

AGRUPAMENTO DE CLIENTES COM BASE NA FICHA DE ANAMNESE ODONTOLÓGICA: UMA APLICAÇÃO DA ART2.

andrey soares

Universidade Federal de Santa Catarina – UFSC
Campus Universitário – Trindade – Florianópolis, SC – 88040-900 – andrey@inf.ufsc.br

marcello thiry, dr.

Universidade do Vale do Itajaí – UNIVALI
Campus VII - Rodovia SC 407 Km 4 – São José, SC – 88122-00 – thiry@sj.univali.br

paulo sérgio da silva borges, dr.

Universidade Federal de Santa Catarina – UFSC
Campus Universitário – Trindade – Florianópolis, SC – 88040-900 – pssb@inf.ufsc.br

Abstract

This work presents the application of Artificial Neural Networks, in particular the ART2, for the customer clustering of a dentistry room. The input data used in the application is based on the odontological anemnesis that is a form with a questionnaire applied about the professional to identify the customer case history. The three proposed customer clustering (good, medium, bad) were created with basis on the buccal hygiene care and the similar habits among the customers. The network is trained using non supervised learning that can be fast or slow learning. Each input data line is formed by the number of interviewed customers (rows) and the answered questions (columns). However, the first step was to transform the answers into binary cells.

Keywords: Anamnese, Adaptative Resonance Theory, Clusters

1 INTRODUÇÃO

Em um consultório odontológico, o exame clínico (ficha de anamnese) é o passo inicial para se realizar um tratamento, como para um melhor relacionamento entre paciente e o profissional. A ficha de anamnese constitui-se de um questionário que é realizado pelo profissional, geralmente na primeira visita (consulta) do cliente ao consultório odontológico, e tem o objetivo de conhecer o histórico clínico do cliente, permitindo o diagnóstico, prognóstico e tratamento das enfermidades da cavidade bucal. Porém, para obter-se êxito em um tratamento é preciso que se conheça os problemas que afligem o paciente.

A técnica de inteligência artificial conhecida por Redes Neurais Artificiais representam um sistema de processamento de informações que tem certas características de desempenho em comum com as Redes Neurais Biológicas (FAUSETT, 1994) (RICH, 1993) (RABUSKE, 1994). As Redes Neurais Artificiais tem sido desenvolvidas como generalizações dos modelos matemáticos, baseado nas suposições que:

- O processamento de informações ocorre em vários elementos simples chamados neurônios;
- Sinais são passados entre neurônios através das conexões de links;
- Cada conexão de link tem um peso associado;

- Cada neurônio aplica uma função de ativação (usualmente não linear) na entrada da rede, determinando o sinal de saída.

Considerando as características do problema, multiplicidade de alternativas, grande dificuldade de estabelecer relação entre os itens componentes da solução do problema, alto grau de incerteza na definição do padrão, alto índice de dificuldade na estruturação da melhor alternativa de solução, optou-se por aplicar uma solução com sistemas conexionistas (CANDIDO et. al., 2000).

A rede neural escolhida foi a Teoria da Ressonância Adaptativa (ART), desenvolvida por Carpenter e Grossberg (FAUSETT, 1994 *apud* CARPENTER et. al., 1987a). Esta rede agrupa entradas usando aprendizagem não supervisionada, ou seja, pode ser usada como um módulo de “descoberta de características”. Os agrupamentos (Figura 1) dos candidatos se formam a partir de características semelhantes entre os mesmos, identificadas pela Rede. A teoria da ressonância adaptativa possui duas formas, a ART1 que é usada para agrupar vetores binários e a ART2 que aceita vetores de valores contínuos. Este artigo aborda a ART2.

Algumas redes alcançam estabilidade reduzindo a taxa de aprendizagem gradualmente, onde o mesmo jogo de padrões de treinamento é apresentado vários vezes. Porém, isto não habilita a rede para aprender um padrão novo, que é apresentado pela primeira vez depois que várias épocas de treinamento já aconteceram. A habilidade de uma rede para responder (aprender) um padrão novo é chamado plasticidade. Para Fausett (1994), Redes com a Teoria de Ressonância Adaptativa são projetadas para serem estáveis e plásticas.

