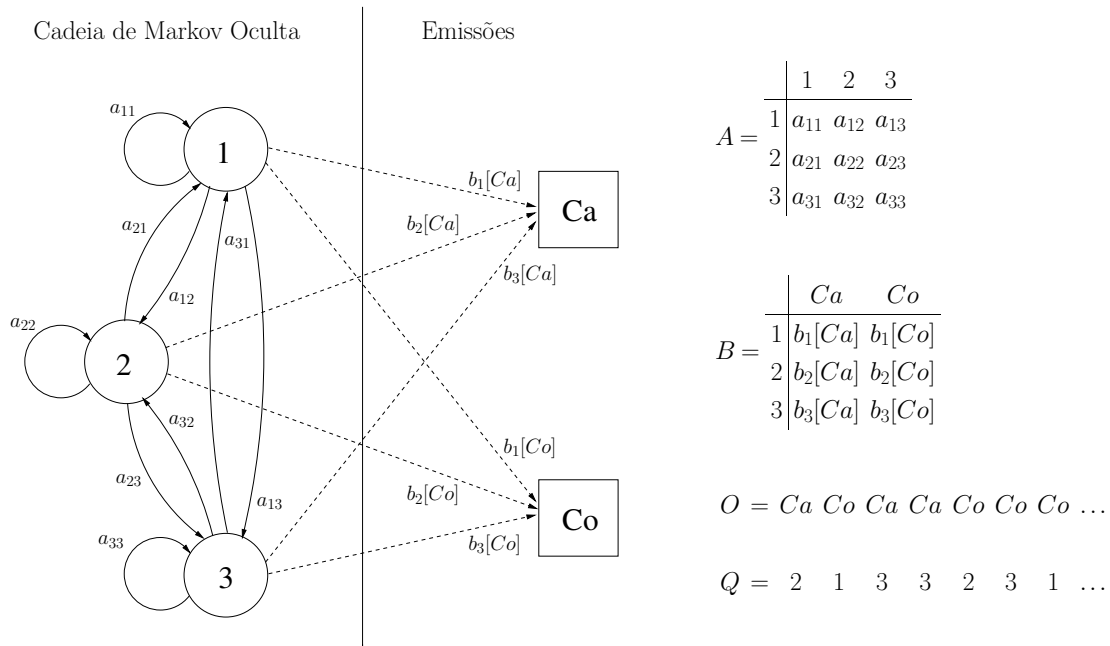


Figura 4.6 – Experimento com Lançamento de Moedas – HMM com Três Moedas

4.5 Exemplo: Bolas em Urnas

Suponha uma sala com N urnas carregadas de bolas de cores variadas. Cada urna tem bolas nas cores azul, branca, preta e rosa em diferentes proporções, ou seja, enquanto que 20% das bolas da urna U1 são azuis, na urna U2 elas representam 30% do total. O processo consiste na escolha aleatória de uma urna, acompanhada do sorteio de uma bola, que em seguida é devolvida à urna de origem; uma nova urna é escolhida, dando continuidade ao processo. A cada sorteio o observador, na sala ao lado, recebe um cartão onde lê a cor da bola sorteada.

Um modelo HMM capaz de descrever esse comportamento é composto de N estados, representando as N urnas, e M sinais, representando as cores das bolas nessas urnas. A parte oculta do processo diz respeito à escolha das urnas, que é descrita por uma cadeia de Markov. A parte observável é justamente a sequência de cores emitidas pelas urnas, que chega ao conhecimento do observador através da leitura dos cartões que lhe são entregues. A Figura 4.7 mostra esse modelo para $N = 3$ urnas e $M = 4$ cores de bolas (azul, branco, preto e rosa).

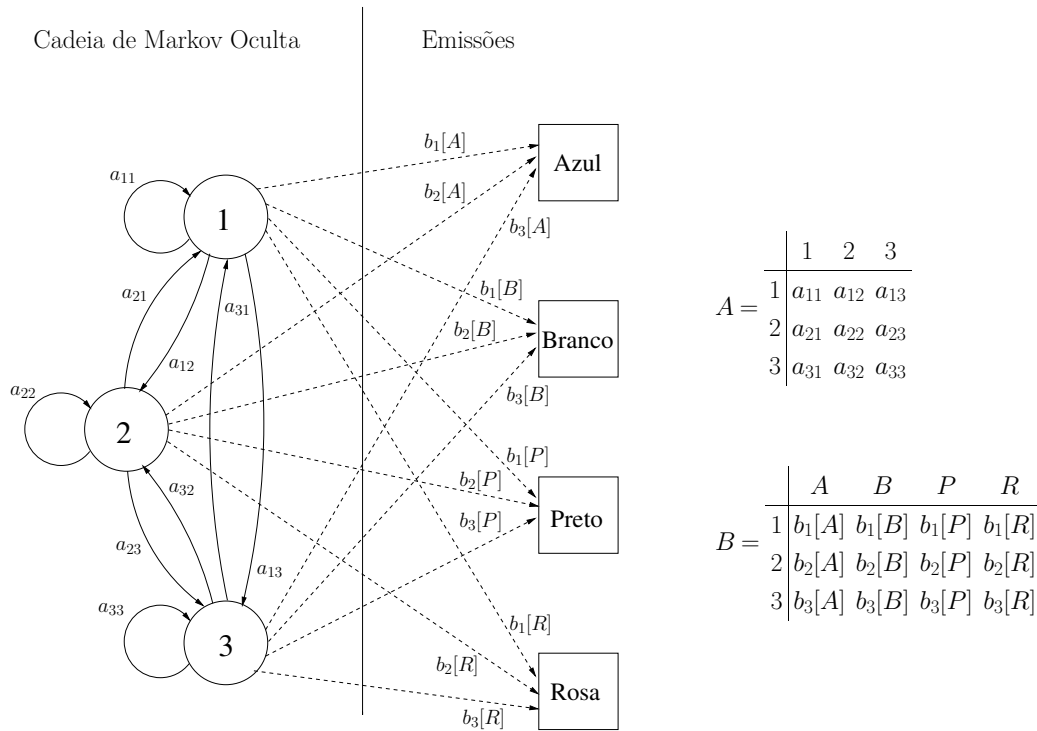


Figura 4.7 – Bolas em Urnas – HMM com $N = 3$ e $M = 4$

4.6 Exemplo: Tabuleiro Numerado

Suponha um tabuleiro quadrado com $N = 400$ posições e uma peça. Esta peça pode andar na direção leste-oeste ou na direção norte-sul, ou seja, movimentos diagonais não são permitidos. Cada posição no tabuleiro é identificada por $T(i, j)$, correspondendo à linha i e coluna j do tabuleiro.

A escolha da posição é dada pelo lançamento de um dado sextavado de proporções equilibradas, isto é: a probabilidade associada a cada face é de $\frac{1}{6}$. O resultado do lançamento do dado indica o número de passos que devem ser feitos na direção e sentido que se desejar. Assim, a transição entre $T(2, 3)$ e $T(2, 10)$ não é possível, pois isso significaria dar 7 passos, o que supera o número máximo de passos permitidos, que é 6.

Cada posição no tabuleiro contém um número, que é o valor absoluto da diferença entre i e j . Esse número é o único sinal emitido nessa posição, contudo, várias posições emitem esse mesmo sinal, por exemplo: $T(2, 3)$, $T(3, 2)$, $T(15, 16)$, dentre outros.

Com o que foi dito até agora, as posições do tabuleiro representam os estados do sistema e os números contidos nessas posições representam as emissões em cada estado. A Tabela 4.1 apresenta em uma matriz como seria a aparência desse tabuleiro.

O processo se dá como segue: alguém é responsável por jogar o dado, decidir o sentido, reposicionar a peça e dizer o valor encontrado na nova posição do tabuleiro; o observador está em outra sala e ouve o que é dito, recebendo uma sequência de sinais emitidos no intervalo de 0 à 19.

Seja a sequência de sinais observada: $O = \{3, 8, 5, 1, 3\}$. Algumas sequências de estados podem tê-la emitido, dentre elas as sequências $Q_1 = \{T(4, 7), T(4, 12), T(7, 12), T(13, 12), T(13, 10)\}$ e $Q_2 = \{T(4, 7), T(4, 12), T(4, 9), T(8, 9), T(8, 5)\}$.

Tabela 4.1 – Tabuleiro Numerado

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13
8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12
9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11
10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10
11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9
12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8
13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7
14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6
15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5
16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4
17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3
18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2
19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1
20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0

5. EQUIVALÊNCIA DE REPRESENTAÇÃO

Neste capítulo prova-se a equivalência entre os formalismos de Modelos Markovianos Ocultos (*Hidden Markov Models* – HMM) e Redes de Autômatos Estocásticos (*Stochastic Automata Networks* – SAN). A estratégia consiste na comparação entre os geradores infinitesimais obtidos a partir desses formalismos; a equivalência é provada ao mostrar a igualdade entre os geradores infinitesimais. Para tal, antes de tudo é necessário formalizar o conceito de um gerador infinitesimal que corresponda ao modelo HMM, visto que a definição clássica do formalismo não contempla este ponto. Para a definição clássica dos formalismos HMM e SAN, consulte os Capítulos 4 e 3.

O capítulo está organizado como segue. Nas Seções 5.1 e 5.2 são apresentados dois métodos para obtenção do gerador infinitesimal para HMM, cuja equivalência entre eles é provada ao final da Seção 5.2; a Seção 5.3 apresenta um método de conversão de modelos HMM em SAN.

5.1 Obtenção do Gerador Infinitesimal para HMM – Primeiro Método

Seja o modelo HMM descrito pelos parâmetros $\lambda = (A, B, \pi)$, onde:

$$\begin{aligned} A = \{a_{ij}\} & \quad \text{matriz de transição da MC}^1 \text{ oculta;} \\ B = \{b_i[k]\} & \quad \text{matriz de emissão de sinais;} \\ \pi = \{\pi_i\} & \quad \text{vetor de distribuição de probabilidade inicial da MC oculta.} \end{aligned}$$

Seja $\mathcal{S} = \{s_i\}$ o conjunto de estados da MC oculta e seja $\mathcal{Y} = \{y_k\}$ o conjunto de sinais. A matriz A é quadrada de ordem $n_A = |\mathcal{S}|$ e a matriz B tem tamanho $n_A \times n_B$, onde $n_B = |\mathcal{Y}|$. Essas matrizes têm seus elementos definidos como:

$$\begin{aligned} a_{ij} = P(s_j | s_i) & \quad \text{probabilidade da transição}^2 \text{ do estado } s_i \text{ para o estado } s_j; \\ b_i[k] = P(y_k | s_i) & \quad \text{probabilidade da emissão do sinal } y_k \text{ quando a MC ocupa o estado } s_i. \end{aligned}$$

Seja

$$P(s_j, y_k | s_i) = a_{ij} b_i[k] \quad \text{probabilidade de emitir } y_k \text{ e sofrer transição de } s_i \text{ para } s_j.$$

Pode-se organizar os valores $P(s_j, y_k | s_i)$ em uma matriz de tamanho $n_A \times (n_A n_B)$. Visando facilitar a compreensão, suponha $n_A = 2$ e $n_B = 3$. Essa matriz tem a seguinte aparência:

¹Cadeia de Markov (*Markov Chain* – MC)

²Este trabalho tem foco em cadeias de Markov homogêneas, onde vale a relação $P(X_{t+1} = s_j | X_t = s_i) = P(X_1 = s_j | X_0 = s_i) = P(s_j | s_i)$.

$$P' = \begin{array}{c|cccccc} & s_1, y_1 & s_1, y_2 & s_1, y_3 & s_2, y_1 & s_2, y_2 & s_2, y_3 \\ \hline s_1 & a_{11}b_1[1] & a_{11}b_1[2] & a_{11}b_1[3] & a_{12}b_1[1] & a_{12}b_1[2] & a_{12}b_1[3] \\ s_2 & a_{21}b_2[1] & a_{21}b_2[2] & a_{21}b_2[3] & a_{22}b_2[1] & a_{22}b_2[2] & a_{22}b_2[3] \end{array}$$

Como visto, a matriz P' tem seus elementos definidos pela probabilidade condicional $P(s_j, y_l | s_i)$, que fornece a probabilidade de emitir o sinal y_l e sofrer transição para o estado s_j , visto que o estado anterior era s_i . Nota-se que a emissão anterior não é considerada no cálculo, então, ao assumir explicitamente a independência entre emissões, pode-se escrever:

$$P(s_j, y_l | s_i, y_k) = P(s_j, y_l | s_i) = a_{ij}b_i[l]$$

Dessa forma, obtém-se a matriz quadrada P , de ordem $n_A n_B$:

$$P = \begin{array}{c|cccccc} & s_1, y_1 & s_1, y_2 & s_1, y_3 & s_2, y_1 & s_2, y_2 & s_2, y_3 \\ \hline s_1, y_1 & a_{11}b_1[1] & a_{11}b_1[2] & a_{11}b_1[3] & a_{12}b_1[1] & a_{12}b_1[2] & a_{12}b_1[3] \\ s_1, y_2 & " & " & " & " & " & " \\ s_1, y_3 & " & " & " & " & " & " \\ s_2, y_1 & a_{21}b_2[1] & a_{21}b_2[2] & a_{21}b_2[3] & a_{22}b_2[1] & a_{22}b_2[2] & a_{22}b_2[3] \\ s_2, y_2 & " & " & " & " & " & " \\ s_2, y_3 & " & " & " & " & " & " \end{array}$$

onde as aspas duplas indicam que o valor da célula é idêntico ao daquela acima. Como exemplo, considere a célula p_{35} , localizada no cruzamento da terceira linha com a quinta coluna. Seu valor é o mesmo que o das células p_{15} e p_{25} , significando que não interessa se o último sinal emitido foi y_1 , y_2 ou y_3 , pois o que importa é que a transição parte de s_1 para s_2 com a emissão do sinal y_2 , emitido em função do estado de partida, s_1 .

Qualquer que seja a ordem de A e o tamanho de B a matriz P é uma matriz estocástica de ordem $n_A n_B$, podendo ser vista como a cadeia de Markov em escala de tempo discreta que corresponde ao modelo HMM em questão. Para obter o gerador infinitesimal correspondente, Q , basta que se subtraia de P a identidade (Seção 2.7, p.36):

$$Q = P - I$$

$$Q = \begin{pmatrix} a_{11}b_1[1] - 1 & a_{11}b_1[2] & a_{11}b_1[3] & a_{12}b_1[1] & a_{12}b_1[2] & a_{12}b_1[3] \\ a_{11}b_1[1] & a_{11}b_1[2] - 1 & a_{11}b_1[3] & a_{12}b_1[1] & a_{12}b_1[2] & a_{12}b_1[3] \\ a_{11}b_1[1] & a_{11}b_1[2] & a_{11}b_1[3] - 1 & a_{12}b_1[1] & a_{12}b_1[2] & a_{12}b_1[3] \\ a_{21}b_2[1] & a_{21}b_2[2] & a_{21}b_2[3] & a_{22}b_2[1] - 1 & a_{22}b_2[2] & a_{22}b_2[3] \\ a_{21}b_2[1] & a_{21}b_2[2] & a_{21}b_2[3] & a_{22}b_2[1] & a_{22}b_2[2] - 1 & a_{22}b_2[3] \\ a_{21}b_2[1] & a_{21}b_2[2] & a_{21}b_2[3] & a_{22}b_2[1] & a_{22}b_2[2] & a_{22}b_2[3] - 1 \end{pmatrix}$$

5.2 Obtenção do Gerador Infinitesimal para HMM – Segundo Método

Seja o modelo HMM descrito pelos parâmetros $\lambda = (A, B, \pi)$, onde:

$$\begin{aligned} A = \{a_{ij}\} & \quad \text{matriz de transição da MC oculta;} \\ B = \{b_i[k]\} & \quad \text{matriz de emissão de sinais;} \\ \pi = \{\pi_i\} & \quad \text{vetor de distribuição de probabilidade inicial da MC oculta.} \end{aligned}$$

Seja $\mathcal{S} = \{s_i\}$ o conjunto de estados da MC oculta e seja $\mathcal{Y} = \{y_k\}$ o conjunto de sinais. A matriz A é quadrada de ordem $n_A = |\mathcal{S}|$ e a matriz B tem tamanho $n_A \times n_B$, onde $n_B = |\mathcal{Y}|$. Essas matrizes têm seus elementos definidos como:

$$\begin{aligned} a_{ij} = P(s_j | s_i) & \quad \text{probabilidade da transição}^3 \text{ do estado } s_i \text{ para o estado } s_j; \\ b_i[k] = P(y_k | s_i) & \quad \text{probabilidade da emissão do sinal } y_k \text{ quando a MC ocupa o estado } s_i. \end{aligned}$$

O método apresentado nesta seção mostra como obter o gerador infinitesimal correspondente ao modelo HMM a partir de seus parâmetros A e B . Para isso é necessário que esses parâmetros sejam transformados em matrizes de probabilidade de transição que, por definição, são matrizes quadradas. As seções seguintes apresentam os procedimentos para realizar essas transformações.

