

**UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA**

**Aprendizado em Modelos de Markov com Variáveis de
Estado Escondidas**

Roberto Castro Alamino

*Tese apresentada ao Instituto
de Física da Universidade de
São Paulo para a obtenção
do título de Doutor em Ciên-
cias.*

Orientador : Prof. Dr. Nestor Felipe Caticha Alfonso

Banca Examinadora:



Prof. Dr. Nestor Felipe Caticha Alfonso (IF-USP)

Profa. Dra. Carmem Pimentel Cintra do Prado (IF-USP)

Prof. Dr. Rogério Rosenfeld (IFT-UNESP)

Prof. Dr. Marco Aurélio Pires Idiart (UFRGS)

Prof. Dr. Renato Vicente (EACH-USP)

SBI-IFUSP



305M810T4409



Prof. Celso Luiz Lima
Presidente da Comissão de Pós-Graduação

SÃO PAULO

2005



FICHA CATALOGRÁFICA
Preparada pelo Serviço de Biblioteca e Informação
do Instituto de Física da Universidade de São Paulo

Alamino, Roberto Castro

Aprendizado em Modelos de Markov com Variáveis de Estado Escondidas. São Paulo, 2005.

Tese (Doutoramento) - Universidade de São Paulo
Instituto de Física – Depto de Física Geral

Orientador: Prof. Dr. Nestor Caticha
Área de Concentração: Física

Unitermos:

1. Mecânica Estatística Clássica-aprendizado em máquinas;
2. Mecânica Estatística Clássica-Modelos Markovianos;
3. Inferência Bayesiana-Inferência estatística.

USP/IF/SBI-082/2005

Sumário

1	Introdução	1
1.1	Visão Geral	1
1.2	Organização da Tese	2
1.3	Convenções	4
2	Fundamentos	5
2.1	Complexidade Computacional	5
2.1.1	Introdução	5
2.1.2	Eficiência Computacional	6
2.1.3	Classes de Complexidade	7
2.2	Processos Markovianos	9
2.2.1	Processos Markovianos Discretos	9
2.2.2	Processos Markovianos Contínuos	12
2.3	Inferência Estatística	14
2.3.1	Introdução	14
2.3.2	Máxima Verossimilhança	15
2.3.3	O Algoritmo EM	16
2.3.4	Inferência Bayesiana	17
2.4	Modelos Gráficos	20
2.4.1	Introdução	20
2.4.2	Modelos Gráficos Direcionados	21
2.4.3	Modelos Gráficos Não-Direcionados	22
2.4.4	Inferência	24
2.5	Aprendizado em Máquinas	24
2.5.1	O Cenário de Aprendizado	26
2.5.2	Algoritmos <i>Off-line</i> e <i>On-line</i>	27
2.5.3	Performance e Generalização	28

2.5.4	O Algoritmo Bayesiano <i>On-line</i>	29
3	Modelos de Markov com Variáveis de Estado Escondidas	32
3.1	Teoria	32
3.2	Algoritmos	36
3.3	Aplicações	41
3.3.1	Bioinformática	41
3.3.2	Econometria	43
3.3.3	Meteorologia	43
3.3.4	Música	44
3.3.5	Reconhecimento de Fala	44
4	Aprendizado Não-Bayesiano	45
4.1	Distância entre HMMs	46
4.2	Simulações	50
4.3	Fórmulas de Reestimação de Baum-Welch	51
4.3.1	Algoritmo	51
4.3.2	Simulações	54
4.4	Método de Baum-Welch <i>On-line</i>	58
4.4.1	Algoritmo	58
4.4.2	Simulações	61
4.5	Algoritmo de Baldi-Chauvin	67
4.5.1	Algoritmo	67
4.5.2	Relação entre BC e BWO	72
4.5.3	Simulações	73
5	Aprendizado Bayesiano	80
5.1	Algoritmo Bayesiano <i>On-line</i>	80
5.1.1	Algoritmo	80
5.1.2	Simulações	84
5.2	Posterior Média	85
5.2.1	Algoritmo	85
5.2.2	Simulações	86
5.3	Estudo Teórico: A Moeda Simples	92
5.3.1	BKL	97
5.3.2	PM	98
5.4	Conceitos Variáveis	98