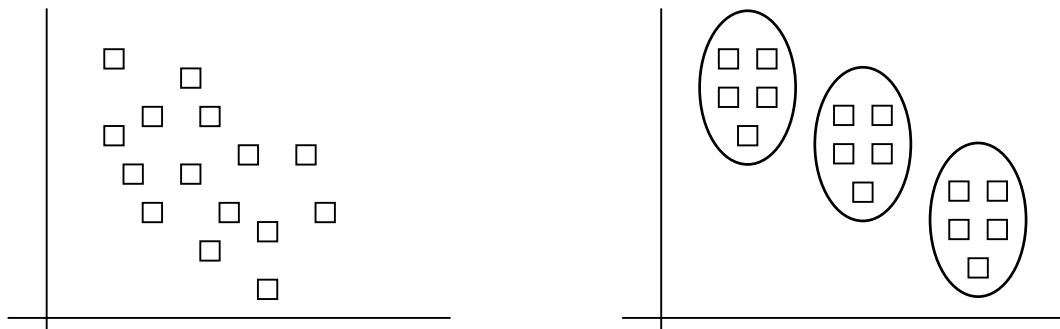


Figura 1: Agrupamento dos candidatos

Este artigo apresenta uma proposta de utilização de Redes Neurais Artificiais (RNA) – ART2 no agrupamento de clientes de um consultório odontológico, baseando-se na ficha de anamnese odontológica.

2 TEORIA DA RESSONÂNCIA ADAPTATIVA

Segundo Fausett (1994), redes com a Teoria de Ressonância Adaptativa, são projetadas para permitir ao usuário controlar o grau de semelhança de padrões colocados no mesmo agrupamento. Para que isso aconteça, os padrões de entrada devem poder diferir, entre eles, no nível de detalhe (número de componentes que são não zero), a semelhança relativa de um padrão de entrada para o vetor de peso para uma unidade de agrupamento, em lugar de utilizar a diferença absoluta entre os vetores.

Como a rede é treinada, cada padrão de treinamento pode ser apresentado várias vezes. Um padrão pode ser colocado em uma unidade de agrupamento na primeira vez que é apresentado e então colocado em um agrupamento diferente quando é apresentado depois (devido a mudanças nos pesos para o primeiro agrupamento se ele aprendeu outros padrões enquanto isso). Uma rede instável não retornará um padrão a um agrupamento prévio; em

outras palavras, um padrão que oscila entre unidades de agrupamento diferentes em fases diferentes de treinamento indica uma rede instável.

A arquitetura básica de uma rede neural de ressonância adaptável envolve três grupos de neurônios: um campo de processamento de entradas (camada F_1), as unidades de agrupamento (Camada F_2), e um mecanismo para controlar o grau de semelhança de padrões colocado no mesmo agrupamento (um mecanismo de *reset*).

Para controlar a semelhança de padrões colocada no mesmo agrupamento, há dois conjuntos de conexão (cada um com seus próprios pesos) entre cada unidade na porção de interface do campo de entrada e cada unidade de agrupamento. A camada F_2 é a camada competitiva.

Na teoria da ressonância adaptativa, as mudanças em ativações de unidades e em pesos são governadas pelo acoplamento de equações diferenciais. A rede está mudando continuamente (dinamicamente) o sistema, mas o processo pode ser simplificado porque é assumido que as ativações mudam muito mais rapidamente que os pesos. Uma vez que uma unidade de agrupamento aceitável foi selecionada para aprender, os sinais *bottom-up* e *top-down* são mantidos por um período estendido, e é durante este tempo que as mudanças de peso acontecem. Esta é a "ressonância" que dá o nome a rede.

Dois tipos de aprendizagem podem ser usadas para redes ART (CANDIDO et. al., 2000):

- **Modo de aprendizagem rápida**, é assumido que a atualização de pesos durante a ressonância aconteça rapidamente, pois relativo à duração de tempo um padrão é apresentado em qualquer tentativa particular. Assim, na aprendizagem rápida, os pesos alcançam equilíbrio em cada tentativa.
- **Modo de aprendizagem lenta**, as mudanças de peso acontecem lentamente relativo à duração de uma tentativa de aprendizagem; os pesos não alcançam equilíbrio em uma tentativa particular. Muitos mais apresentações de padrões são requeridos para a aprendizagem lenta do que para aprendizagem rápida, mas menos cálculos acontecem em cada tentativa de aprendizagem.