5.2.1 Método para Obtenção de uma Matriz de Probabilidade de Transição entre Estados

Nesta seção é feita a operação trivial de transformação do parâmetro de entrada A de um modelo HMM na matriz \hat{A} , que será usada mais adiante. Essa transformação está aqui pelo rigor formal, uma vez que, como será visto, $\hat{A} = A$.

Sejam

$$\begin{aligned} \mathcal{S} & \quad \text{conjunto finito de estados da cadeia de Markov;} \\ |\mathcal{S}| & \quad \text{cardinalidade de } \mathcal{S}. \end{aligned}$$

Definição 5.2.1. A função T , definida por

$$T : \mathcal{S} \times \mathcal{S} \rightarrow [0, 1],$$

é chamada função de transição e associa uma probabilidade de ocorrência no intervalo $[0, 1]$ a cada transição entre estados do conjunto \mathcal{S} .

³Este trabalho tem foco em cadeias de Markov homogêneas, onde vale a relação $P(X_{t+1} = s_j | X_t = s_i) = P(X_1 = s_j | X_0 = s_i) = P(s_j | s_i)$.

Sejam

- $T(i, j)$ probabilidade de transição do estado s_i para o estado s_j , onde $s_i, s_j \in \mathcal{S}$;
- \hat{A} matriz de probabilidade de transição entre estados de uma DTMC⁴, cuja ordem é $n_{\hat{A}} = |\mathcal{S}|$;
- \hat{a}_{ij} elemento da matriz \hat{A} posicionado na intersecção entre a linha que corresponde ao estado s_i e a coluna que corresponde ao estado s_j .

Definição 5.2.2. A matriz de probabilidade de transição entre os estados de uma DTMC, \hat{A} , tem seus elementos definidos por:

$$\hat{a}_{ij} = T(i, j), \quad \forall s_i, s_j \in \mathcal{S}$$

Como a matriz $A = \{a_{ij}\}$, $s_i, s_j \in \mathcal{S}$, do modelo HMM carrega justamente as probabilidades de transição entre os estados da cadeia de Markov que está oculta, segue que

$$\hat{a}_{ij} = a_{ij} \quad \forall s_i, s_j \in \mathcal{S}$$

Para facilitar a compreensão, considere o caso em que $|\mathcal{S}| = 2$. As matrizes A e \hat{A} têm a seguinte aparência:

$$A = \{a_{ij}\} = \begin{matrix} & \begin{matrix} s_1 & s_2 \end{matrix} \\ \begin{matrix} s_1 \\ s_2 \end{matrix} & \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \end{matrix}; \quad \hat{A} = \{\hat{a}_{ij}\} = \begin{matrix} & \begin{matrix} s_1 & s_2 \end{matrix} \\ \begin{matrix} s_1 \\ s_2 \end{matrix} & \begin{pmatrix} \hat{a}_{11} & \hat{a}_{12} \\ \hat{a}_{21} & \hat{a}_{22} \end{pmatrix} \end{matrix} \longrightarrow \hat{A} = A$$

5.2.2 Método para Obtenção de uma Matriz de Probabilidade de Transição entre Sinais

Nesta seção é feita a transformação do parâmetro de entrada B de um modelo HMM na matriz $\hat{B}(\hat{A})$, que será usada mais adiante. Esta transformação é necessária para obter uma matriz de transição entre sinais, que por definição deve ser quadrada. Como resultado, $\hat{B}(\hat{A})$ é uma matriz cujos elementos são todos funcionais e dependentes da matriz \hat{A} , obtida na Seção 5.2.1.

Sejam

- \mathcal{Y} conjunto de sinais;
- $|\mathcal{Y}|$ cardinalidade de \mathcal{Y} .

⁴Cadeia de Markov em Tempo Discreto (*Discrete-Time Markov Chain* – DTMC)

Definição 5.2.3. A função E , definida por

$$E : \mathcal{S} \times \mathcal{Y} \rightarrow [0, 1],$$

é chamada função de emissão e associa uma probabilidade de emissão no intervalo $[0, 1]$ a cada par (s_i, y_k) da HMM, onde $s_i \in \mathcal{S}$ e $y_k \in \mathcal{Y}$.

Sejam

- $E(i, k)$ probabilidade de emissão do sinal y_k visto que a MC oculta está no estado s_i ;
- $\hat{B}(\hat{A})$ matriz de probabilidade de transição entre os sinais emitidos em uma HMM, cuja ordem é $n_{\hat{B}} = |\mathcal{Y}|$. Note a dependência⁵ explícita sobre \hat{A} ;
- $\hat{b}_{kl}(\hat{a}_i)$ elemento da matriz $\hat{B}(\hat{A})$ posicionado na intersecção entre a linha que corresponde ao sinal y_k e a coluna que corresponde ao sinal y_l , onde $y_k, y_l \in \mathcal{Y}$.

Definição 5.2.4. A matriz de probabilidade de transição entre os sinais emitidos em uma HMM, $\hat{B}(\hat{A})$, tem seus elementos definidos por:

$$\hat{b}_{kl}(\hat{a}_i) = E(i, l), \quad \forall y_k, y_l \in \mathcal{Y}, \quad \forall s_i \in \mathcal{S}$$

Assim, os elementos de $\hat{B}(\hat{A})$ são todos funcionais, evidenciando a relação de dependência existente no modelo HMM, onde as emissões acontecem em função dos estados da MC oculta.

Como a matriz $B = \{b_i[k]\}$, $s_i \in \mathcal{S}$ e $y_k \in \mathcal{Y}$, do modelo HMM carrega justamente as probabilidades de emissão de sinais em função do estado ocupado pela cadeia de Markov que está oculta, segue que

$$\hat{b}_{kl}(\hat{a}_i) = b_i[l], \quad \forall y_k, y_l \in \mathcal{Y}, \quad \forall s_i \in \mathcal{S}$$

Dessa forma, supondo que o conjunto de sinais tenha cardinalidade $|\mathcal{Y}| = 3$, as matrizes B e $\hat{B}(\hat{A})$ têm a seguinte aparência:

$$B = \{b_i[k]\} = \begin{matrix} & \begin{matrix} y_1 & y_2 & y_3 \end{matrix} \\ \begin{matrix} s_1 \\ s_2 \\ \vdots \\ s_N \end{matrix} & \begin{pmatrix} b_1[1] & b_1[2] & b_1[3] \\ b_2[1] & b_2[2] & b_2[3] \\ \vdots & \vdots & \vdots \\ b_N[1] & b_N[2] & b_N[3] \end{pmatrix} \end{matrix}; \quad \hat{B}(\hat{A}) = \{\hat{b}_{kl}(\hat{a}_i)\} = \begin{matrix} & \begin{matrix} y_1 & y_2 & y_3 \end{matrix} \\ \begin{matrix} y_1 \\ y_2 \\ y_3 \end{matrix} & \begin{pmatrix} \hat{b}_{11}(\hat{a}_i) & \hat{b}_{12}(\hat{a}_i) & \hat{b}_{13}(\hat{a}_i) \\ \hat{b}_{21}(\hat{a}_i) & \hat{b}_{22}(\hat{a}_i) & \hat{b}_{23}(\hat{a}_i) \\ \hat{b}_{31}(\hat{a}_i) & \hat{b}_{32}(\hat{a}_i) & \hat{b}_{33}(\hat{a}_i) \end{pmatrix} \end{matrix}$$

⁵Esta notação é própria da Álgebra Tensorial Generalizada, introduzida em [16]. Para mais detalhes, consulte o Apêndice A.

$$\hat{B}(\hat{A}) = \{b_i[l]\} = \begin{matrix} & y_1 & y_2 & y_3 \\ y_1 & \begin{pmatrix} b_i[1] & b_i[2] & b_i[3] \end{pmatrix} \\ y_2 & \begin{pmatrix} b_i[1] & b_i[2] & b_i[3] \end{pmatrix} \\ y_3 & \begin{pmatrix} b_i[1] & b_i[2] & b_i[3] \end{pmatrix} \end{matrix}$$

$$b_i[1] = \begin{cases} b_1[1] & \text{para } s_i = s_1 \\ b_2[1] & \text{para } s_i = s_2 \\ \vdots & \\ b_N[1] & \text{para } s_i = s_N \end{cases} \quad b_i[2] = \begin{cases} b_1[2] & \text{para } s_i = s_1 \\ b_2[2] & \text{para } s_i = s_2 \\ \vdots & \\ b_N[2] & \text{para } s_i = s_N \end{cases} \quad b_i[3] = \begin{cases} b_1[3] & \text{para } s_i = s_1 \\ b_2[3] & \text{para } s_i = s_2 \\ \vdots & \\ b_N[3] & \text{para } s_i = s_N \end{cases}$$

onde a cardinalidade do espaço de estados \mathcal{S} é $|\mathcal{S}| = N$.

5.2.3 Método para Obtenção do Gerador Infinitesimal da HMM

Aqui conclui-se a exposição do método da Seção 5.2. A obtenção do gerador infinitesimal para HMM é um passo importante pois serve de termo de comparação na prova da equivalência entre os formalismos HMM e SAN. Para melhor compreender o que segue, recomenda-se a leitura do Apêndice A, onde são apresentados conceitos fundamentais da Álgebra Tensorial.

Sejam

- \hat{A} matriz de probabilidade de transição entre estados da MC oculta;
- $\hat{B}(\hat{A})$ matriz de probabilidade de transição entre os sinais emitidos pelo modelo HMM;
- Q gerador infinitesimal correspondente ao modelo HMM.

Dadas as características das matrizes \hat{A} e $\hat{B}(\hat{A})$, vistas nas Seções 5.2.1 e 5.2.2, com o auxílio da Álgebra Tensorial Generalizada a matriz de transição global pode ser obtida pela expressão tensorial:

$$P = \hat{A} \otimes_g \hat{B}(\hat{A}) \quad (5.1)$$

A Equação 5.1 fornece a DTMC que corresponde ao modelo HMM. Para obter o gerador infinitesimal, basta que se subtraia a identidade. Assim, o gerador infinitesimal correspondente ao modelo HMM é obtido pela expressão tensorial:

$$Q = \hat{A} \otimes_g \hat{B}(\hat{A}) - I_{(n_{\hat{A}} n_{\hat{B}})}, \quad (5.2)$$

onde $I_{(n_{\hat{A}} n_{\hat{B}})}$ é a matriz identidade de ordem $n_{\hat{A}} n_{\hat{B}}$.

As cardinalidades de \mathcal{S} e \mathcal{Y} podem ser quaisquer. Para exemplificar, suponha um modelo HMM onde $|\mathcal{S}| = 2$ e $|\mathcal{Y}| = 3$. As matrizes \hat{A} e $\hat{B}(\hat{A})$ têm a seguinte aparência:

$$A = \{a_{ij}\} = \begin{matrix} & \begin{matrix} s_1 & s_2 \end{matrix} \\ \begin{matrix} s_1 \\ s_2 \end{matrix} & \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \end{matrix} \quad \longrightarrow \quad \hat{A} = \{\hat{a}_{ij}\} = \begin{matrix} & \begin{matrix} s_1 & s_2 \end{matrix} \\ \begin{matrix} s_1 \\ s_2 \end{matrix} & \begin{pmatrix} \hat{a}_{11} & \hat{a}_{12} \\ \hat{a}_{21} & \hat{a}_{22} \end{pmatrix} \end{matrix}$$

$$B = \{b_i[k]\} = \begin{matrix} & \begin{matrix} y_1 & y_2 & y_3 \end{matrix} \\ \begin{matrix} s_1 \\ s_2 \end{matrix} & \begin{pmatrix} b_1[1] & b_1[2] & b_1[3] \\ b_2[1] & b_2[2] & b_2[3] \end{pmatrix} \end{matrix} \quad \longrightarrow \quad \hat{B}(\hat{A}) = \{\hat{b}_{kl}(\hat{a}_i)\} = \begin{matrix} & \begin{matrix} y_1 & y_2 & y_3 \end{matrix} \\ \begin{matrix} y_1 \\ y_2 \\ y_3 \end{matrix} & \begin{pmatrix} \hat{b}_{11}(\hat{a}_i) & \hat{b}_{12}(\hat{a}_i) & \hat{b}_{13}(\hat{a}_i) \\ \hat{b}_{21}(\hat{a}_i) & \hat{b}_{22}(\hat{a}_i) & \hat{b}_{23}(\hat{a}_i) \\ \hat{b}_{31}(\hat{a}_i) & \hat{b}_{32}(\hat{a}_i) & \hat{b}_{33}(\hat{a}_i) \end{pmatrix} \end{matrix}$$

Após resolver a Equação 5.1, tem-se:

$$P = \left(\begin{array}{ccc|ccc} \hat{a}_{11}\hat{b}_{11}(\hat{a}_i) & \hat{a}_{11}\hat{b}_{12}(\hat{a}_i) & \hat{a}_{11}\hat{b}_{13}(\hat{a}_i) & \hat{a}_{12}\hat{b}_{11}(\hat{a}_i) & \hat{a}_{12}\hat{b}_{12}(\hat{a}_i) & \hat{a}_{12}\hat{b}_{13}(\hat{a}_i) \\ \hat{a}_{11}\hat{b}_{21}(\hat{a}_i) & \hat{a}_{11}\hat{b}_{22}(\hat{a}_i) & \hat{a}_{11}\hat{b}_{23}(\hat{a}_i) & \hat{a}_{12}\hat{b}_{21}(\hat{a}_i) & \hat{a}_{12}\hat{b}_{22}(\hat{a}_i) & \hat{a}_{12}\hat{b}_{23}(\hat{a}_i) \\ \hat{a}_{11}\hat{b}_{31}(\hat{a}_i) & \hat{a}_{11}\hat{b}_{32}(\hat{a}_i) & \hat{a}_{11}\hat{b}_{33}(\hat{a}_i) & \hat{a}_{12}\hat{b}_{31}(\hat{a}_i) & \hat{a}_{12}\hat{b}_{32}(\hat{a}_i) & \hat{a}_{12}\hat{b}_{33}(\hat{a}_i) \\ \hline \hat{a}_{21}\hat{b}_{11}(\hat{a}_i) & \hat{a}_{21}\hat{b}_{12}(\hat{a}_i) & \hat{a}_{21}\hat{b}_{13}(\hat{a}_i) & \hat{a}_{22}\hat{b}_{11}(\hat{a}_i) & \hat{a}_{22}\hat{b}_{12}(\hat{a}_i) & \hat{a}_{22}\hat{b}_{13}(\hat{a}_i) \\ \hat{a}_{21}\hat{b}_{21}(\hat{a}_i) & \hat{a}_{21}\hat{b}_{22}(\hat{a}_i) & \hat{a}_{21}\hat{b}_{23}(\hat{a}_i) & \hat{a}_{22}\hat{b}_{21}(\hat{a}_i) & \hat{a}_{22}\hat{b}_{22}(\hat{a}_i) & \hat{a}_{22}\hat{b}_{23}(\hat{a}_i) \\ \hat{a}_{21}\hat{b}_{31}(\hat{a}_i) & \hat{a}_{21}\hat{b}_{32}(\hat{a}_i) & \hat{a}_{21}\hat{b}_{33}(\hat{a}_i) & \hat{a}_{22}\hat{b}_{31}(\hat{a}_i) & \hat{a}_{22}\hat{b}_{32}(\hat{a}_i) & \hat{a}_{22}\hat{b}_{33}(\hat{a}_i) \end{array} \right)$$