5.5	Aplicação : Previsão de Índices	100
5.6	Aprendizado e Quebra de Simetria	106
6	Conclusões e Perspectivas	111
6.1	Conclusões	111
6.1.1	Distâncias entre HMMs	111
6.1.2	Algoritmos	112
6.2	Perspectivas	112
6.2.1	Estudos Teóricos	113
6.2.2	Aplicações	114
A	O Método de Descida pelo Gradiente	115
B	Distribuições de Dirichlet	116
B.1	Definições	116
B.2	Normalização	116
B.3	Momentos	119
B.4	Obtenção pelo Princípio da Máxima Entropia	121
B.5	Mistura de Dirichlets	122
C	O Sistema Digama	123
D	Projeção de uma Distribuição em uma Família Paramétrica	124
D.1	Obtenção da Família Paramétrica via Máxima Entropia	124
D.2	Projeção de uma Distribuição na Família Paramétrica	125
E	Notação	127
E.1	Abreviações	127
E.2	Símbolos	127

Lista de Figuras

2.1	O problema $P \times NP$: A figura <i>a</i> mostra o caso em que existem 3 classes e a figura <i>b</i> o caso onde apenas duas existem.	9
2.2	Grafo de Transição de uma CMH: Os círculos denotam os estados com os respectivos nomes e as setas as possíveis transições com os valores das probabilidades.	11
2.3	Distribuição a Priori: Exemplo de distribuição a priori da variável <i>A</i> no intervalo real $[a, b]$ tal que os valores mais prováveis de <i>A</i> se encontram perto do ponto <i>a</i>	17
2.4	Modelos Gráficos: Modelo gráfico representando relações de interdependência entre 4 variáveis aleatórias diferentes.	20
2.5	Modelos Gráficos Direcionados: Modelo gráfico direcionado representando os 4 primeiros passos de um processo markoviano de ordem 2.	22
2.6	Modelos Gráficos Não-Direcionados: Modelo gráfico representando o modelo de Ising em uma rede 3×3	23
2.7	Percéptron: Representação gráfica de um perceptron.	25
2.8	Percéptron: Representação gráfica de um perceptron com uma camada intermediária (chamada de camada escondida).	26
2.9	Algoritmo Bayesiano <i>On-line</i>: a) 1o. passo: quando um novo dado é incorporado, a distribuição antiga é modificada de tal modo que sai do espaço formado por sua família paramétrica. b) 2o. passo: a distribuição é então projetada de volta no espaço das distribuições paramétricas originais através da minimização da distância de Kullback-Leibler.	30
3.1	HMM: Representação gráfica de um HMM.	33
3.2	Topologias de um HMM: a) modelo <i>left-right</i> ; b) modelo de <i>loop</i>	35

- 4.1 **Algoritmo de Baum-Welch:** a) Gráfico mostrando a distância euclidiana entre o professor e o aluno após a aplicação do algoritmo para uma seqüência de T símbolos fazendo a média em 110 professores com 2 estados escondidos e 3 estados observados. A linha tracejada mostra o limite do aprendizado; b) os mesmos dados do gráfico *a* subtraindo-se o valor limite 0.3, porém em log-log. A linha tracejada indica o ajuste de uma lei de potência $(d - 0.3) = \alpha T^{-\xi}$ para os últimos 100 pontos com resultados $\alpha = 0.25$ e $\xi = 0.69$ 56
- 4.2 **Algoritmo de Baum-Welch:** a) O gráfico ilustra o efeito de aumentar o número de estados escondidos do HMM; b) mesmo gráfico em escala log-log. 57
- 4.3 **Algoritmo de Baum-Welch:** a) O gráfico mostra a comparação entre duas simulações com alunos iniciais diferentes do algoritmo de Baum-Welch: os círculos mostram o resultado obtido com um aluno inicial simétrico e os pontos com as respectivas barras de erro mostram a média em 50 alunos iniciais aleatórios; b) o mesmo gráfico em escala log-log. 59
- 4.4 **Método de Baum-Welch *On-line*:** a) Gráfico mostrando a distância de Kullback-Leibler entre o professor e o aluno após a aplicação do algoritmo para 4000 seqüências de 2 símbolos fazendo a média em 500 professores com 2 estados escondidos e 3 estados observados. Os números próximos a cada linha indicam o valor de η_{BW} utilizado. b) O mesmo que o gráfico *a*, porém em log-log. 62
- 4.5 **Método de Baum-Welch *On-line*:** a) Gráfico mostrando a distância Euclidiana entre o professor e o aluno após a aplicação do algoritmo para 4000 seqüências de 2 símbolos fazendo a média em 500 professores com 2 estados escondidos e 3 estados observados. Os números próximos a cada linha indicam o valor de η_{BW} utilizado. b) O mesmo que o gráfico *a*, porém em log-log. 63
- 4.6 **Método de Baum-Welch *On-line*:** a) Gráfico comparando a curva de aprendizado de um aluno inicial simétrico (linha tracejada) com a média de 60 alunos aleatórios (linha contínua); b) o mesmo gráfico em escala log-log. 64
- 4.7 **Método de Baum-Welch *On-line*:** Comparação entre o algoritmo BW e o algoritmo BWO adaptado para aprender uma seqüência de tamanho T quebrando-a em subseqüências menores com tamanho indicado junto às curvas correspondentes. 65