3 ART2

Segundo Fausett (1994), as redes ART2 são projetadas para executar, com vetores de entrada contínuos, o mesmo tipo de tarefas que a ART1 faz para vetores de entrada binários. As diferenças entre ART2 e ART1 refletem as modificações necessárias para acomodar padrões com componentes de valores contínuos. O campo F_1 (mais complexo na ART2) é necessário porque vetores de entrada contínuos podem ter valores arbitrários. Este campo inclui uma combinação de normalização e supressão de ruído, além da comparação de sinais *bottom-up* e *top-down* necessárias para o mecanismo de *reset*.

3.1 Arquitetura

A arquitetura típica da ART2 é ilustrada na (Figura 2). A camada F_1 consiste de seis tipos de unidades (W , X , U , V , P e Q). Só uma unidade de cada tipo é mostrada na Figura. Uma unidade suplementar entre as unidades W e as unidades X recebe sinais de todas as unidades W , computam a norma do vetor w , e enviam este (inibitório) sinal para cada uma das unidades X . Cada um destes recebem também um sinal excitatório da unidade W correspondente.

Uma unidade suplementar semelhante executa o mesmo papel entre as unidades P e as unidades Q , e outra faz o mesmo entre as unidades V e as unidades U . Cada unidade X é conectada à uma unidade V correspondente, e cada unidade Q também é conectada à uma unidade V correspondente.

Os símbolos na rota de conexão entre as várias unidades na camada F_1 , indicam a transformação que ocorre no sinal ao passar de um tipo de unidade para o próximo; eles não indicam multiplicação pela quantidade determinada. Porém, as conexões entre as unidades P_i (da camada F_1) e Y_j (da camada F_2) mostram os pesos que multiplicam o sinal transmitido em cima dessas rotas. A ativação do vencedor da unidade F_2 é d , onde $0 < d < 1$.

O símbolo $\cdot \triangleright$ indica normalização; isto é, o vetor de ativação q das unidades Q é justamente o mesmo vetor de ativação das unidades P , normalizado para aproximadamente o comprimento (duração) da unidade. As unidades competem em um “concurso” pelo direito de aprender cada padrão de entrada. O aprendizado ocorre somente se o vetor de pesos *top-down* da unidade vencedora, for suficientemente similar ao vetor de entradas.

As unidades X_i e Q_j , aplicam uma função de ativação para suas entradas da rede; esta função suprime qualquer componente dos vetores de ativações que cai abaixo dos níveis do valor θ . Os caminhos de conexão de W para U e de Q para V fixam os pesos a e b , respectivamente.

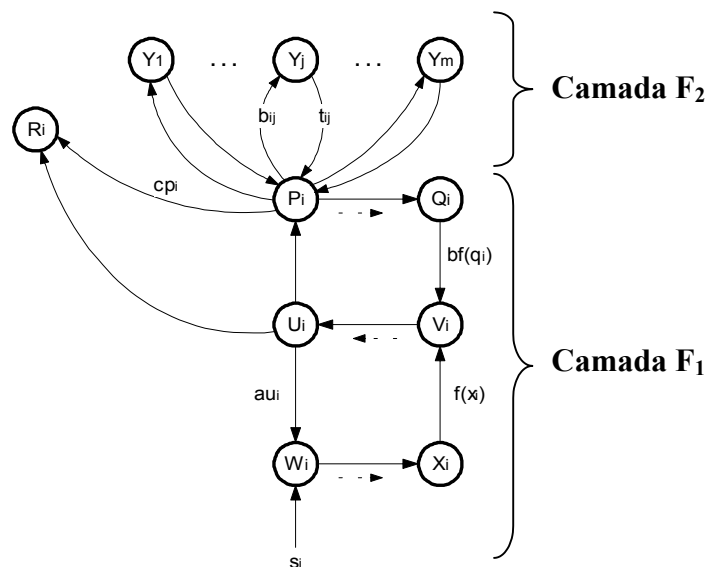


Figura 2: Arquitetura típica ART2

3.2 Algoritmo de Treinamento

O algoritmo (passos) seguinte pode ser usado para aprendizagem rápida ou aprendizagem lenta (FAUSETT, 1994). Em aprendizagem lenta, somente uma iteração da equação de atualização de pesos ocorre em cada processo de aprendizagem. Um grande número de apresentações de cada padrão é requerido, mas relativamente pouca computação é feita em cada processo. No algoritmo a seguir, estas apresentações que repetem são tratadas como épocas. Em aprendizagem rápida, a atualização dos pesos continua até o equilíbrio dos pesos em cada processo. Somente algumas épocas são requeridas, mas um grande número de iterações através da atualização dos pesos do algoritmo podem ser desempenhados em cada processo de aprendizagem.

A unidade J é o nó vencedor da camada F_2 depois da competição. Se nenhuma unidade vencedora for escolhida, d será 0 (zero) para todas as unidades.

Passo 0. Inicialize os parâmetros: $a, b, \theta, c, d, e, \alpha, p$.

Passo 1. Faça os passos 2 a 12 tantas vezes quantas forem os números de épocas. Execute o número especificado de épocas de treinamento.

Passo 2. Para cada vetor de entrada s_i , faça os passos 3 a 11.

Passo 3. Atualize as ativações da unidade F_1 :

$$\begin{aligned} u_i &= 0 & x_i &= \frac{s_i}{e + \|s\|} \\ w_i &= s_i & q_i &= 0 \\ p_i &= 0 & v_i &= f(x_i) \end{aligned}$$

Atualize as ativações da unidade F_1 novamente:

$$\begin{aligned} u_i &= \frac{v_i}{e + \|v\|} & w_i &= s_i + au_i \\ x_i &= \frac{w_i}{e + \|w\|} & q_i &= \frac{p_i}{e + \|p\|} \\ p_i &= u_i & v_i &= f(x_i) + bf(q_i) \end{aligned}$$

Passo 4. Calcule os sinais para as unidades F_1 :

$$y_i = \sum_i b_{ij} p_i$$

Passo 5. Enquanto o reajuste (*Reset*) é verdadeiro, faça os passos 6 – 7.

Passo 6. Ache Y_j da unidade F_2 com sinal maior. (Defina J tal que $y_j \geq y_1$ para $j=1, \dots, m$)

Passo 7. Confira o reajuste (*Reset*):

$$\begin{aligned} u_i &= \frac{v_i}{e + \|v\|} & r_i &= \frac{u_i + cp_i}{e + \|u\| + e\|p\|} \\ p_i &= u_i + dt_{ij} \end{aligned}$$

Se $\|r\| < p - e$, então

$$y_i = -1 \quad (\text{inibe } J)$$

Reset verdadeiro; repita o passo 5;

Se $\|r\| \geq p - e$, então

$$\begin{aligned} x_i &= \frac{w_i}{e + \|w\|} & q_i &= \frac{p_i}{e + \|p\|} \\ w_i &= s_i + au_i & v_i &= f(x_i) + bf(q_i) \end{aligned}$$

Reajuste falso; Proceda o passo 8.

Passo 8. Faça os passos 9 a 11 tantas vezes quanto forem os números de iterações. Execute o número especificado de repetições de aprendizagem.

Passo 9. Atualize pesos para unidade vencedora J :

$$\begin{aligned} t_{ij} &= \alpha du_i + \{1 + \alpha d(d-1)\}t_{ij} \\ b_{ij} &= \alpha du_i + \{1 + \alpha d(d-1)\}b_{ij} \end{aligned}$$

Passo 10. Atualize as ativações de F_1 :

$$\begin{aligned} u_i &= \frac{v_i}{e + \|v\|} & w_i &= s_i + au_i \\ x_i &= \frac{w_i}{e + \|w\|} & q_i &= \frac{p_i}{e + \|p\|} \\ p_i &= u_i + dt_{ji} & v_i &= f(x_i) + bf(q_i) \end{aligned}$$

Passo 11. Teste para condição de parada de atualizações de pesos.

Passo 12. Teste para condição de parada para o número de épocas.

Parâmetros

- n** número de unidades de entrada (camada F_1);
- m** número de unidades de agrupamento (camada F_2);
- a, b** pesos fixos na camada F_1 ; valores de amostra são $a = 10, b = 10$. Ou fixando $a = 0$ ou $b = 0$ produzem instabilidade na rede. A rede não é particularmente sensível aos valores escolhidos.
- c** peso fixo usado em teste para reajuste; um valor de amostra é $c = 0.1$. Um c pequeno dá uma maior faixa do parâmetro de vigilância.
- d** Ativação do vencedor da unidade F_2 ; um valor de amostra é $d = 0.9$. Note que c e d devem ser escolhidos de modo que satisfaça a desigualdade.

$$\frac{cd}{1-d} \leq 1$$

(para impedir para que um reajuste aconteça durante uma tentativa de aprendizagem). A relação deveria ser escolhida perto de 1 para alcançar uma faixa maior para vigilância.

e um parâmetro pequeno introduzido para prevenir divisão por zero quando a norma de um vetor é zero. Este valor impede a normalização para unidade de ser exata.

θ parâmetro de supressão de ruído, um valor de amostra é:

$$\theta = \frac{1}{\sqrt{n}}$$

O valor da amostra pode ser maior que o desejada em algumas aplicações. Componentes do vetor de entrada normalizado (e outros vetores do *loop F₁*) que são menores que este valor são configurados para zero.

α taxa de aprendizagem. Um valor menor reduzirá a velocidade da aprendizagem no modo de aprendizagem rápida ou modo de aprendizagem lenta. Porém, um valor menor assegurará que os pesos (como também a colocação de padrões em agrupamentos) eventualmente alcancem equilíbrio no modo de aprendizagem lento.

p parâmetro vigilância. Junto com os pesos *bottom-up*, este parâmetro determina quantos agrupamentos serão formados. Embora, teoricamente, valores de 0 a 1 são permitidos, só valores entre, aproximadamente, 0.7 e 1 executam qualquer papel útil no controle do número de agrupamentos. (Qualquer valor menor que 0.7 terá o mesmo efeito como se fosse fixado para zero.)

Pesos iniciais

t_{ij}(0) pesos iniciais *top-down* (tem que ser pequeno para assegurar que nenhum reajuste aconteça para o primeiro padrão colocado em uma unidade de agrupamento);

$$t_{ij}(0) = 0$$

b_{ij}(0) pesos iniciais *bottom-up* (deve ser escolhido satisfazer a inequação);

$$b_{ij}(0) \leq \frac{1}{(1-d)\sqrt{n}}$$

para prevenir a possibilidade de um vencedor novo que é escolhido durante a "ressonância" como a mudança de pesos. Valores maiores de *b_{ij}* encorajam a rede para formar mais agrupamentos.

4 APLICAÇÃO DESENVOLVIDA

Quando um cliente visita o consultório odontológico pela primeira vez ou a critério do profissional para fins de atualização, ele preenche uma ficha de anamnese odontológica que tem o objetivo de identificar o perfil do cliente, relativo à sua saúde bucal.

Baseado nas respostas, os clientes são agrupados em 3 grupos:

- ✓ **BOM**: os clientes que cuidam efetivamente da saúde bucal;
- ✓ **MÉDIO**: os clientes que apresentam alguma preocupação com a saúde bucal;
- ✓ **RUIM**: os clientes que preocupam-se esporadicamente com a saúde bucal.

No dia-a-dia do consultório fica ao encargo do profissional a interpretação das respostas apresentadas e a classificação do cliente em um determinado grupo, ou seja o profissional precisa recorrer a sua experiência. Com a aplicação proposta, o tempo dispendido pelo profissional para analisar as fichas de anamnese poderiam ser aproveitados para outros fins. Por exemplo, o cliente poderia responder a ficha de anamnese antes de iniciar a consulta e quando esta ocorrer, o sistema apresentaria uma sugestão de classificação do cliente em um dos grupos definidos.

As 21 perguntas da ficha de anamnese foram elaboradas seguindo cinco enfoques: Alimentação, Higiene, Doenças, Informações e Medicamentos. Cada resposta possui algumas alternativas, sendo as de maior peso (maior influência) dispostas mais a esquerda, devido ao seu alto valor após a conversão de binário para decimal.

HIGIENE			
1 – Quantas vezes por dia pratica higiene oral			
<input type="checkbox"/> Raramente	<input type="checkbox"/> 2 vezes	<input type="checkbox"/> 3 vezes	<input type="checkbox"/> Após refeições
2 – Quando usa o fio dental			
<input type="checkbox"/> Nunca	<input type="checkbox"/> As vezes	<input type="checkbox"/> Regularmente	
3 – Qual tipo de escova dental usada			
<input type="checkbox"/> Dura	<input type="checkbox"/> Média	<input type="checkbox"/> Macia	<input type="checkbox"/> Extra macia
4 – Qual a frequência que troca a escova dental			
<input type="checkbox"/> Menos de 1 mês	<input type="checkbox"/> 2 meses	<input type="checkbox"/> 3 meses ou +	
5 – Quando usa antisséptico bucal			
<input type="checkbox"/> Nunca	<input type="checkbox"/> As vezes	<input type="checkbox"/> Pouco	
6 – Visita regularmente o dentista			
<input type="checkbox"/> Emergência	<input type="checkbox"/> As vezes	<input type="checkbox"/> Regularmente	

Figura 3: Perguntas do enfoque Higiene

As respostas são tratadas como células binárias, ou seja, se um item for respondido com **X**, este receberá o valor **1** e se não for respondido receberá o valor **0**. Se um cliente responder a questão 2 do enfoque Higiene como o exemplo a seguir (Figura 4), então o valor binário da resposta será **100**. Este valor é convertido para decimal e obtém-se o valor **4**. Calcula-se também o valor binário para o caso de todos os itens da resposta serem selecionados, neste caso **111**. Este valor também é convertido para decimal, obtendo-se o valor **7**.

2 – Quando usa o fio dental		
<input checked="" type="checkbox"/> Nunca	<input type="checkbox"/> As vezes	<input type="checkbox"/> Regularmente

Figura 4: Exemplo de resposta da questão 2

Para simplificar os cálculos, estabeleceu-se um coeficiente de redução para que o maior valor fosse 1 e o menor valor próximo de zero. Sendo assim, dividiu-se o valor decimal da resposta pelo valor decimal da resposta com todos os itens selecionados (CANDIDO et. al., 2000).

Neste caso a fórmula para a redução é $4 / 7 = 0,571429$.

Desse modo a matriz de entrada é construída (Tabela 1), onde as linhas representam o número de candidatos entrevistados e as colunas representam o número de questões da ficha de anamnese.

Entrevistados	Questões					
	1	2	3	4	...	21
1	0.133333	0.428571	0.066667	0.285714	...	0.200000
2	0.133333	0.428571	0.066667	0.142857	...	0.400000
...
31	0.266667	0.285714	0.133333	0.285714	...	1.000000

Tabela 1: Matriz de entrada

Após a execução do algoritmo de treinamento os clientes foram agrupados em três grupos, sendo 7 clientes pertencentes ao grupo 1, 11 clientes pertencentes ao grupo 2 e 13 clientes pertencentes ao grupo 3. Após analisar a ficha de anamnese de alguns clientes o profissional identificava a qual grupo este pertencia. Assim foi possível estabelecer que o grupo 1 refere-se ao grupo MÉDIO, o grupo 2 refere-se ao grupo BOM e o grupo 3 refere-se ao grupo RUIM.

5 CONCLUSÕES E PROPOSTAS FUTURAS

Este trabalho demonstrou a utilização das redes neurais, mais especificamente a ART2 como meio de seleção e agrupamento de clientes de um consultório odontológico. Tarefa esta que é tipicamente executada pelo profissional. Algumas melhorias e acréscimos podem ser sugeridos para este trabalho.

Por exemplo, a partir dos agrupamentos formados, pode-se aplicar outros tipos de ferramentas de inteligência artificial para a identificação do padrão dos conjuntos agrupados. Com isto seria possível conhecer quais características tem maior influência na formação dos grupos.

Outro tópico a ser estudado é a relação de uma resposta com outra, por exemplo identificar qual das situações a seguir tem mais influência na classificação em um grupo: se o cliente ingere muito alimento, mas a quantidade em nutrientes é equilibrada ou se o cliente ingere pouco alimento mas faz 5 refeições por dia e ingere grande quantidade de açúcar.

6 BIBLIOGRAFIA

- CANDIDO, A., CERUTTI, F., JUNIOR, O. BASTOS, L, **Aplicação de Redes Neurais na seleção de estagiários para a ciência da computação**, Aplicações de Redes Conexionistas – PPGEP/UFSC, 2000.
- FAUSETT, L. **Fundamentals of Neural Networks – Architectures, Algorithms, and Applications**. New Jersey: Prentice-Hall, 1994
- RABUSKE, R. **Inteligência Artificial**. Florianópolis: Ed. UFSC, 1995.
- RICH, E. KNIGHT, K. **Inteligência Artificial**. São Paulo: Makron Books, 1993.
- RUSSEL, S. J., NORVIG, P. **Artificial Intelligence: a modern approach**. New Jersey: Prentice-Hall, 1995