Visto que \hat{A} e $\hat{B}(\hat{A})$ são como seguem, após avaliar as funções em P , tem-se:

$$\hat{A} = A; \quad \hat{B}(\hat{A}) = \{\hat{b}_{kl}(\hat{a}_i)\} = \{b_i[l]\} = \begin{matrix} & \begin{matrix} y_1 & y_2 & y_3 \end{matrix} \\ \begin{matrix} y_1 \\ y_2 \\ y_3 \end{matrix} & \begin{pmatrix} b_i[1] & b_i[2] & b_i[3] \\ b_i[1] & b_i[2] & b_i[3] \\ b_i[1] & b_i[2] & b_i[3] \end{pmatrix} \end{matrix}$$

$$b_i[1] = \begin{cases} b_1[1] & \text{para } s_i = s_1 \\ b_2[1] & \text{para } s_i = s_2 \end{cases} \quad b_i[2] = \begin{cases} b_1[2] & \text{para } s_i = s_1 \\ b_2[2] & \text{para } s_i = s_2 \end{cases} \quad b_i[3] = \begin{cases} b_1[3] & \text{para } s_i = s_1 \\ b_2[3] & \text{para } s_i = s_2 \end{cases}$$

$$P = \left(\begin{array}{ccc|ccc} a_{11}b_1[1] & a_{11}b_1[2] & a_{11}b_1[3] & a_{12}b_1[1] & a_{12}b_1[2] & a_{12}b_1[3] \\ a_{11}b_1[1] & a_{11}b_1[2] & a_{11}b_1[3] & a_{12}b_1[1] & a_{12}b_1[2] & a_{12}b_1[3] \\ a_{11}b_1[1] & a_{11}b_1[2] & a_{11}b_1[3] & a_{12}b_1[1] & a_{12}b_1[2] & a_{12}b_1[3] \\ \hline a_{21}b_2[1] & a_{21}b_2[2] & a_{21}b_2[3] & a_{22}b_2[1] & a_{22}b_2[2] & a_{22}b_2[3] \\ a_{21}b_2[1] & a_{21}b_2[2] & a_{21}b_2[3] & a_{22}b_2[1] & a_{22}b_2[2] & a_{22}b_2[3] \\ a_{21}b_2[1] & a_{21}b_2[2] & a_{21}b_2[3] & a_{22}b_2[1] & a_{22}b_2[2] & a_{22}b_2[3] \end{array} \right)$$

A matriz recém obtida é idêntica àquela obtida na Seção 5.1 (Página 75), representando a matriz de probabilidade de transição correspondente ao modelo HMM. Para transformá-la no gerador infinitesimal correspondente, basta que se subtraia a identidade, como visto na Equação 5.2:

$$Q = \left(\begin{array}{ccc|ccc} a_{11}b_1[1] - 1 & a_{11}b_1[2] & a_{11}b_1[3] & a_{12}b_1[1] & a_{12}b_1[2] & a_{12}b_1[3] \\ a_{11}b_1[1] & a_{11}b_1[2] - 1 & a_{11}b_1[3] & a_{12}b_1[1] & a_{12}b_1[2] & a_{12}b_1[3] \\ a_{11}b_1[1] & a_{11}b_1[2] & a_{11}b_1[3] - 1 & a_{12}b_1[1] & a_{12}b_1[2] & a_{12}b_1[3] \\ \hline a_{21}b_2[1] & a_{21}b_2[2] & a_{21}b_2[3] & a_{22}b_2[1] - 1 & a_{22}b_2[2] & a_{22}b_2[3] \\ a_{21}b_2[1] & a_{21}b_2[2] & a_{21}b_2[3] & a_{22}b_2[1] & a_{22}b_2[2] - 1 & a_{22}b_2[3] \\ a_{21}b_2[1] & a_{21}b_2[2] & a_{21}b_2[3] & a_{22}b_2[1] & a_{22}b_2[2] & a_{22}b_2[3] - 1 \end{array} \right)$$

Assim prova-se que o método aqui apresentado é equivalente àquele da Seção 5.1 (Página 75). Portanto, qualquer dos métodos pode ser usado na obtenção do gerador infinitesimal correspondente ao modelo HMM, tendo o método desta seção a vantagem da representação tensorial (Equação 5.2).

5.3 Método de Conversão de Modelos HMM em SAN

Seja o modelo HMM descrito pelos parâmetros $\lambda = (A, B, \pi)$, onde:

$$\begin{aligned} A = \{a_{ij}\} & \quad \text{matriz de transição da MC oculta;} \\ B = \{b_i[k]\} & \quad \text{matriz de emissão de sinais;} \\ \pi = \{\pi_i\} & \quad \text{vetor de distribuição de probabilidade inicial da MC oculta.} \end{aligned}$$

Seja $\mathcal{S} = \{s_i\}$ o conjunto de estados da MC oculta e seja $\mathcal{Y} = \{y_k\}$ o conjunto de sinais. A matriz A é quadrada de ordem $n_A = |\mathcal{S}|$ e a matriz B tem tamanho $n_A \times n_B$, onde $n_B = |\mathcal{Y}|$. Essas matrizes têm seus elementos definidos como:

$$\begin{aligned} a_{ij} &= P(s_j | s_i) \quad \text{probabilidade da transição}^6 \text{ do estado } s_i \text{ para o estado } s_j; \\ b_i[k] &= P(y_k | s_i) \quad \text{probabilidade da emissão do sinal } y_k \text{ quando a MC ocupa o estado } s_i. \end{aligned}$$

O método de conversão de HMM em SAN aqui apresentado gera um modelo SAN composto de dois autômatos e um único evento que sincroniza todas as transições do modelo.

Definição 5.3.1. *O conjunto de eventos é unitário $\mathcal{E} = \{e\}$. A esse evento associa-se uma tupla de evento, (e, μ) , onde:*

$$\begin{aligned} e & \quad \text{identificador do evento;} \\ \mu & \quad \text{taxa de ocorrência do evento.} \end{aligned}$$

⁶Este trabalho tem foco em cadeias de Markov homogêneas, onde vale a relação $P(X_{t+1} = s_j | X_t = s_i) = P(X_1 = s_j | X_0 = s_i) = P(s_j | s_i)$.

5.3.1 Método para Obtenção do Autômato Correspondente à Cadeia de Markov Oculta

O método de conversão de modelos HMM em SAN apresentado nesta dissertação consiste na obtenção de uma rede composta por dois autômatos estocásticos. Esta seção trata da obtenção do autômato que vai substituir a cadeia de Markov oculta que faz parte do modelo HMM original.

Sejam

\mathcal{S} conjunto finito de estados da cadeia de Markov;

$|\mathcal{S}|$ cardinalidade de \mathcal{S} .

Definição 5.3.2. A função T , definida por

$$T : \mathcal{S} \times \mathcal{S} \rightarrow [0, 1],$$

é chamada função de transição e associa uma probabilidade de ocorrência no intervalo $[0, 1]$ a cada transição entre estados do conjunto \mathcal{S} .

Sejam

$T(i, j)$ probabilidade de transição do estado s_i para o estado s_j , onde $s_i, s_j \in \mathcal{S}$;

\hat{A} matriz de probabilidade de transição entre estados de uma DTMC, cuja ordem é $n_{\hat{A}} = |\mathcal{S}|$;

\hat{a}_{ij} elemento da matriz \hat{A} posicionado na intersecção entre a linha que corresponde ao estado s_i e a coluna que corresponde ao estado s_j .

Definição 5.3.3. A matriz de probabilidade de transição entre os estados de uma DTMC, \hat{A} , tem seus elementos definidos por:

$$\hat{a}_{ij} = T(i, j), \quad \forall s_i, s_j \in \mathcal{S}$$

Como a matriz $A = \{a_{ij}\}$, $s_i, s_j \in \mathcal{S}$, do modelo HMM carrega justamente as probabilidades de transição entre os estados da cadeia de Markov que está oculta, segue que

$$\hat{a}_{ij} = a_{ij} \quad \forall s_i, s_j \in \mathcal{S}$$

Definição 5.3.4. A todo $s_i, s_j \in \mathcal{S}$ associa-se uma tupla de transição local *partindo de s_i para s_j* , notada por $(e, \pi_e(s_i, s_j))$, onde:

$$e \quad \text{identificador do evento;}$$

$$\pi_e(s_i, s_j) = \hat{a}_{ij} \quad \text{probabilidade de roteamento.}$$

Definição 5.3.5. O autômato \mathcal{A} é definido pelo conjunto de estados \mathcal{S} , matriz de transição entre os estados \hat{A} e correspondentes tuplas de transição local.

5.3.2 Método para Obtenção do Autômato Correspondente à Emissão de Sinais

O método de conversão de modelos HMM em SAN apresentado nesta dissertação consiste na obtenção de uma rede composta por dois autômatos estocásticos. Esta seção trata da obtenção do autômato que vai substituir as emissões produzidas pelo modelo HMM original.

Sejam

$$\mathcal{Y} \quad \text{conjunto de sinais;}$$

$$|\mathcal{Y}| \quad \text{cardinalidade de } \mathcal{Y}.$$

Definição 5.3.6. A função E , definida por

$$E : \mathcal{S} \times \mathcal{Y} \rightarrow [0, 1],$$

é chamada função de emissão e associa uma probabilidade de emissão no intervalo $[0, 1]$ a cada par (s_i, y_k) da HMM, onde $s_i \in \mathcal{S}$ e $y_k \in \mathcal{Y}$.

Sejam

$$E(i, k) \quad \text{probabilidade de emissão do sinal } y_k \text{ visto que a MC oculta está no estado } s_i;$$

$$\hat{B}(\hat{A}) \quad \text{matriz de probabilidade de transição entre os sinais emitidos em uma HMM, cuja ordem é } n_{\hat{B}} = |\mathcal{Y}|. \text{ Note a dependência}^7 \text{ explícita sobre } \hat{A};$$

$$\hat{b}_{kl}(\hat{a}_i) \quad \text{elemento da matriz } \hat{B}(\hat{A}) \text{ posicionado na intersecção entre a linha que corresponde ao sinal } y_k \text{ e a coluna que corresponde ao sinal } y_l, \text{ onde } y_k, y_l \in \mathcal{Y}.$$

⁷Esta notação é própria da Álgebra Tensorial Generalizada, introduzida em [16]. Para mais detalhes, consulte o Apêndice A.

Definição 5.3.7. A matriz de probabilidade de transição entre os sinais emitidos em uma HMM, $\hat{B}(\hat{A})$, tem seus elementos definidos por:

$$\hat{b}_{kl}(\hat{a}_i) = E(i, l), \quad \forall y_k, y_l \in \mathcal{Y}, \quad \forall s_i \in \mathcal{S}$$

Assim, os elementos de $\hat{B}(\hat{A})$ são todos funcionais, evidenciando a relação de dependência existente no modelo HMM, onde as emissões acontecem em função dos estados da MC oculta.

Como a matriz $B = \{b_i[k]\}$, $s_i \in \mathcal{S}$ e $y_k \in \mathcal{Y}$, do modelo HMM carrega justamente as probabilidades de emissão de sinais em função do estado ocupado pela cadeia de Markov que está oculta, segue que

$$\hat{b}_{kl}(\hat{a}_i) = b_i[l], \quad \forall y_k, y_l \in \mathcal{Y}, \quad \forall s_i \in \mathcal{S}$$

Definição 5.3.8. A todo $y_k, y_l \in \mathcal{Y}$ associa-se uma tupla de transição local *partindo de y_k para y_l* , notada por $(e, \pi_e(y_k, y_l))$, onde:

e identificador do evento;

$\pi_e(y_k, y_l) = \hat{b}_{kl}(\hat{a}_i)$ probabilidade de roteamento.

Definição 5.3.9. O autômato \mathcal{B} é definido pelo conjunto de emissões \mathcal{Y} , matriz de transição entre as emissões $\hat{B}(\hat{A})$ e correspondentes tuplas de transição local.

Na notação da ferramenta PEPS⁸ [11], a dependência funcional do autômato \mathcal{B} sobre o autômato \mathcal{A} pode ser reescrita como segue:

$$\pi_e(y_k, y_l) = \hat{b}_{kl}(\hat{a}_i) = b_i[l] = \sum_i (st \mathcal{A} == s_i) * b_i[l], \quad \forall y_k, y_l \in \mathcal{Y}, \quad \forall s_i \in \mathcal{S}$$

Neste capítulo, a probabilidade de roteamento em \mathcal{B} é identificada pela notação:

$$\begin{aligned} \pi_e(y_k, y_l) &= _y_l, & \forall y_k, y_l \in \mathcal{Y} \\ _y_l &= \sum_i (st \mathcal{A} == s_i) * b_i[l], & \forall y_k, y_l \in \mathcal{Y}, \quad \forall s_i \in \mathcal{S} \end{aligned}$$

onde $_y_l$ é a probabilidade de roteamento para y_l , não importando o sinal y_k de origem.

⁸PEPS é uma ferramenta para solução de modelos SAN.

5.3.3 Representação Gráfica do Modelo SAN Obtido a partir de um Modelo HMM

Esta seção mostra a representação gráfica do modelo SAN formado pelos dois autômatos criados nas Seções 5.3.1 e 5.3.2. Do ponto de vista formal, este passo não é obrigatório, mas auxilia na compreensão.

Sejam

- $\lambda = (A, B, \pi)$ modelo HMM;
- \mathcal{A} autômato que representa a MC oculta do modelo HMM;
- \mathcal{B} autômato que representa as emissões do modelo HMM.

A Figura 5.1 mostra a representação gráfica de uma SAN genérica com $|\mathcal{S}| = 2$ e $|\mathcal{Y}| = 3$:

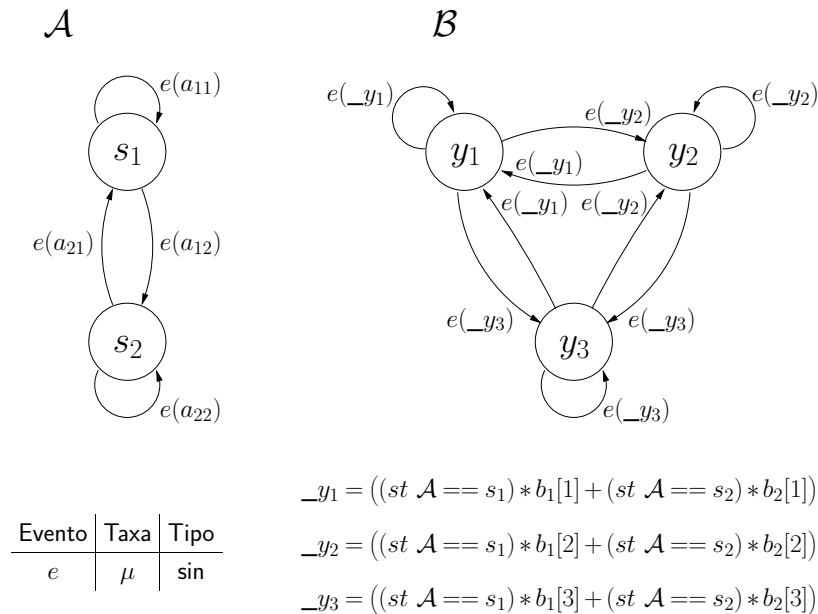


Figura 5.1 – Modelo SAN para uma HMM Genérica com $|\mathcal{S}| = 2$ e $|\mathcal{Y}| = 3$

5.3.4 Obtenção do Descritor Markoviano

Ao autômato \mathcal{A} são associadas:

- uma matriz de transição $Q_{e+}^{\mathcal{A}}$ correspondendo à *parte positiva* do evento sincronizante nesse autômato e suas respectivas tuplas de transição;
- uma matriz de transição $Q_{e-}^{\mathcal{A}}$ correspondendo à *parte negativa* do evento sincronizante nesse autômato e suas respectivas tuplas de transição.

Seja

$Q_x^A(i, j)$ o elemento do tensor Q_x^A na linha i e na coluna j , onde $x \in \{e^+, e^-\}$.

Definição 5.3.10. Os elementos da matriz de transição $Q_{e^+}^A$, referente à ocorrência do evento sincronizante $e \in \mathcal{E}$ no autômato \mathcal{A} , são definidos por:

- $\forall s_i, s_j \in \mathcal{S}$

$$Q_{e^+}^A(i, j) = \mu(\pi_e(s_i, s_j)) = \mu(\hat{a}_{ij}),$$

de forma que $Q_{e^+}^A = \mu \times \hat{A}$.

Definição 5.3.11. Os elementos da matriz $Q_{e^-}^A$, que representam o ajuste necessário à ocorrência do evento $e \in \mathcal{E}$ no autômato \mathcal{A} , são definidos por:

- $\forall s_i, s_j \in \mathcal{S}$

$$Q_{e^-}^A(i, i) = -\mu\left(\sum_{s_j} \pi_e(s_i, s_j)\right) = -\mu\left(\sum_j \hat{a}_{ij}\right);$$

- $\forall s_i, s_j \in \mathcal{S}$, tal que $s_i \neq s_j$

$$Q_{e^-}^A(i, j) = 0.$$

Visto que \hat{A} é uma matriz de probabilidade de transição, então o somatório sobre qualquer linha da matriz resulta em "1", de forma que $Q_{e^-}^A = -\mu \times I_{n_{\hat{A}}}$.

Ao autômato \mathcal{B} são associadas:

- uma matriz de transição $Q_{e^+}^B$ correspondendo à *parte positiva* do evento sincronizante nesse autômato e suas respectivas tuplas de transição;
- uma matriz de transição $Q_{e^-}^B$ correspondendo à *parte negativa* do evento sincronizante nesse autômato e suas respectivas tuplas de transição.

Seja

$Q_x^B(k, l)$ o elemento do tensor Q_x^B na linha k e na coluna l , onde $x \in \{e^+, e^-\}$.

Definição 5.3.12. Os elementos da matriz de transição $Q_{e^+}^B$, referente à ocorrência do evento sincronizante $e \in \mathcal{E}$ no autômato \mathcal{B} , são definidos por:

- $\forall y_k, y_l \in \mathcal{Y}$

$$Q_{e^+}^B(k, l) = \pi_e(y_k, y_l) = \hat{b}_{kl}(\hat{a}_i)$$

de forma que $Q_{e^+}^B = \hat{B}(\hat{A})$.

Definição 5.3.13. Os elementos da matriz $Q_{e^-}^{\mathcal{B}}$, que representam o ajuste necessário à ocorrência do evento $e \in \mathcal{E}$ no autômato \mathcal{B} , são definidos por:

- $\forall y_k, y_l \in \mathcal{Y}$

$$Q_{e^-}^{\mathcal{B}}(k, k) = \sum_{y_l} \pi_e(y_k, y_l) = \sum_l \hat{b}_{kl}(\hat{a}_i);$$

- $\forall y_k, y_l \in \mathcal{Y}$, tal que $y_k \neq y_l$

$$Q_{e^-}^{\mathcal{Y}}(k, l) = 0.$$

Visto que $\hat{B}(\hat{A})$ é uma matriz de probabilidade de transição, então o somatório sobre qualquer linha da matriz resulta em "1", de forma que $Q_{e^-}^{\mathcal{B}} = I_{n_{\hat{B}}}$.

Teorema 5.3.1. O descritor Markoviano do modelo SAN obtido pelo método de conversão de modelos HMM (Seção 5.3) é dado pela fórmula tensorial:

$$Q = \mu \left(\hat{A} \otimes_g \hat{B}(\hat{A}) - I_{n_{\hat{A}} n_{\hat{B}}} \right) \quad (5.3)$$

onde μ é a taxa associada ao evento e .

Prova 5.3.1. O método de conversão de HMM em SAN apresentado nesta seção gera um modelo SAN com dois autômatos, \mathcal{A} e \mathcal{B} , e um evento sincronizante, e . Na ausência de eventos locais, o descritor Markoviano desse modelo é dado pela expressão:

$$Q = Q_{e^+}^{\mathcal{A}} \otimes_g Q_{e^+}^{\mathcal{B}} + Q_{e^-}^{\mathcal{A}} \otimes_g Q_{e^-}^{\mathcal{B}}. \quad (5.4)$$

Pelas Definições 5.3.10 e 5.3.11, referentes ao autômato \mathcal{A} , e Definições 5.3.12 e 5.3.13, referentes ao autômato \mathcal{B} , a Equação 5.4 equivale à Equação 5.3:

$$Q = \mu \hat{A} \otimes_g \hat{B}(\hat{A}) - \mu I_{n_{\hat{A}}} \otimes_g I_{n_{\hat{B}}} \quad (5.5)$$

$$Q = \mu \left(\hat{A} \otimes_g \hat{B}(\hat{A}) - I_{n_{\hat{A}} n_{\hat{B}}} \right) \quad (5.6)$$

□

As Equações 5.3 e 5.2 (Página 80) são proporcionais, sendo a taxa do evento, μ , a própria constante de proporcionalidade. Se $\mu = 1$, então essas equações são idênticas. Geradores infinitesimais proporcionais possuem a mesma solução estacionária, ficando assim provada a equivalência entre o modelo HMM original e o modelo SAN criado segundo o método aqui apresentado.

6. EXEMPLOS DE CONVERSÃO DE HMM EM SAN

Neste capítulo são aplicados os conceitos desenvolvidos no Capítulo 5, onde dois métodos para obtenção do gerador infinitesimal de um modelo HMM (*Hidden Markov Models*) foram apresentados, além de um método de conversão de modelos HMM em modelos SAN (*Stochastic Automata Networks*). O capítulo é dividido em duas seções, cada uma dedicada a um exemplo: a Seção 6.1 mostra um modelo de máquina de refrigerante proposto por Manning e Schütze [24]; a Seção 6.2 é dedicada ao estudo de um exemplo adaptado da *Wikipedia* [19, 32], onde existem pares de estados sem transição prevista e onde os estados não emitem o mesmo conjunto de sinais.

6.1 Exemplo: *Crazy Soda Machine*

Seja uma máquina de refrigerantes abastecida com *Coca-Cola* e *Ice-Tea*. Toda vez que um cliente insere uma moeda, a máquina entrega um refrigerante e escolhe qual será o próximo a ser entregue quando uma nova moeda for inserida. A escolha é estocástica, dada pelo processo ilustrado na Figura 6.1, onde “CC” e “IT” representam *Coca-Cola* e *Ice-Tea*. Para exemplificar, suponha que o estado atual da máquina é “CC”, significando que o próximo refrigerante a ser entregue é uma *Coca-Cola*. Assim que o cliente inserir a moeda, a *Coca-Cola* será entregue e a máquina escolherá seu novo estado, que pode ser “CC” ou “IT”, com 70% e 30% de chance, respectivamente. Se o sorteio do novo estado resultar em “IT”, o próximo cliente receberá uma lata de *Ice-Tea*.

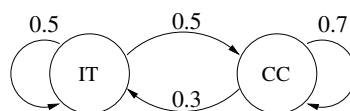


Figura 6.1 – Máquina de Refrigerantes – Modelagem em MC

O problema descreve uma *cadeia de Markov* pois, além de satisfazer a propriedade Markoviana e possuir espaço de estados discreto, o estado atual do sistema pode ser diretamente observado, ou seja, é possível verificar a sequência exata de estados ocupados durante o processo. Essa verificação se dá pela inserção de moedas, visto que a máquina sempre entrega o refrigerante que corresponde ao seu estado atual.

Considere agora uma máquina de refrigerantes diferente. Essa máquina, além de *Coca-Cola* e *Ice-Tea*, também entrega *Fanta*, e seus estados indicam apenas uma *tendência* na entrega ou *emissão* de refrigerantes. Essa tendência estabelece os novos estados dessa máquina: “PCC” e “PIT”, significando “*Preferência: Coca-Cola*” e “*Preferência: Ice-Tea*”. Note que não há um estado preferencial para a entrega de *Fanta*. A Figura 6.2 apresenta essa nova máquina, incluindo as probabilidades de emissão de refrigerante em função do estado preferencial em que ela se encontra. Assim, se a máquina está no estado “PIT” no momento em que o cliente insere a moeda, as probabilidades são

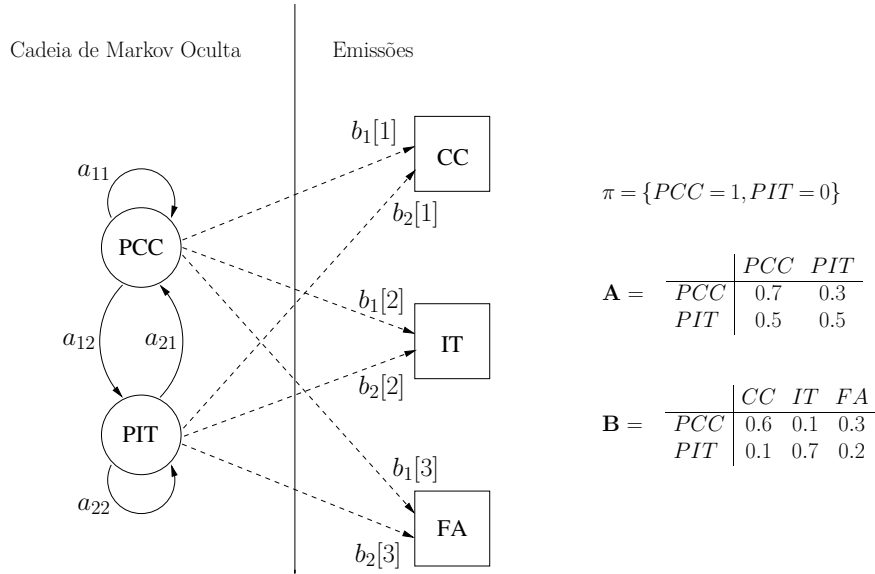


Figura 6.2 – Crazy Soda Machine – Modelagem em HMM

de 10%, 70% ou 20% de que seja entregue Coca-Cola, Ice-Tea ou Fanta, respectivamente.

Tão logo seja emitido um refrigerante, a máquina sorteia seu novo estado preferencial, “PCC” ou “PIT”, de acordo com as probabilidades de transição a partir do estado atual. Logo, estando a máquina no estado “PIT” no momento da emissão, seu próximo estado preferencial será “PCC” ou novamente “PIT” com iguais chances de escolha, indicadas por a_{21} e a_{22} .

Note que, diferentemente do exemplo anterior, referente à Figura 6.1, não é possível aqui ter certeza sobre a sequência de escolhas de estado preferencial nessa máquina a partir da sequência observada de emissões de refrigerantes, visto que aqui não há uma correspondência única entre o refrigerante entregue ao cliente e o estado preferencial em que a máquina se encontra.

Essa nova máquina, cujo modelo HMM é apresentado na Figura 6.2, foi introduzida por Manning e Schütze [24] em seu livro sobre linguagem natural. Para referências futuras, essa máquina recebe o nome de *Crazy Soda Machine*.

6.1.1 Obtenção do Gerador Infinitesimal do Modelo HMM da Figura 6.2

Um modelo HMM é definido pelos seus parâmetros, $\lambda = (A, B, \pi)$. A Figura 6.2 mostra a representação gráfica e os parâmetros de um modelo HMM para o exemplo *Crazy Soda Machine*, tal que:

$$A = \begin{matrix} & \text{PCC} & \text{PIT} \\ \text{PCC} & \begin{pmatrix} 0.7 & 0.3 \end{pmatrix} \\ \text{PIT} & \begin{pmatrix} 0.5 & 0.5 \end{pmatrix} \end{matrix}; \quad B = \begin{matrix} & \text{CC} & \text{IT} & \text{FA} \\ \text{PCC} & \begin{pmatrix} 0.6 & 0.1 & 0.3 \end{pmatrix} \\ \text{PIT} & \begin{pmatrix} 0.1 & 0.7 & 0.2 \end{pmatrix} \end{matrix}$$

Conhecendo o modelo HMM, pode-se obter o gerador infinitesimal correspondente usando qualquer dos dois métodos apresentados nas Seções 5.1 e 5.2.

Primeiro Método (Seção 5.1, p.75)

A partir das matrizes A e B obtém-se a matriz P' , a seguir, que indica a probabilidade conjunta da transição de um estado a outro acompanhada de uma emissão em função do estado de partida.

$$P' = \begin{array}{c|cccccc} & \text{PCC,CC} & \text{PCC,IT} & \text{PCC,FA} & \text{PIT,CC} & \text{PIT,IT} & \text{PIT,FA} \\ \hline \text{PCC} & 0.7*0.6 & 0.7*0.1 & 0.7*0.3 & 0.3*0.6 & 0.3*0.1 & 0.3*0.3 \\ \text{PIT} & 0.5*0.1 & 0.5*0.7 & 0.5*0.2 & 0.5*0.1 & 0.5*0.7 & 0.5*0.2 \end{array}$$

$$P' = \begin{array}{c|cccccc} & \text{PCC,CC} & \text{PCC,IT} & \text{PCC,FA} & \text{PIT,CC} & \text{PIT,IT} & \text{PIT,FA} \\ \hline \text{PCC} & 0.42 & 0.07 & 0.21 & 0.18 & 0.03 & 0.09 \\ \text{PIT} & 0.05 & 0.35 & 0.10 & 0.05 & 0.35 & 0.10 \end{array}$$

Dada a independência entre emissões consecutivas, a matriz P' é estendida, gerando a matriz P , onde os valores repetidos em uma mesma coluna explicitam essa independência, garantindo que as emissões só dependam do estado de partida.

$$P = \begin{array}{c|cccccc} & \text{PCC,CC} & \text{PCC,IT} & \text{PCC,FA} & \text{PIT,CC} & \text{PIT,IT} & \text{PIT,FA} \\ \hline \text{PCC,CC} & 0.42 & 0.07 & 0.21 & 0.18 & 0.03 & 0.09 \\ \text{PCC,IT} & 0.42 & 0.07 & 0.21 & 0.18 & 0.03 & 0.09 \\ \text{PCC,FA} & 0.42 & 0.07 & 0.21 & 0.18 & 0.03 & 0.09 \\ \text{PIT,CC} & 0.05 & 0.35 & 0.10 & 0.05 & 0.35 & 0.10 \\ \text{PIT,IT} & 0.05 & 0.35 & 0.10 & 0.05 & 0.35 & 0.10 \\ \text{PIT,FA} & 0.05 & 0.35 & 0.10 & 0.05 & 0.35 & 0.10 \end{array}$$

A matriz P é a matriz de probabilidade de transição correspondente ao modelo HMM. Para transformá-la no gerador infinitesimal Q , basta que dela se subtraia a identidade:

$$Q = P - I$$

$$Q = \begin{pmatrix} -0.58 & 0.07 & 0.21 & 0.18 & 0.03 & 0.09 \\ 0.42 & -0.93 & 0.21 & 0.18 & 0.03 & 0.09 \\ 0.42 & 0.07 & -0.79 & 0.18 & 0.03 & 0.09 \\ 0.05 & 0.35 & 0.10 & -0.95 & 0.35 & 0.10 \\ 0.05 & 0.35 & 0.10 & 0.05 & -0.65 & 0.10 \\ 0.05 & 0.35 & 0.10 & 0.05 & 0.35 & -0.90 \end{pmatrix}$$

Segundo Método (Seção 5.2, p.77)

A aplicação deste método requer a transformação dos parâmetros iniciais A e B do modelo HMM, como segue:

$$A = \begin{matrix} & \text{PCC} & \text{PIT} \\ \text{PCC} & \begin{pmatrix} 0.7 & 0.3 \end{pmatrix} \\ \text{PIT} & \begin{pmatrix} 0.5 & 0.5 \end{pmatrix} \end{matrix} \longrightarrow \hat{A} = A$$

$$B = \begin{matrix} & \text{CC} & \text{IT} & \text{FA} \\ \text{PCC} & \begin{pmatrix} 0.6 & 0.1 & 0.3 \end{pmatrix} \\ \text{PIT} & \begin{pmatrix} 0.1 & 0.7 & 0.2 \end{pmatrix} \end{matrix} \longrightarrow \hat{B}(\hat{A}) = \begin{matrix} & \text{CC} & \text{IT} & \text{FA} \\ \text{CC} & \begin{pmatrix} _CC & _IT & _FA \end{pmatrix} \\ \text{IT} & \begin{pmatrix} _CC & _IT & _FA \end{pmatrix} \\ \text{FA} & \begin{pmatrix} _CC & _IT & _FA \end{pmatrix} \end{matrix}$$

$$_CC = \begin{cases} 0.6 & \text{para } s_i = \text{PCC} \\ 0.1 & \text{para } s_i = \text{PIT} \end{cases} \quad _IT = \begin{cases} 0.1 & \text{para } s_i = \text{PCC} \\ 0.7 & \text{para } s_i = \text{PIT} \end{cases} \quad _FA = \begin{cases} 0.3 & \text{para } s_i = \text{PCC} \\ 0.2 & \text{para } s_i = \text{PIT} \end{cases}$$

Após obtidos \hat{A} e $\hat{B}(\hat{A})$, o gerador infinitesimal Q é calculado pela Equação 5.2 (Página 80):

$$Q = \hat{A} \otimes_g \hat{B}(\hat{A}) - I_{n_{\hat{A}} n_{\hat{B}}}$$

$$Q = \left(\begin{array}{ccc|ccc} 0.7*_CC - 1 & 0.7*_IT & 0.7*_FA & 0.3*_CC & 0.3*_IT & 0.3*_FA \\ 0.7*_CC & 0.7*_IT - 1 & 0.7*_FA & 0.3*_CC & 0.3*_IT & 0.3*_FA \\ 0.7*_CC & 0.7*_IT & 0.7*_FA - 1 & 0.3*_CC & 0.3*_IT & 0.3*_FA \\ \hline 0.5*_CC & 0.5*_IT & 0.5*_FA & 0.5*_CC - 1 & 0.5*_IT & 0.5*_FA \\ 0.5*_CC & 0.5*_IT & 0.5*_FA & 0.5*_CC & 0.5*_IT - 1 & 0.5*_FA \\ 0.5*_CC & 0.5*_IT & 0.5*_FA & 0.5*_CC & 0.5*_IT & 0.5*_FA - 1 \end{array} \right)$$

Após avaliar as funções $_CC$, $_IT$ e $_FA$ em respeito ao estado ocupado pela cadeia de Markov oculta, PCC ou PIT, tem-se o gerador infinitesimal:

$$Q = \begin{pmatrix} -0.58 & 0.07 & 0.21 & 0.18 & 0.03 & 0.09 \\ 0.42 & -0.93 & 0.21 & 0.18 & 0.03 & 0.09 \\ 0.42 & 0.07 & -0.79 & 0.18 & 0.03 & 0.09 \\ 0.05 & 0.35 & 0.10 & -0.95 & 0.35 & 0.10 \\ 0.05 & 0.35 & 0.10 & 0.05 & -0.65 & 0.10 \\ 0.05 & 0.35 & 0.10 & 0.05 & 0.35 & -0.90 \end{pmatrix}$$

6.1.2 Obtenção do Gerador Infinitesimal do Modelo SAN da Figura 6.3

Seja a modelagem HMM da Figura 6.2 para o problema *Crazy Soda Machine*. Pela aplicação do método apresentado na Seção 5.3 é possível converter esse modelo em um modelo SAN equivalente, cuja representação gráfica é vista na Figura 6.3.

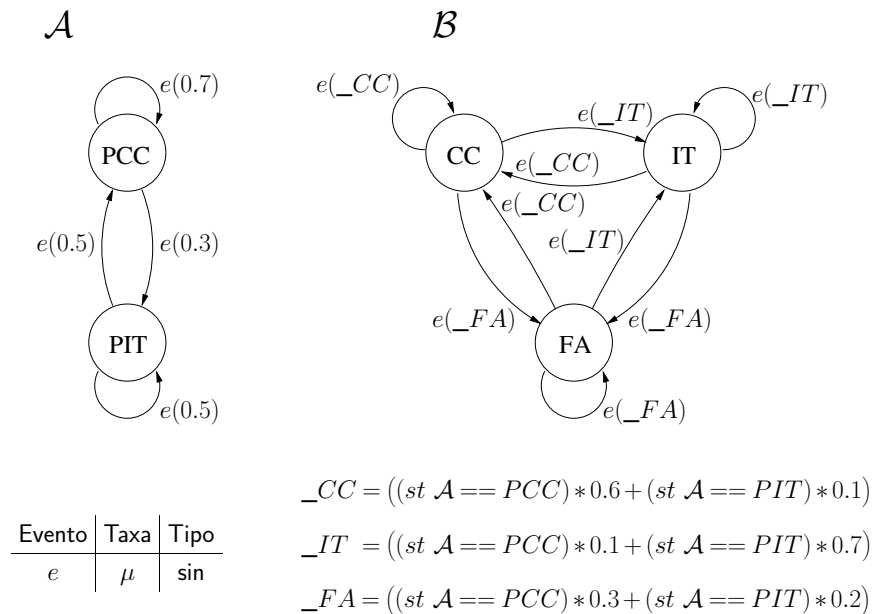


Figura 6.3 – *Crazy Soda Machine* – Modelagem em SAN

Esse método de conversão gera uma SAN de dois autômatos cuja dinâmica é regulada por um único evento sincronizante, e . O descritor Markoviano desse modelo é dado pela Equação 5.4 (Página 88):

$$Q = Q_{e^+}^A \otimes Q_{e^+}^B + Q_{e^-}^A \otimes Q_{e^-}^B \quad (6.1)$$

Os primeiros dois fatores a seguir compõem o termo tensorial concernente à parte positiva do evento e . Os outros dois fatores dizem respeito ao termo tensorial da parte negativa desse evento, onde $F = _CC + _IT + _FA$.

$$Q_{e^+}^A = \mu \begin{pmatrix} 0.7 & 0.3 \\ 0.5 & 0.5 \end{pmatrix}; \quad Q_{e^+}^B = \begin{pmatrix} _CC & _IT & _FA \\ _CC & _IT & _FA \\ _CC & _IT & _FA \end{pmatrix}; \quad Q_{e^-}^A = -\mu \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \quad Q_{e^-}^B = \begin{pmatrix} F & 0 & 0 \\ 0 & F & 0 \\ 0 & 0 & F \end{pmatrix}.$$

A função F sempre retorna “1” quando avaliada, pois:

$$\begin{aligned}
 F &= _CC + _IT + _FA \\
 &= (st \hat{A} == PCC) * (0.6 + 0.1 + 0.3) + (st \hat{A} == PIT) * (0.1 + 0.7 + 0.2) \\
 &= (st \hat{A} == PCC) * 1.0 + (st \hat{A} == PIT) * 1.0 \\
 F &= 1
 \end{aligned}$$

Então:

$$Q_e^B = \begin{pmatrix} F & 0 & 0 \\ 0 & F & 0 \\ 0 & 0 & F \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Ao resolver a Equação 6.1, obtém-se o gerador infinitesimal do modelo SAN para *Crazy Soda Machine*, como segue:

$$Q = \mu \left(\begin{array}{ccc|ccc}
 0.7*_CC - 1 & 0.7*_IT & 0.7*_FA & 0.3*_CC & 0.3*_IT & 0.3*_FA \\
 0.7*_CC & 0.7*_IT - 1 & 0.7*_FA & 0.3*_CC & 0.3*_IT & 0.3*_FA \\
 0.7*_CC & 0.7*_IT & 0.7*_FA - 1 & 0.3*_CC & 0.3*_IT & 0.3*_FA \\
 \hline
 0.5*_CC & 0.5*_IT & 0.5*_FA & 0.5*_CC - 1 & 0.5*_IT & 0.5*_FA \\
 0.5*_CC & 0.5*_IT & 0.5*_FA & 0.5*_CC & 0.5*_IT - 1 & 0.5*_FA \\
 0.5*_CC & 0.5*_IT & 0.5*_FA & 0.5*_CC & 0.5*_IT & 0.5*_FA - 1
 \end{array} \right)$$

Note que μ é um fator multiplicativo do gerador infinitesimal, não importando o seu valor, pois geradores infinitesimais proporcionais possuem a mesma solução estacionária. Então, após avaliar as probabilidades funcionais e fazer $\mu = 1$, obtém-se:

$$Q = \begin{pmatrix} -0.58 & 0.07 & 0.21 & 0.18 & 0.03 & 0.09 \\ 0.42 & -0.93 & 0.21 & 0.18 & 0.03 & 0.09 \\ 0.42 & 0.07 & -0.79 & 0.18 & 0.03 & 0.09 \\ 0.05 & 0.35 & 0.10 & -0.95 & 0.35 & 0.10 \\ 0.05 & 0.35 & 0.10 & 0.05 & -0.65 & 0.10 \\ 0.05 & 0.35 & 0.10 & 0.05 & 0.35 & -0.90 \end{pmatrix}$$

que confere com os resultados obtidos na Seção 6.1.1. Note que esse resultado poderia ter sido obtido diretamente pela Equação 5.3 (Teorema 5.3.1):

$$Q = \mu \left(\hat{A} \otimes_g \hat{B}(\hat{A}) - I_{n_A n_B} \right)$$

Entretanto, optou-se por mostrar o processo completo de resolução de um modelo SAN.

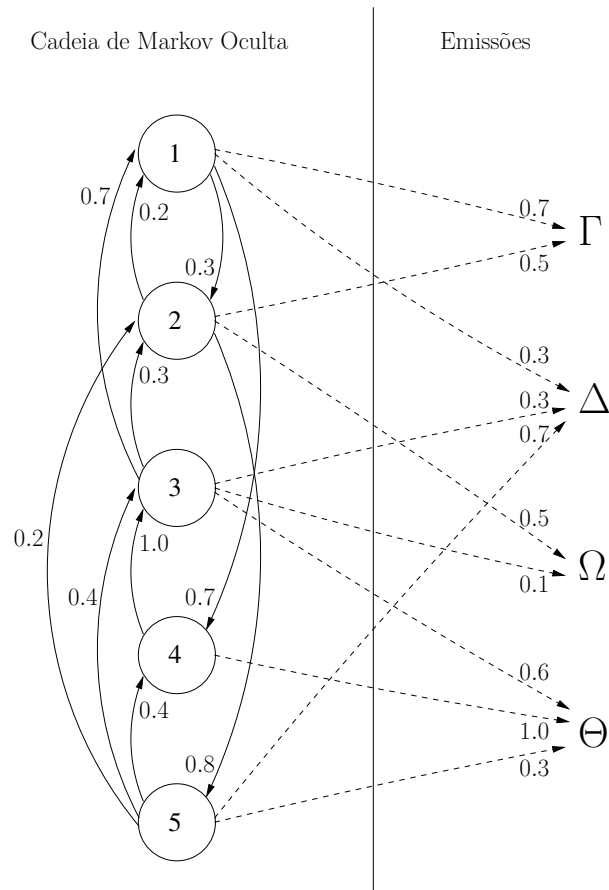


Figura 6.5 – *Greek Speller* – Modelagem em HMM

6.2.1 Obtenção do Gerador Infinitesimal do Modelo HMM da Figura 6.5

Um modelo HMM é definido pelos seus parâmetros, $\lambda = (A, B, \pi)$. A Figura 6.5 mostra a representação gráfica e os parâmetros do modelo HMM para o exemplo *Greek Speller*, tal que:

$$A = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{pmatrix} 0 & 0.3 & 0 & 0.7 & 0 \\ 0.2 & 0 & 0 & 0 & 0.8 \\ 0.7 & 0.3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0.2 & 0.4 & 0.4 & 0 \end{pmatrix} \end{matrix}; \quad B = \begin{matrix} & \begin{matrix} \Gamma & \Delta & \Omega & \Theta \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{pmatrix} 0.7 & 0.3 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.3 & 0.1 & 0.6 \\ 0 & 0 & 0 & 1 \\ 0 & 0.7 & 0 & 0.3 \end{pmatrix} \end{matrix}$$

Conhecendo o modelo HMM, pode-se obter o gerador infinitesimal correspondente usando qualquer dos dois métodos apresentados nas Seções 5.1 e 5.2.

Primeiro Método (Seção 5.1, p.75)

A partir das matrizes A e B obtém-se a matriz P' , a seguir, que indica a probabilidade conjunta da transição de um estado a outro acompanhada de uma emissão em função do estado de partida.

	1, Γ	1, Δ	1, Ω	1, Θ	2, Γ	2, Δ	2, Ω	2, Θ	3, Γ	3, Δ	3, Ω	3, Θ	4, Γ	4, Δ	4, Ω	4, Θ	5, Γ	5, Δ	5, Ω	5, Θ
1	0	0	0	0	0.3*0.7	0.3*0.3	0	0	0	0	0	0	0.7*0.7	0.7*0.3	0	0	0	0	0	0
2	0.2*0.5	0	0.2*0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8*0.5	0	0.8*0.5	0
3	0	0.7*0.3	0.7*0.1	0.7*0.6	0	0.3*0.3	0.3*0.1	0.3*0.6	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	1.0*1.0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0.2*0.7	0	0.2*0.3	0	0.4*0.7	0	0.4*0.3	0	0.4*0.7	0	0.4*0.3	0	0	0	0

Note que algumas colunas estão completamente zeradas, significando que aquele *estado global* específico não é atingível. Esse é o caso das colunas “3, Γ ”, “3, Ω ”, “4, Ω ”, “5, Δ ” e “5, Θ ”, que podem ser eliminadas, obtendo a seguinte matriz:

	1, Γ	1, Δ	1, Ω	1, Θ	2, Γ	2, Δ	2, Ω	2, Θ	3, Δ	3, Θ	4, Γ	4, Δ	4, Θ	5, Γ	5, Ω
1	0	0	0	0	0.21	0.09	0	0	0	0	0.49	0.21	0	0	0
2	0.10	0	0.10	0	0	0	0	0	0	0	0	0	0	0.40	0.40
3	0	0.21	0.07	0.42	0	0.09	0.03	0.18	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
5	0	0	0	0	0	0.14	0	0.06	0.28	0.12	0	0.28	0.12	0	0

Dada a independência entre emissões consecutivas, a matriz P' é estendida, gerando a matriz P , onde os valores repetidos em uma mesma coluna explicitam essa independência, garantindo que as emissões só dependam do estado de partida.

	1, Γ	1, Δ	1, Ω	1, Θ	2, Γ	2, Δ	2, Ω	2, Θ	3, Δ	3, Θ	4, Γ	4, Δ	4, Θ	5, Γ	5, Ω
1, Γ	0	0	0	0	0.21	0.09	0	0	0	0	0.49	0.21	0	0	0
1, Δ	0	0	0	0	0.21	0.09	0	0	0	0	0.49	0.21	0	0	0
1, Ω	0	0	0	0	0.21	0.09	0	0	0	0	0.49	0.21	0	0	0
1, Θ	0	0	0	0	0.21	0.09	0	0	0	0	0.49	0.21	0	0	0
2, Γ	0.10	0	0.10	0	0	0	0	0	0	0	0	0	0	0.40	0.40
2, Δ	0.10	0	0.10	0	0	0	0	0	0	0	0	0	0	0.40	0.40
2, Ω	0.10	0	0.10	0	0	0	0	0	0	0	0	0	0	0.40	0.40
2, Θ	0.10	0	0.10	0	0	0	0	0	0	0	0	0	0	0.40	0.40
3, Δ	0	0.21	0.07	0.42	0	0.09	0.03	0.18	0	0	0	0	0	0	0
3, Θ	0	0.21	0.07	0.42	0	0.09	0.03	0.18	0	0	0	0	0	0	0
4, Γ	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
4, Δ	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
4, Θ	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
5, Γ	0	0	0	0	0	0.14	0	0.06	0.28	0.12	0	0.28	0.12	0	0
5, Ω	0	0	0	0	0	0.14	0	0.06	0.28	0.12	0	0.28	0.12	0	0

A matriz P é a matriz de probabilidade de transição correspondente ao modelo HMM. Para transformá-la no gerador infinitesimal Q , basta que dela se subtraia a identidade:

$$Q = P - I$$

$$Q = \begin{array}{c|cccccccccccccccc} & 1,\Gamma & 1,\Delta & 1,\Omega & 1,\Theta & 2,\Gamma & 2,\Delta & 2,\Omega & 2,\Theta & 3,\Delta & 3,\Theta & 4,\Gamma & 4,\Delta & 4,\Theta & 5,\Gamma & 5,\Omega \\ \hline 1,\Gamma & -1 & 0 & 0 & 0 & 0.21 & 0.09 & 0 & 0 & 0 & 0 & 0.49 & 0.21 & 0 & 0 & 0 \\ 1,\Delta & 0 & -1 & 0 & 0 & 0.21 & 0.09 & 0 & 0 & 0 & 0 & 0.49 & 0.21 & 0 & 0 & 0 \\ 1,\Omega & 0 & 0 & -1 & 0 & 0.21 & 0.09 & 0 & 0 & 0 & 0 & 0.49 & 0.21 & 0 & 0 & 0 \\ 1,\Theta & 0 & 0 & 0 & -1 & 0.21 & 0.09 & 0 & 0 & 0 & 0 & 0.49 & 0.21 & 0 & 0 & 0 \\ 2,\Gamma & 0.10 & 0 & 0.10 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.40 & 0.40 \\ 2,\Delta & 0.10 & 0 & 0.10 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.40 & 0.40 \\ 2,\Omega & 0.10 & 0 & 0.10 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0.40 & 0.40 \\ 2,\Theta & 0.10 & 0 & 0.10 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0.40 & 0.40 \\ 3,\Delta & 0 & 0.21 & 0.07 & 0.42 & 0 & 0.09 & 0.03 & 0.18 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3,\Theta & 0 & 0.21 & 0.07 & 0.42 & 0 & 0.09 & 0.03 & 0.18 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 4,\Gamma & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 4,\Delta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 4,\Theta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 5,\Gamma & 0 & 0 & 0 & 0 & 0 & 0.14 & 0 & 0.06 & 0.28 & 0.12 & 0 & 0.28 & 0.12 & -1 & 0 \\ 5,\Omega & 0 & 0 & 0 & 0 & 0 & 0.14 & 0 & 0.06 & 0.28 & 0.12 & 0 & 0.28 & 0.12 & 0 & -1 \end{array}$$

Segundo Método (Seção 5.2, p.77)

A aplicação deste método requer a transformação dos parâmetros iniciais A e B do modelo HMM, como segue:

$$A = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{pmatrix} 0 & 0.3 & 0 & 0.7 & 0 \\ 0.2 & 0 & 0 & 0 & 0.8 \\ 0.7 & 0.3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0.2 & 0.4 & 0.4 & 0 \end{pmatrix} \longrightarrow \hat{A} = A$$

$$B = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{pmatrix} \Gamma & \Delta & \Omega & \Theta \\ 0.7 & 0.3 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.3 & 0.1 & 0.6 \\ 0 & 0 & 0 & 1 \\ 0 & 0.7 & 0 & 0.3 \end{pmatrix} \longrightarrow \hat{B}(\hat{A}) = \begin{array}{c} \Gamma \\ \Delta \\ \Omega \\ \Theta \end{array} \begin{pmatrix} \Gamma & \Delta & \Omega & \Theta \\ -\Gamma & -\Delta & -\Omega & -\Theta \\ -\Gamma & -\Delta & -\Omega & -\Theta \\ -\Gamma & -\Delta & -\Omega & -\Theta \\ -\Gamma & -\Delta & -\Omega & -\Theta \end{pmatrix}$$

$$\begin{aligned}
-\Gamma &= \begin{cases} 0.7 & \text{para } s_i = 1 \\ 0.5 & \text{para } s_i = 2 \end{cases} &
-\Omega &= \begin{cases} 0.5 & \text{para } s_i = 2 \\ 0.1 & \text{para } s_i = 3 \end{cases} &
-\Delta &= \begin{cases} 0.3 & \text{para } s_i = 1 \\ 0.3 & \text{para } s_i = 3 \\ 0.7 & \text{para } s_i = 5 \end{cases} &
-\Theta &= \begin{cases} 0.6 & \text{para } s_i = 3 \\ 1.0 & \text{para } s_i = 4 \\ 0.3 & \text{para } s_i = 5 \end{cases}
\end{aligned}$$

Após obtidos \hat{A} e $\hat{B}(\hat{A})$, o gerador infinitesimal Q é calculado pela Equação 5.2 (Página 80):

$$Q = \hat{A} \otimes_g \hat{B}(\hat{A}) - I_{n_{\hat{A}} n_{\hat{B}}}$$

	1,Γ	1,Δ	1,Ω	1,Θ	2,Γ	2,Δ	2,Ω	2,Θ	3,Γ	3,Δ	3,Ω	3,Θ	4,Γ	4,Δ	4,Ω	4,Θ	5,Γ	5,Δ	5,Ω	5,Θ
1,Γ	-1	0	0	0	0.3_Γ	0.3_Δ	0.3_Ω	0.3_Θ	0	0	0	0	0.7_Γ	0.7_Δ	0.7_Ω	0.7_Θ	0	0	0	0
1,Δ	0	-1	0	0	0.3_Γ	0.3_Δ	0.3_Ω	0.3_Θ	0	0	0	0	0.7_Γ	0.7_Δ	0.7_Ω	0.7_Θ	0	0	0	0
1,Ω	0	0	-1	0	0.3_Γ	0.3_Δ	0.3_Ω	0.3_Θ	0	0	0	0	0.7_Γ	0.7_Δ	0.7_Ω	0.7_Θ	0	0	0	0
1,Θ	0	0	0	-1	0.3_Γ	0.3_Δ	0.3_Ω	0.3_Θ	0	0	0	0	0.7_Γ	0.7_Δ	0.7_Ω	0.7_Θ	0	0	0	0
2,Γ	0.2_Γ	0.2_Δ	0.2_Ω	0.2_Θ	-1	0	0	0	0	0	0	0	0	0	0	0	0.8_Γ	0.8_Δ	0.8_Ω	0.8_Θ
2,Δ	0.2_Γ	0.2_Δ	0.2_Ω	0.2_Θ	0	-1	0	0	0	0	0	0	0	0	0	0	0.8_Γ	0.8_Δ	0.8_Ω	0.8_Θ
2,Ω	0.2_Γ	0.2_Δ	0.2_Ω	0.2_Θ	0	0	-1	0	0	0	0	0	0	0	0	0	0.8_Γ	0.8_Δ	0.8_Ω	0.8_Θ
2,Θ	0.2_Γ	0.2_Δ	0.2_Ω	0.2_Θ	0	0	0	-1	0	0	0	0	0	0	0	0	0.8_Γ	0.8_Δ	0.8_Ω	0.8_Θ
3,Γ	0.7_Γ	0.7_Δ	0.7_Ω	0.7_Θ	0.3_Γ	0.3_Δ	0.3_Ω	0.3_Θ	-1	0	0	0	0	0	0	0	0	0	0	0
3,Δ	0.7_Γ	0.7_Δ	0.7_Ω	0.7_Θ	0.3_Γ	0.3_Δ	0.3_Ω	0.3_Θ	0	-1	0	0	0	0	0	0	0	0	0	0
3,Ω	0.7_Γ	0.7_Δ	0.7_Ω	0.7_Θ	0.3_Γ	0.3_Δ	0.3_Ω	0.3_Θ	0	0	-1	0	0	0	0	0	0	0	0	0
3,Θ	0.7_Γ	0.7_Δ	0.7_Ω	0.7_Θ	0.3_Γ	0.3_Δ	0.3_Ω	0.3_Θ	0	0	0	-1	0	0	0	0	0	0	0	0
4,Γ	0	0	0	0	0	0	0	0	Γ	Δ	Ω	Θ	-1	0	0	0	0	0	0	0
4,Δ	0	0	0	0	0	0	0	0	Γ	Δ	Ω	Θ	0	-1	0	0	0	0	0	0
4,Ω	0	0	0	0	0	0	0	0	Γ	Δ	Ω	Θ	0	0	-1	0	0	0	0	0
4,Θ	0	0	0	0	0	0	0	0	Γ	Δ	Ω	Θ	0	0	0	-1	0	0	0	0
5,Γ	0	0	0	0	0.2_Γ	0.2_Δ	0.2_Ω	0.2_Θ	0.4_Γ	0.4_Δ	0.4_Ω	0.4_Θ	0.4_Γ	0.4_Δ	0.4_Ω	0.4_Θ	-1	0	0	0
5,Δ	0	0	0	0	0.2_Γ	0.2_Δ	0.2_Ω	0.2_Θ	0.4_Γ	0.4_Δ	0.4_Ω	0.4_Θ	0.4_Γ	0.4_Δ	0.4_Ω	0.4_Θ	0	-1	0	0
5,Ω	0	0	0	0	0.2_Γ	0.2_Δ	0.2_Ω	0.2_Θ	0.4_Γ	0.4_Δ	0.4_Ω	0.4_Θ	0.4_Γ	0.4_Δ	0.4_Ω	0.4_Θ	0	0	-1	0
5,Θ	0	0	0	0	0.2_Γ	0.2_Δ	0.2_Ω	0.2_Θ	0.4_Γ	0.4_Δ	0.4_Ω	0.4_Θ	0.4_Γ	0.4_Δ	0.4_Ω	0.4_Θ	0	0	0	-1

	1,Γ	1,Δ	1,Ω	1,Θ	2,Γ	2,Δ	2,Ω	2,Θ	3,Γ	3,Δ	3,Ω	3,Θ	4,Γ	4,Δ	4,Ω	4,Θ	5,Γ	5,Δ	5,Ω	5,Θ
1,Γ	-1	0	0	0	0.21	0.09	0	0	0	0	0	0	0.49	0.21	0	0	0	0	0	0
1,Δ	0	-1	0	0	0.21	0.09	0	0	0	0	0	0	0.49	0.21	0	0	0	0	0	0
1,Ω	0	0	-1	0	0.21	0.09	0	0	0	0	0	0	0.49	0.21	0	0	0	0	0	0
1,Θ	0	0	0	-1	0.21	0.09	0	0	0	0	0	0	0.49	0.21	0	0	0	0	0	0
2,Γ	0.10	0	0.10	0	-1	0	0	0	0	0	0	0	0	0	0	0	0.40	0	0.40	0
2,Δ	0.10	0	0.10	0	0	-1	0	0	0	0	0	0	0	0	0	0	0.40	0	0.40	0
2,Ω	0.10	0	0.10	0	0	0	-1	0	0	0	0	0	0	0	0	0	0.40	0	0.40	0
2,Θ	0.10	0	0.10	0	0	0	0	-1	0	0	0	0	0	0	0	0	0.40	0	0.40	0
3,Γ	0	0.21	0.07	0.42	0	0.09	0.03	0.18	-1	0	0	0	0	0	0	0	0	0	0	0
3,Δ	0	0.21	0.07	0.42	0	0.09	0.03	0.18	0	-1	0	0	0	0	0	0	0	0	0	0
3,Ω	0	0.21	0.07	0.42	0	0.09	0.03	0.18	0	0	-1	0	0	0	0	0	0	0	0	0
3,Θ	0	0.21	0.07	0.42	0	0.09	0.03	0.18	0	0	0	-1	0	0	0	0	0	0	0	0
4,Γ	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0
4,Δ	0	0	0	0	0	0	0	0	0	0	0	1	0	-1	0	0	0	0	0	0
4,Ω	0	0	0	0	0	0	0	0	0	0	0	1	0	0	-1	0	0	0	0	0
4,Θ	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	-1	0	0	0	0
5,Γ	0	0	0	0	0	0.14	0	0.06	0	0.28	0	0.12	0	0.28	0	0.12	-1	0	0	0
5,Δ	0	0	0	0	0	0.14	0	0.06	0	0.28	0	0.12	0	0.28	0	0.12	0	-1	0	0
5,Ω	0	0	0	0	0	0.14	0	0.06	0	0.28	0	0.12	0	0.28	0	0.12	0	0	-1	0
5,Θ	0	0	0	0	0	0.14	0	0.06	0	0.28	0	0.12	0	0.28	0	0.12	0	0	0	-1

Ao avaliar as funções, percebe-se que algumas colunas estão zeradas, indicando que o *estado global* não é atingível. Esse é o caso das colunas “3, Γ ”, “3, Ω ”, “4, Ω ”, “5, Δ ” e “5, Θ ”. Ao excluir essas colunas e correspondentes linhas, obtém-se a seguinte matriz:

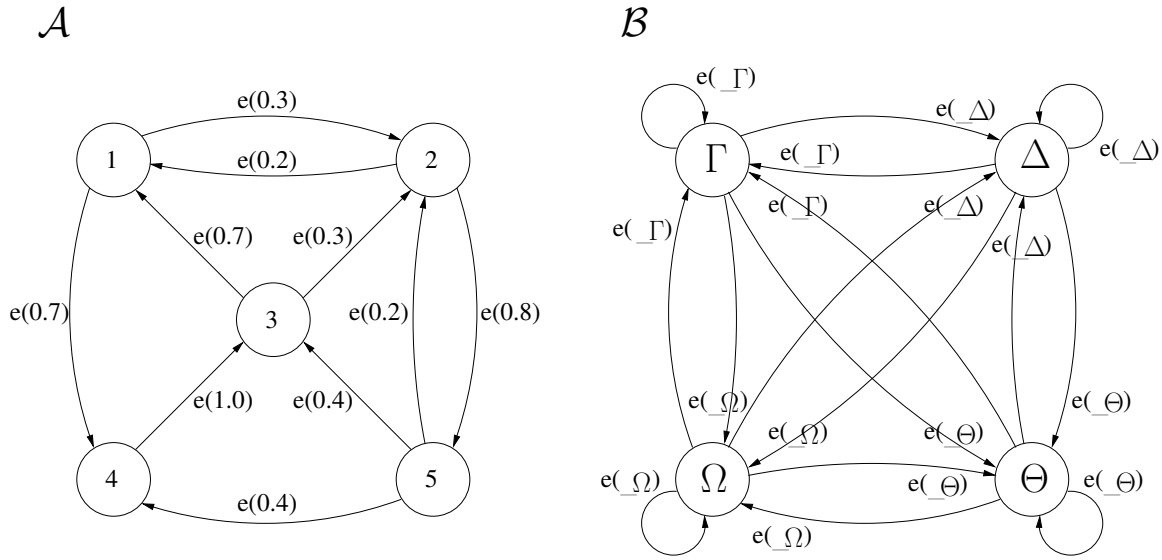
$$Q = \begin{array}{c|cccccccccccccccc} & 1,\Gamma & 1,\Delta & 1,\Omega & 1,\Theta & 2,\Gamma & 2,\Delta & 2,\Omega & 2,\Theta & 3,\Delta & 3,\Theta & 4,\Gamma & 4,\Delta & 4,\Theta & 5,\Gamma & 5,\Omega \\ \hline 1,\Gamma & -1 & 0 & 0 & 0 & 0.21 & 0.09 & 0 & 0 & 0 & 0 & 0.49 & 0.21 & 0 & 0 & 0 \\ 1,\Delta & 0 & -1 & 0 & 0 & 0.21 & 0.09 & 0 & 0 & 0 & 0 & 0.49 & 0.21 & 0 & 0 & 0 \\ 1,\Omega & 0 & 0 & -1 & 0 & 0.21 & 0.09 & 0 & 0 & 0 & 0 & 0.49 & 0.21 & 0 & 0 & 0 \\ 1,\Theta & 0 & 0 & 0 & -1 & 0.21 & 0.09 & 0 & 0 & 0 & 0 & 0.49 & 0.21 & 0 & 0 & 0 \\ 2,\Gamma & 0.10 & 0 & 0.10 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.40 & 0.40 \\ 2,\Delta & 0.10 & 0 & 0.10 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.40 & 0.40 \\ 2,\Omega & 0.10 & 0 & 0.10 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0.40 & 0.40 \\ 2,\Theta & 0.10 & 0 & 0.10 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0.40 & 0.40 \\ 3,\Delta & 0 & 0.21 & 0.07 & 0.42 & 0 & 0.09 & 0.03 & 0.18 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3,\Theta & 0 & 0.21 & 0.07 & 0.42 & 0 & 0.09 & 0.03 & 0.18 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 4,\Gamma & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 4,\Delta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 4,\Theta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 5,\Gamma & 0 & 0 & 0 & 0 & 0 & 0.14 & 0 & 0.06 & 0.28 & 0.12 & 0 & 0.28 & 0.12 & -1 & 0 \\ 5,\Omega & 0 & 0 & 0 & 0 & 0 & 0.14 & 0 & 0.06 & 0.28 & 0.12 & 0 & 0.28 & 0.12 & 0 & -1 \end{array}$$

6.2.2 Obtenção do Gerador Infinitesimal do Modelo SAN da Figura 6.5

Seja a modelagem HMM da Figura 6.5 para o problema *Greek Speller*. Pela aplicação do método apresentado na Seção 5.3 é possível converter esse modelo em um modelo SAN equivalente, cuja representação gráfica é vista na Figura 6.6.

Neste modelo, o autômato \mathcal{A} representa a cadeia de Markov oculta, enquanto que o autômato \mathcal{B} representa as emissões. Existe apenas um evento sincronizante, forçando transições simultâneas em ambos os autômatos. No autômato \mathcal{B} , as transições com destino a $y_l \in \mathcal{Y}$ possuem a mesma probabilidade de roteamento funcional, identificada por $_y_l$. Isso significa que as transições entre sinais independem do sinal y_k de partida. Entetando, as probabilidades funcionais $_y_l$ determinam que as transições entre sinais dependam do estado em que se encontra o autômato \mathcal{A} .

Dessa forma, no modelo SAN da Figura 6.6 cada transição entre estados da cadeia de Markov oculta força a emissão de um sinal, sendo esta emissão dependente do estado de onde parte a transição na MC oculta. A forma como as probabilidades funcionais $_y_l$ são distribuídas no modelo garante a independência entre emissões. Assim, pode-se dizer que há semelhança semântica entre o modelo SAN proposto e o modelo HMM original. Por exemplo, quando o autômato \mathcal{A} está no estado “1”, as únicas emissões possíveis são “ Γ ” e “ Δ ”, visto que as únicas probabilidades funcionais não-nulas são “ $_\Gamma$ ” e “ $_\Delta$ ”. Este resultado é compatível com o modelo HMM da Figura 6.5.



Evento	Taxa	Tipo
e	μ	sin

$$_ \Gamma = ((stA == 1) * 0.7 + (stA == 2) * 0.5)$$

$$_ \Delta = (((stA == 1) || (stA == 3)) * 0.3 + (stA == 5) * 0.7)$$

$$_ \Omega = ((stA == 2) * 0.5 + (stA == 3) * 0.1)$$

$$_ \Theta = ((stA == 3) * 0.6 + (stA == 4) * 1.0 + (stA == 5) * 0.3)$$

Figura 6.6 – *Greek Speller* – Modelagem em SAN

Como estabelece o Teorema 5.3.1 (Página 88), esse método de conversão gera uma SAN cujo descritor Markoviano pode ser obtido diretamente pela fórmula tensorial

$$Q = \mu(\hat{A} \otimes \hat{B}(\hat{A}) - I_{n_{\hat{A}} n_{\hat{B}}}), \quad (6.2)$$

cujos termos tensoriais e respectivas funções são dados por:

$$\hat{A} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{pmatrix} 0 & 0.3 & 0 & 0.7 & 0 \\ 0.2 & 0 & 0 & 0 & 0.8 \\ 0.7 & 0.3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0.2 & 0.4 & 0.4 & 0 \end{pmatrix} \end{matrix}$$

$$\hat{B}(\hat{A}) = \begin{matrix} & \begin{matrix} \Gamma & \Delta & \Omega & \Theta \end{matrix} \\ \begin{matrix} \Gamma \\ \Delta \\ \Omega \\ \Theta \end{matrix} & \begin{pmatrix} _ \Gamma & _ \Delta & _ \Omega & _ \Theta \\ _ \Gamma & _ \Delta & _ \Omega & _ \Theta \\ _ \Gamma & _ \Delta & _ \Omega & _ \Theta \\ _ \Gamma & _ \Delta & _ \Omega & _ \Theta \end{pmatrix} \end{matrix}$$

$$_ \Gamma = \begin{cases} 0.7 & \text{para } s_i = 1 \\ 0.5 & \text{para } s_i = 2 \end{cases} \quad _ \Omega = \begin{cases} 0.5 & \text{para } s_i = 2 \\ 0.1 & \text{para } s_i = 3 \end{cases}$$

$$_ \Delta = \begin{cases} 0.3 & \text{para } s_i = 1 \\ 0.3 & \text{para } s_i = 3 \\ 0.7 & \text{para } s_i = 5 \end{cases} \quad _ \Theta = \begin{cases} 0.6 & \text{para } s_i = 3 \\ 1.0 & \text{para } s_i = 4 \\ 0.3 & \text{para } s_i = 5 \end{cases}$$

Note que μ é um fator multiplicativo do gerador infinitesimal, não importando o seu valor, pois, como dito anteriormente, geradores infinitesimais proporcionais possuem a mesma solução estacionária. Então, ao resolver a Equação 6.2 atribuindo $\mu = 1$, obtém-se:

	1, Γ	1, Δ	1, Ω	1, Θ	2, Γ	2, Δ	2, Ω	2, Θ	3, Γ	3, Δ	3, Ω	3, Θ	4, Γ	4, Δ	4, Ω	4, Θ	5, Γ	5, Δ	5, Ω	5, Θ
1, Γ	-F	0	0	0	0.3 Γ	0.3 Δ	0.3 Ω	0.3 Θ	0	0	0	0	0.7 Γ	0.7 Δ	0.7 Ω	0.7 Θ	0	0	0	0
1, Δ	0	-F	0	0	0.3 Γ	0.3 Δ	0.3 Ω	0.3 Θ	0	0	0	0	0.7 Γ	0.7 Δ	0.7 Ω	0.7 Θ	0	0	0	0
1, Ω	0	0	-F	0	0.3 Γ	0.3 Δ	0.3 Ω	0.3 Θ	0	0	0	0	0.7 Γ	0.7 Δ	0.7 Ω	0.7 Θ	0	0	0	0
1, Θ	0	0	0	-F	0.3 Γ	0.3 Δ	0.3 Ω	0.3 Θ	0	0	0	0	0.7 Γ	0.7 Δ	0.7 Ω	0.7 Θ	0	0	0	0
2, Γ	0.2 Γ	0.2 Δ	0.2 Ω	0.2 Θ	-F	0	0	0	0	0	0	0	0	0	0	0	0.8 Γ	0.8 Δ	0.8 Ω	0.8 Θ
2, Δ	0.2 Γ	0.2 Δ	0.2 Ω	0.2 Θ	0	-F	0	0	0	0	0	0	0	0	0	0	0.8 Γ	0.8 Δ	0.8 Ω	0.8 Θ
2, Ω	0.2 Γ	0.2 Δ	0.2 Ω	0.2 Θ	0	0	-F	0	0	0	0	0	0	0	0	0	0.8 Γ	0.8 Δ	0.8 Ω	0.8 Θ
2, Θ	0.2 Γ	0.2 Δ	0.2 Ω	0.2 Θ	0	0	0	-F	0	0	0	0	0	0	0	0	0.8 Γ	0.8 Δ	0.8 Ω	0.8 Θ
3, Γ	0.7 Γ	0.7 Δ	0.7 Ω	0.7 Θ	0.3 Γ	0.3 Δ	0.3 Ω	0.3 Θ	-F	0	0	0	0	0	0	0	0	0	0	0
3, Δ	0.7 Γ	0.7 Δ	0.7 Ω	0.7 Θ	0.3 Γ	0.3 Δ	0.3 Ω	0.3 Θ	0	-F	0	0	0	0	0	0	0	0	0	0
3, Ω	0.7 Γ	0.7 Δ	0.7 Ω	0.7 Θ	0.3 Γ	0.3 Δ	0.3 Ω	0.3 Θ	0	0	-F	0	0	0	0	0	0	0	0	0
3, Θ	0.7 Γ	0.7 Δ	0.7 Ω	0.7 Θ	0.3 Γ	0.3 Δ	0.3 Ω	0.3 Θ	0	0	0	-F	0	0	0	0	0	0	0	0
4, Γ	0	0	0	0	0	0	0	0	Γ	Δ	Ω	Θ	-F	0	0	0	0	0	0	0
4, Δ	0	0	0	0	0	0	0	0	Γ	Δ	Ω	Θ	0	-F	0	0	0	0	0	0
4, Ω	0	0	0	0	0	0	0	0	Γ	Δ	Ω	Θ	0	0	-F	0	0	0	0	0
4, Θ	0	0	0	0	0	0	0	0	Γ	Δ	Ω	Θ	0	0	0	-F	0	0	0	0
5, Γ	0	0	0	0	0.2 Γ	0.2 Δ	0.2 Ω	0.2 Θ	0.4 Γ	0.4 Δ	0.4 Ω	0.4 Θ	0.4 Γ	0.4 Δ	0.4 Ω	0.4 Θ	-F	0	0	0
5, Δ	0	0	0	0	0.2 Γ	0.2 Δ	0.2 Ω	0.2 Θ	0.4 Γ	0.4 Δ	0.4 Ω	0.4 Θ	0.4 Γ	0.4 Δ	0.4 Ω	0.4 Θ	0	-F	0	0
5, Ω	0	0	0	0	0.2 Γ	0.2 Δ	0.2 Ω	0.2 Θ	0.4 Γ	0.4 Δ	0.4 Ω	0.4 Θ	0.4 Γ	0.4 Δ	0.4 Ω	0.4 Θ	0	0	-F	0
5, Θ	0	0	0	0	0.2 Γ	0.2 Δ	0.2 Ω	0.2 Θ	0.4 Γ	0.4 Δ	0.4 Ω	0.4 Θ	0.4 Γ	0.4 Δ	0.4 Ω	0.4 Θ	0	0	0	-F

	1, Γ	1, Δ	1, Ω	1, Θ	2, Γ	2, Δ	2, Ω	2, Θ	3, Γ	3, Δ	3, Ω	3, Θ	4, Γ	4, Δ	4, Ω	4, Θ	5, Γ	5, Δ	5, Ω	5, Θ
1, Γ	-1	0	0	0	0.21	0.09	0	0	0	0	0	0	0.49	0.21	0	0	0	0	0	0
1, Δ	0	-1	0	0	0.21	0.09	0	0	0	0	0	0	0.49	0.21	0	0	0	0	0	0
1, Ω	0	0	-1	0	0.21	0.09	0	0	0	0	0	0	0.49	0.21	0	0	0	0	0	0
1, Θ	0	0	0	-1	0.21	0.09	0	0	0	0	0	0	0.49	0.21	0	0	0	0	0	0
2, Γ	0.10	0	0.10	0	-1	0	0	0	0	0	0	0	0	0	0	0	0.40	0	0.40	0
2, Δ	0.10	0	0.10	0	0	-1	0	0	0	0	0	0	0	0	0	0	0.40	0	0.40	0
2, Ω	0.10	0	0.10	0	0	0	-1	0	0	0	0	0	0	0	0	0	0.40	0	0.40	0
2, Θ	0.10	0	0.10	0	0	0	0	-1	0	0	0	0	0	0	0	0	0.40	0	0.40	0
3, Γ	0	0.21	0.07	0.42	0	0.09	0.03	0.18	-1	0	0	0	0	0	0	0	0	0	0	0
3, Δ	0	0.21	0.07	0.42	0	0.09	0.03	0.18	0	-1	0	0	0	0	0	0	0	0	0	0
3, Ω	0	0.21	0.07	0.42	0	0.09	0.03	0.18	0	0	-1	0	0	0	0	0	0	0	0	0
3, Θ	0	0.21	0.07	0.42	0	0.09	0.03	0.18	0	0	0	-1	0	0	0	0	0	0	0	0
4, Γ	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0
4, Δ	0	0	0	0	0	0	0	0	0	0	0	1	0	-1	0	0	0	0	0	0
4, Ω	0	0	0	0	0	0	0	0	0	0	0	1	0	0	-1	0	0	0	0	0
4, Θ	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	-1	0	0	0	0
5, Γ	0	0	0	0	0	0.14	0	0.06	0	0.28	0	0.12	0	0.28	0	0.12	-1	0	0	0
5, Δ	0	0	0	0	0	0.14	0	0.06	0	0.28	0	0.12	0	0.28	0	0.12	0	-1	0	0
5, Ω	0	0	0	0	0	0.14	0	0.06	0	0.28	0	0.12	0	0.28	0	0.12	0	0	-1	0
5, Θ	0	0	0	0	0	0.14	0	0.06	0	0.28	0	0.12	0	0.28	0	0.12	0	0	0	-1

Ao avaliar as probabilidades funcionais percebe-se que algumas colunas são zeradas. Eliminando essas colunas e correspondentes linhas, obtém-se a matriz que segue, a qual confere com os resultados obtidos na Seção 6.2.1:

	1, Γ	1, Δ	1, Ω	1, Θ	2, Γ	2, Δ	2, Ω	2, Θ	3, Δ	3, Θ	4, Γ	4, Δ	4, Θ	5, Γ	5, Ω
1, Γ	-1	0	0	0	0.21	0.09	0	0	0	0	0.49	0.21	0	0	0
1, Δ	0	-1	0	0	0.21	0.09	0	0	0	0	0.49	0.21	0	0	0
1, Ω	0	0	-1	0	0.21	0.09	0	0	0	0	0.49	0.21	0	0	0
1, Θ	0	0	0	-1	0.21	0.09	0	0	0	0	0.49	0.21	0	0	0
2, Γ	0.10	0	0.10	0	-1	0	0	0	0	0	0	0	0	0.40	0.40
2, Δ	0.10	0	0.10	0	0	-1	0	0	0	0	0	0	0	0.40	0.40
2, Ω	0.10	0	0.10	0	0	0	-1	0	0	0	0	0	0	0.40	0.40
2, Θ	0.10	0	0.10	0	0	0	0	-1	0	0	0	0	0	0.40	0.40
3, Δ	0	0.21	0.07	0.42	0	0.09	0.03	0.18	-1	0	0	0	0	0	0
3, Θ	0	0.21	0.07	0.42	0	0.09	0.03	0.18	0	-1	0	0	0	0	0
4, Γ	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0
4, Δ	0	0	0	0	0	0	0	0	0	1	0	-1	0	0	0
4, Θ	0	0	0	0	0	0	0	0	0	1	0	0	-1	0	0
5, Γ	0	0	0	0	0	0.14	0	0.06	0.28	0.12	0	0.28	0.12	-1	0
5, Ω	0	0	0	0	0	0.14	0	0.06	0.28	0.12	0	0.28	0.12	0	-1

6.2.3 *Greek Speller* – Representação Gráfica da CTMC Correspondente

Qualquer que seja o exemplo modelado, a partir do gerador infinitesimal é possível obter a representação gráfica da CTMC correspondente. Contudo, como visto na Figura 6.7, nem sempre isso é desejável, visto o quão confusa essa representação pode ser, mesmo para modelos pequenos, como é o caso do modelo *Greek Speller* (com vinte estados globais). A busca por clareza na representação gráfica é um bom argumento para adotar SAN como ferramenta de modelagem.

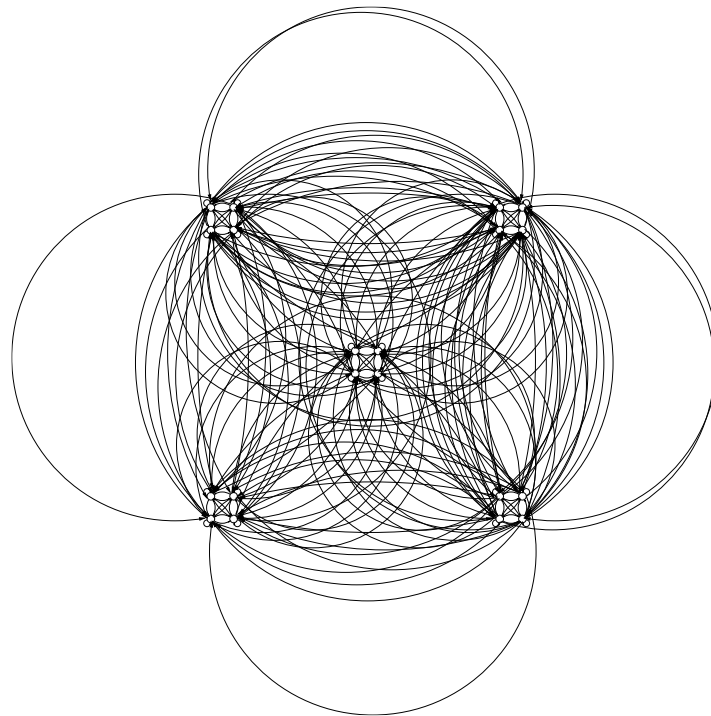


Figura 6.7 – *Greek Speller* – Cadeia de Markov Equivalente (com 20 estados)

7. CONSIDERAÇÕES FINAIS E PERSPECTIVAS

Este trabalho teve como objetivo formalizar uma proposta de representação tensorial para Modelos Markovianos Ocultos (*Hidden Markov Models* – HMM). A estratégia usada foi identificar cada estágio da HMM como um autômato e modelar o processo usando o formalismo SAN (*Stochastic Automata Networks*), que foi escolhido por possuir um formato tensorial conhecido. Algumas restrições foram impostas, de forma que a HMM deveria ser discreta¹ e finita, tanto no que diz respeito ao espaço de estados quanto ao conjunto de sinais.

O método de conversão proposto é sistemático, gerando sempre dois autômatos cujas transições são gerenciadas por um único evento sincronizante. Como consequência, a fórmula tensorial obtida é sempre a mesma, independente do modelo HMM original, desde que em observância às restrições iniciais, comentadas no parágrafo anterior. Como visto nos exemplos do Capítulo 6, o método se aplica a casos em que a cadeia de Markov oculta é ergódica, entretanto, entende-se que isso não se constitui em uma restrição, sendo possível trabalhar com modelos HMM do tipo *left-right*, tão comuns em reconhecimento de fala [26]. O argumento que embasa esta afirmativa reside no fato que um dos autômatos da SAN obtida pela aplicação desse método é justamente a própria cadeia de Markov oculta do modelo HMM, que não necessariamente é ergódica. Outro caso importante e não explicitamente abordado nesta monografia diz respeito à não obrigatoriedade de emissão, em que, por vezes um estado é visitado mas deixa de emitir algum sinal. No método de conversão proposto, isso poderia ser tratado com a inserção de um *sinal vazio* no conjunto de sinais do modelo.

Faz-se importante colocar que a busca de uma estruturação ótima para HMM esteve, desde o começo, fora do escopo do trabalho, que se deteve ao estudo de sua viabilidade. Entretanto, uma vez provada a viabilidade pelo método de conversão de HMM em SAN, pode-se pensar em métodos mais eficazes de estruturação, de forma a usufruir dos possíveis benefícios da modelagem em SAN. A estruturação tem na clareza de representação a sua maior virtude, tendendo a gerar modelos mais compreensíveis. O formato tensorial, por sua vez, é uma alternativa à redução do uso de memória computacional na solução de modelos [12], contudo, não há garantias, mas fortes indícios [13, 31], de que promova redução no custo computacional.

A contribuição desta dissertação está na prova da viabilidade de uma estruturação e correspondente formato tensorial para Modelos Markovianos Ocultos. Uma possibilidade baseada na conversão de HMM em SAN foi apresentada. Ficam para trabalhos futuros a busca de novos métodos de conversão e, quem sabe, uma conversão ótima.

Existem expectativas sobre a criação do que poderia ser chamado de HSMM, ou *Hidden Structured Markov Models*², onde ambos os níveis da HMM poderiam vir a ser estruturados. Para dar um pouco de perspectiva, considere uma HMM cuja cadeia de Markov oculta possui milhares de estados. Não seria interessante estruturá-la em alguns poucos autômatos, com cerca de uma dezena

¹Uma HMM é dita discreta se seus observáveis formam um conjunto discreto de sinais.

²Nome sugerido pelo orientador desta dissertação.

de estados, cada? Um modelo como este, além de mais compreensível, teria maiores chances de ser tratado computacionalmente de forma eficiente.

Também consta do conjunto de trabalhos futuros a adaptação dos algoritmos usados na solução dos problemas clássicos de HMM (*Forward-Backward*, *Viterbi* e *Baum-Welch*), assim como a extensão da ferramenta PEPS para comportar a solução desses problemas. Dito isto, esta dissertação lança as bases destes trabalhos ao estabelecer uma possibilidade de estruturação e formatação tensorial para HMM.

Bibliografia

- [1] A. G. Adami. Automatic Speech Recognition: From the Beginning to the Portuguese Language. *Tutorial at the 9th International Conference on Computational Processing of the Portuguese Language*, 2010. <http://www.inf.pucrs.br/~propor2010/proceedings/tutorials/Adami.pdf>.
- [2] V. Amoia, G. D. Micheli, M. Santomauro. "Computer-Oriented Formulation of Transition-Rate Matrices via Kronecker Algebra". *IEEE Transactions on Reliability*, R-30(2):123–132, 1981.
- [3] L. Bahl, F. Jelinek. "Decoding for channels with insertions, deletions, and substitutions with applications to speech recognition". *IEEE Transactions on Information Theory*, 21(4):404 – 411, 1975.
- [4] J. Baker. "The DRAGON system—An overview". *IEEE Transactions on Acoustics, Speech and Signal Processing*, 23(1):24–29, Fevereiro 1975.
- [5] L. Baum. "An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes". *Inequalities*, 3:1–8, 1972.
- [6] L. E. Baum, J. A. Egon. "An inequality with applications to statistical estimation for probabilistic functions of a Markov process and to a model for ecology". *Bulletin of the American Meteorological Society*, 73:360–363, 1967.
- [7] L. E. Baum, T. Petrie. "Statistical Inference for Probabilistic Functions of Finite State Markov Chains". *The Annals of Mathematical Statistics*, 37(6):1554–1563, 1966.
- [8] L. E. Baum, T. Petrie, G. Soules, N. Weiss. "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains". *The Annals of Mathematical Statistics*, 41(1):164–171, 1970.
- [9] L. E. Baum, G. R. Sell. "Growth functions for transformations on manifolds". *Pacific Journal Of Mathematics*, 27(2):211–227, 1968.
- [10] L. Brenner. "Réseaux d'Automates Stochastiques: Analyse transitoire en temps continu et Algèbre tensorielle pour une sémantique en temps discret". PhD thesis, Institut National Polytechnique de Grenoble, France, 2009.
- [11] L. Brenner, P. Fernandes, B. Plateau, I. Sbeity. "PEPS 2007 - Stochastic Automata Networks Software Tool". In *Proceedings of the 4th International Conference on Quantitative Evaluation of Systems (QEST 2007)*, pages 163–164. IEEE Press, Setembro 2007.

- [12] L. Brenner, P. Fernandes, A. Sales. "The Need for and the Advantages of Generalized Tensor Algebra for Kronecker Structured Representations". *International Journal of Simulation: Systems, Science & Technology (IJSIM)*, 6(3-4):52–60, Fevereiro 2005. 20th Annual UK Performance Engineering Workshop (UKPEW 2004).
- [13] R. M. Czekster. "Solução numérica de descritores Markovianos a partir de re-estruturações de termos tensoriais". PhD thesis, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Brasil, Maio 2010.
- [14] M. Davio. "Kronecker Products and Shuffle Algebra". *IEEE Transactions on Computers*, C-30(2):116–125, 1981.
- [15] R. Durbin, S. R. Eddy, A. Krogh, G. Mitchison. "Biological sequence analysis: Probabilistic models of proteins and nucleic acids". Cambridge University Press, 1998.
- [16] P. Fernandes. "Méthodes numériques pour la solution de systèmes Markoviens à grand espace d'états". PhD thesis, Institut National Polytechnique de Grenoble, France, 1998.
- [17] P. Fernandes, B. Plateau, W. J. Stewart. "Efficient descriptor-vector multiplication in stochastic automata networks". *Journal of the ACM (JACM)*, 45(3):381–414, Maio 1998.
- [18] G. A. Fink. "Markov Models for Pattern Recognition, From Theory to Applications". Springer, Heidelberg, 2008.
- [19] Hakeem.gadi. HMM Sequence Production. <http://en.wikipedia.org/wiki/File:HMMsequence.svg>, Outubro 2008.
- [20] F. Jelinek. "Fast Sequential Decoding Algorithm Using a Stack". *IBM Journal of Research and Development*, 13(6):675 –685, 1969.
- [21] F. Jelinek. "Statistical Methods for Speech Recognition". The MIT Press, 1998.
- [22] F. Jelinek, L. Bahl, R. Mercer. "Design of a linguistic statistical decoder for the recognition of continuous speech". *IEEE Transactions on Information Theory*, 21(3):250 – 256, 1975.
- [23] L. Kronecker. "Über einige Interpolationsformeln für ganze Funktionen mehrerer Variablen". In *L. Kroneckers Werke*, volume I, pages 133–141, Lectures at the academy of sciences, Dezembro 21 1865. Chelsea Publishing Company.
- [24] C. D. Manning, H. Schütze. "Foundations of Statistical Natural Language Processing". The MIT Press, Cambridge, Massachusetts, London, England, 1999.
- [25] B. Plateau. "On the stochastic structure of parallelism and synchronization models for distributed algorithms". *ACM SIGMETRICS Performance Evaluation Review*, 13(2):147–154, Agosto 1985.

- [26] L. R. Rabiner. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition". *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [27] L. E. Reichl. "A Modern Course in Statistical Physics". WILEY-VCH, second edition, 2004.
- [28] A. Sales. "Formalismos Estruturados de Modelagem para Sistemas Markovianos Complexos". Master's thesis, PUCRS-FACIN-PPGCC, Porto Alegre, Dezembro 2003.
- [29] A. Sales. *Réseaux d'Automates Stochastiques: Génération de l'espace d'états atteignables et Multiplication vecteur-descripteur pour une sémantique en temps discret*. PhD thesis, Institut National Polytechnique de Grenoble, France, 2009. <http://tel.archives-ouvertes.fr/tel-00436020/en/>.
- [30] W. J. Stewart. "Introduction to the numerical solution of Markov chains". Princeton University Press, 1994.
- [31] T. Webber. "Reducing the Impact of State Space Explosion in Stochastic Automata Networks". PhD thesis, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Brasil, Abril 2009.
- [32] Wikipedia. Hidden Markov Model. http://en.wikipedia.org/wiki/Hidden_Markov_model, Janeiro 2011.

A. ÁLGEBRA TENSORIAL

No século *XIX* o matemático alemão Leopold Kronecker [23] propôs uma operação baseada em tensores. Tal operação foi agregada ao conjunto de operações e propriedades definidas em álgebra linear, e passou a ser representada por um único operador chamado *produto Kronecker* ou *produto tensorial*.

Sua aplicação na área de ciência da computação, data, de acordo com Brenner, Fernandes e Sales [12], do final da década de 1970. Os trabalhos de Amoia, De Micheli e Santomauro [2] e de Davio [14] merecem destaque por terem estendido o trabalho de Kronecker com a inclusão da *soma tensorial*, levando à criação de uma nova álgebra, composta de produto e soma tensorial, conhecida como Álgebra Tensorial Clássica (*Classical Tensor Algebra* – CTA).

Redes de Autômatos Estocásticos (*Stochastic Automata Networks* – SAN), proposto em 1984 por Plateau [25], foi o primeiro formalismo para modelagem de sistemas a usar Álgebra Tensorial Clássica. Trata-se de uma representação estruturada de cadeias de Markov e, por possuir uma álgebra tensorial associada, propicia uma modelagem mais compacta do que aquela obtida pelo uso de cadeias de Markov.

A definição de CTA foi estendida em 1998, por Fernandes, Plateau e Stewart [17], com a introdução do conceito de *funções* nas modelagens feitas em SAN como uma forma alternativa de modelar o comportamento local de um autômato em decorrência de alterações em outros autômatos do modelo. Dessa forma, criou-se a Álgebra Tensorial Generalizada (*Generalized Tensor Algebra* – GTA), que não difere de CTA no que diz respeito à capacidade de modelagem e solução de problemas, mas sim no que diz respeito à facilidade e eficiência com que isso é feito [12].

Nas seções que seguem, são apresentados os conceitos e propriedades da Álgebra tensorial. A Seção A.1 é dedicada à Álgebra Tensorial Clássica [2,14], enquanto que a Seção A.2 trata de Álgebra Tensorial Generalizada [16]. O que segue foi extraído da dissertação de Sales [28].

A.1 CTA – Álgebra Tensorial Clássica

Define-se duas matrizes A e B a saber:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad B = \begin{pmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \end{pmatrix}$$

O *produto tensorial* $C = A \otimes B$ é dado por:

$$C = \begin{pmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{pmatrix}$$

Em geral, para se definir o produto tensorial de duas matrizes: A de dimensões $(\rho_1 \times \gamma_1)$ e B de dimensões $(\rho_2 \times \gamma_2)$, é conveniente observar que a matriz resultante do produto tensorial possui dimensões $(\rho_1 \rho_2 \times \gamma_1 \gamma_2)$ e pode ser considerada como composta de $\rho_1 \gamma_1$ blocos cada um com dimensões $(\rho_2 \gamma_2)$, *i.e.*, as dimensões de B . Para se especificar um elemento em particular, basta apenas especificar o bloco e a posição dentro do bloco que o elemento em questão se encontra. Desta forma, conforme definido anteriormente, o elemento c_{47} ($a_{22} b_{13}$) está no bloco (2, 2) e na posição interna (1, 3) do mesmo. O produto tensorial $C = A \otimes B$ é definido pela associação do elemento de C que está na posição interna (k, l) do bloco (i, j) , *i.e.*:

$$C_{[ik][jl]} = a_{ij} b_{kl}.$$

A soma tensorial de duas matrizes quadradas A e B é definida em relação ao produto tensorial que segue:

$$A \oplus B = A \otimes I_{n_B} + I_{n_A} \otimes B$$

onde n_A é a ordem de A ; n_B é a ordem de B ; I_{n_i} é a matriz identidade de ordem n_i e “+” representa a operação usual de soma de matrizes. Portanto, a soma tensorial é definida somente por matrizes quadradas, visto que ambos os lados desta operação (soma de matrizes) devem possuir dimensões idênticas. O valor designado para o elemento $C_{[ik][jl]}$ da soma tensorial $C = A \oplus B$ é definida como:

$$C_{[ik][jl]} = a_{ij} \delta_{kl} + b_{kl} \delta_{ij},$$

onde δ_{ij} é o elemento da i -ésima linha e j -ésima coluna de uma matriz identidade¹ definido como:

$$\delta_{ij} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}$$

Algumas propriedades importantes das operações produto e soma tensorial definidos por Davio [2,14] são:

- Associatividade:

$$A \otimes (B \otimes C) = (A \otimes B) \otimes C \text{ e } A \oplus (B \oplus C) = (A \oplus B) \oplus C$$

- Distributividade em relação à soma:

$$(A + B) \otimes (C + D) = (A \otimes C) + (B \otimes C) + (A \otimes D) + (B \otimes D)$$

- Compatibilidade com a multiplicação:

$$(A \times B) \otimes (C \times D) = (A \otimes C) \times (B \otimes D)$$

- Decomposição em fatores normais:

$$A \otimes B = (A \otimes I_{n_B}) \times (I_{n_A} \otimes B)$$

¹ δ é geralmente conhecido como operador de Kronecker.

- Comutatividade dos fatores normais²:

$$(A \otimes I_{n_B}) \times (I_{n_A} \otimes B) = (I_{n_A} \otimes B) \times (A \otimes I_{n_B})$$

A.2 GTA – Álgebra Tensorial Generalizada

Álgebra Tensorial Generalizada é uma extensão da Álgebra Tensorial Clássica. A principal diferença da GTA em relação à CTA é o acréscimo do conceito de *elementos funcionais*. Entretanto, uma matriz pode ser composta por elementos constantes (pertencentes a \mathbb{R}) ou elementos funcionais. Um elemento funcional é uma função avaliada em \mathbb{R} de acordo com um conjunto de parâmetros compostos por linhas de uma ou mais matrizes. O produto tensorial generalizado é representado pelo símbolo \otimes_g . O valor associado para o elemento $C_{[ik][jl]}$ de um produto tensorial generalizado $C = A(\mathcal{B}) \otimes_g B(\mathcal{A})$ é definido como:

$$C_{[ik][jl]} = a_{ij}(b_k)b_{kl}(a_i).$$

A soma tensorial generalizada também é análoga à soma tensorial clássica, e ela é representada pelo símbolo \oplus_g . Os elementos de uma soma tensorial generalizada $C = A(\mathcal{B}) \oplus_g B(\mathcal{A})$ são definidos como:

$$C_{[ik][jl]} = a_{ij}(b_k)\delta_{kl} + b_{kl}(a_i)\delta_{ij}.$$

As propriedades da GTA definidas por Fernandes, Plateau e Stewart [17] são:

- Associatividade:

$$\begin{aligned} [A(\mathcal{B}, \mathcal{C}) \otimes_g B(\mathcal{A}, \mathcal{C})] \otimes_g C(\mathcal{A}, \mathcal{B}) &= A(\mathcal{B}, \mathcal{C}) \otimes_g [B(\mathcal{A}, \mathcal{C}) \otimes_g C(\mathcal{A}, \mathcal{B})] \\ [A(\mathcal{B}, \mathcal{C}) \oplus_g B(\mathcal{A}, \mathcal{C})] \oplus_g C(\mathcal{A}, \mathcal{B}) &= A(\mathcal{B}, \mathcal{C}) \oplus_g [B(\mathcal{A}, \mathcal{C}) \oplus_g C(\mathcal{A}, \mathcal{B})] \end{aligned}$$

- Distributividade em relação à soma:

$$[A(\mathcal{C}, \mathcal{D}) + B(\mathcal{C}, \mathcal{D})] \otimes_g [C(\mathcal{A}, \mathcal{B}) + D(\mathcal{A}, \mathcal{B})] = A(\mathcal{C}, \mathcal{D}) \otimes_g C(\mathcal{A}, \mathcal{B}) + A(\mathcal{C}, \mathcal{D}) \otimes_g D(\mathcal{A}, \mathcal{B}) + B(\mathcal{C}, \mathcal{D}) \otimes_g C(\mathcal{A}, \mathcal{B}) + B(\mathcal{C}, \mathcal{D}) \otimes_g D(\mathcal{A}, \mathcal{B})$$

- Decomposição em fatores normais I:

$$A \otimes_g B(\mathcal{A}) = I_{n_A} \otimes_g B(\mathcal{A}) \times A \otimes_g I_{n_B}$$

- Decomposição em fatores normais II:

$$A(\mathcal{B}) \otimes_g B = A(\mathcal{B}) \otimes_g I_{n_B} \times I_{n_A} \otimes_g B$$

- Decomposição do Produto Tensorial Generalizado para o Produto Tensorial Clássico:

$$A \otimes_g B(\mathcal{A}) = \sum_{k=1}^{n_A} \ell_k(A) \otimes B(a_k)$$

²Embora esta propriedade possa ser deduzida da propriedade *Compatibilidade com a multiplicação*, ela foi definida por Fernandes, Plateau e Stewart [17].