- 4.8 **Método de Baum-Welch *On-line*:** Valor médio da distância entre aluno e professor nos últimos 100 pontos das curvas de aprendizado de BWO para seqüências únicas com tamanho de 2 até 100 símbolos em função do tamanho das subseqüências. 66
- 4.9 **Algoritmo de Baldi-Chauvin:** a) Gráfico mostrando a distância de Kullback-Leibler entre o professor e o aluno após a aplicação do algoritmo para 4000 seqüências de 2 símbolos fazendo a média em 200 professores com 2 estados escondidos e 3 estados observados. Os números próximos a cada linha indicam o valor de η_{BC} utilizado. b) O mesmo que o gráfico *a*, porém em log-log e para 10000 seqüências. 74
- 4.10 **Algoritmo de Baldi-Chauvin:** a) Gráfico mostrando a distância Euclideana entre o professor e o aluno após a aplicação do algoritmo para 4000 seqüências de 2 símbolos fazendo a média em 200 professores com 2 estados escondidos e 3 estados observados. Os números próximos a cada linha indicam o valor de η_{BC} utilizado. b) O mesmo que o gráfico *a*, porém em log-log e para 10000 seqüências. 75
- 4.11 **Algoritmo de Baldi-Chauvin:** Gráfico mostrando a comparação entre as curvas de aprendizado para o algoritmo BC (linha contínua) e para o algoritmo BWO com a taxa de aprendizado estimada (linha tracejada) e com a melhor taxa obtida por tentativa e erro (linha pontilhada). As curvas são traçadas para $\eta_{BC} = 1$ fixo e para 3 valores distintos de λ , de baixo para cima, 0.01, 0.001 e 0.0001. 76
- 4.12 **Algoritmo de Baldi-Chauvin:** Gráfico mostrando a distância euclideana entre aluno e professor para o algoritmo BC aplicado ao aprendizado de uma única seqüência de tamanho T utilizando cinco tamanhos diferentes de subseqüências indicadas ao lado de cada curva comparando o resultado com o obtido utilizando o algoritmo BW. 77
- 4.13 **Algoritmo de Baldi-Chauvin:** Gráfico mostrando a distância euclideana entre aluno e professor para o algoritmo BC aplicado ao aprendizado de uma única seqüência de tamanho T utilizando subseqüências de tamanho 6 para cinco diferentes valores da taxa de aprendizado η_{BC} indicados ao lado de cada curva. 78

- 5.1 **Posterior Média:** a) Gráfico mostrando e distância de Kullback-Leibler entre o professor e o aluno após a aplicação do algoritmo para 10000 seqüências de 2 símbolos fazendo a média em 500 professores com 2 estados escondidos e 3 estados observados. b) O mesmo que o gráfico *a*, porém em log-log. 87
- 5.2 **Posterior Média:** a) Gráfico mostrando e distância Euclideana entre o professor e o aluno após a aplicação do algoritmo para 10000 seqüências de 2 símbolos fazendo a média em 500 professores com 2 estados escondidos e 3 estados observados. b) O mesmo que o gráfico *a*, porém em log-log. 88
- 5.3 **Posterior Média:** a) A curva superior representa o valor da variância de π ao longo do aprendizado, as duas curvas inferiores, as variâncias de *A*. b) Variâncias de *B* ao longo do aprendizado. 90
- 5.4 **Posterior Média:** O gráfico log-log acima mostra a distância d_{KL} em função do número de seqüências aprendidas para cinco valores iniciais de u : 10^{-4} , 0.5, 1.0, 2.0 e 20.0. 91
- 5.5 **Posterior Média:** a) Comparação entre PM e BC. BC foi rodado com $\eta_{BC} = 0.5$ e para 5 diferentes valores de λ indicados junto às respectivas curvas. b) Comparação entre PM e BWO. BWO foi rodado para 4 diferentes valores de η_{BW} indicados junto às respectivas curvas. Para os dois gráficos: $n = 2$, $m = 3$, $T = 2$ e $N_\tau = 500$ e a linha inferior corresponde à curva de aprendizado de PM. 93
- 5.6 **Posterior Média:** a) Comparação entre PM e BC. BC foi rodado com $\eta_{BC} = 0.5$ e para 5 diferentes valores de λ indicados junto às respectivas curvas. Escalas em log-log. b) Comparação entre PM e BWO. BWO foi rodado para 4 diferentes valores de η_{BW} indicados junto às respectivas curvas. Para os dois gráficos: $n = 2$, $m = 3$, $T = 2$ e $N_\tau = 500$ e a linha inferior corresponde à curva de aprendizado de PM. Escalas em log-log. 94
- 5.7 **Posterior Média:** a) Gráfico comparativo da performance de PM versus BKL. Ambos foram rodados para os seguintes valores: $n = 2$, $m = 3$, $T = 2$ e $N_\tau = 150$, para um total de 300 seqüências. Os pequenos círculos mostram o resultado de BKL, enquanto a linha tracejada mostra o resultado de PM; b) mesmo gráfico em escala log-log. 95

5.8	Posterior Média: a) Gráfico mostrando as curvas de aprendizado para uma situação onde há apenas uma única seqüência de tamanho T a ser aprendida dividida em diferentes tamanhos de subseqüências indicados na legenda; b) mesmo gráfico em escala log-log.	96
5.9	Teste de Wisconsin para HMMs: Gráfico comparativo da performance de PM versus BC. A linha contínua representa PM e a linha tracejada representa BC. O professor é mudado aleatoriamente a cada 500 seqüências.	100
5.10	Mudança Aleatória Contínua: Gráfico comparativo da performance de PM versus BC. A linha contínua representa PM e a linha tracejada representa BC. O professor sofre uma mudança aleatória pequena a cada seqüência.	101
5.11	Previsão do Ibovespa: a) Gráfico comparativo da performance de 4 diferentes estratégias: apostando aleatoriamente (Aleat.), apostando de acordo com o resultado do dia anterior (Dia Ant.), e usando os HMMs treinados pelos algoritmos PM e BC com subseqüências de tamanho 2. b) Gráfico mostrando a evolução no tempo do índice Ibovespa deflacionado.	105
5.12	Quebra de Simetria: Gráfico mostrando a distância entre aluno e professor e os valores das entradas dos parâmetros do HMM aluno em função do número de seqüências aprendidas para o algoritmo BC.	107
5.13	Quebra de Simetria: Gráfico mostrando a distância entre aluno e professor e os valores das entradas dos parâmetros do HMM aluno em função do número de seqüências aprendidas para o algoritmo PM.	108
5.14	Quebra de Simetria: Gráfico mostrando a distância entre aluno e professor e os valores das entradas dos parâmetros do HMM aluno em função do número de seqüências aprendidas para o algoritmo BWO.	109
B.1	Normalização da Distribuição de Dirichlet: Caminho de integração no plano complexo utilizado para o caso $k > 0$	118
B.2	Normalização da Distribuição de Dirichlet: Caminho de integração no plano complexo utilizado para o caso $k < 0$	119

Gracias por visitar este Libro Electrónico

Puedes leer la versión completa de este libro electrónico en diferentes formatos:

- HTML(Gratis / Disponible a todos los usuarios)
- PDF / TXT(Disponible a miembros V.I.P. Los miembros con una membresía básica pueden acceder hasta 5 libros electrónicos en formato PDF/TXT durante el mes.)
- Epub y Mobipocket (Exclusivos para miembros V.I.P.)

Para descargar este libro completo, tan solo seleccione el formato deseado, abajo:

