

•
•
•
•
•

Teoria da Ressonância Adaptativa

ART - *Adaptive Resonance Theory*

Aluizio Fausto Ribeiro Araújo

Universidade Federal de Pernambuco

Centro de Informática



• • • • •

Conteúdo

- Introdução;
- ART1 e ART2:
 - Características;
 - Topologia;
 - Algoritmo;
 - Aplicações;
- Referências.

Introdução

- Motivação: O que se deseja com a rede ART?
 - Um algoritmo que se adapte em resposta a estímulos relevantes e não mude para estímulos irrelevantes.
 - Um algoritmo capaz de mudar de um modelo de atuação estável para um outro plástico sem deixar de convergir.
 - Um algoritmo capaz de aprendizagem incremental (ou online).
 - Um algoritmo que preserve conhecimento prévio aprendido quando adiciona novo conhecimento.
 - Um algoritmo plástico adapta seu comportamento considerando novos e relevantes eventos.
 - Um algoritmo estável preserva comportamento após eventos irrelevantes.

Introdução

- Motivação: Dilema plasticidade X estabilidade
 - Afirmção de Grossberg: A maioria dos algoritmos de redes neurais existentes são fixos e estáveis ou plásticos e instáveis.
 - Por exemplo, a MLP-BP em sua forma estável, depois de seus parâmetros livres terem convergido, não é plástica, isto é, não pode se adaptar para responder a novos estímulos do meio ambiente.
 - Em sua forma plástica, a MLP-BP pode ser retreinada para novas entradas. Esta forma é instável pois haverá perda de conhecimento anterior. Portanto, o algoritmo é plástico pois modifica informação anterior mas não é estável pois perde conhecimento anterior.

Introdução

- Características gerais da família ART:
 - Propostas por Grossberg e associados desde 1987;
 - É um algoritmo auto-organizável para agrupamento com aprendizagem on-line;
 - É uma rede neural recorrente;
 - Possui camada de saída competitiva;
 - Trata da dicotomia: estabilidade versus plasticidade;
 - A rede é provadamente convergente.

Introdução

- Arquitetura:
 - Duas camadas processáveis;
 - Conexões entre nodos de diferentes camadas;
 - Conjunto de pesos entrada-saída diferente do conjunto de pesos saída-entrada;
 - Existência de unidades de monitoramento e controle de funções da rede;
 - Conhecimento de memória de longo prazo está nos pesos da rede;
 - Conhecimento de memória de curto prazo está nos padrões de ativação da rede.

Introdução

- Parte central do algoritmo:

Apresente novo padrão de entrada

Determine o protótipo mais próximo

Se padrão for próximo o suficiente do protótipo vencedor

 Ajuste os pesos do vencedor

De outro modo

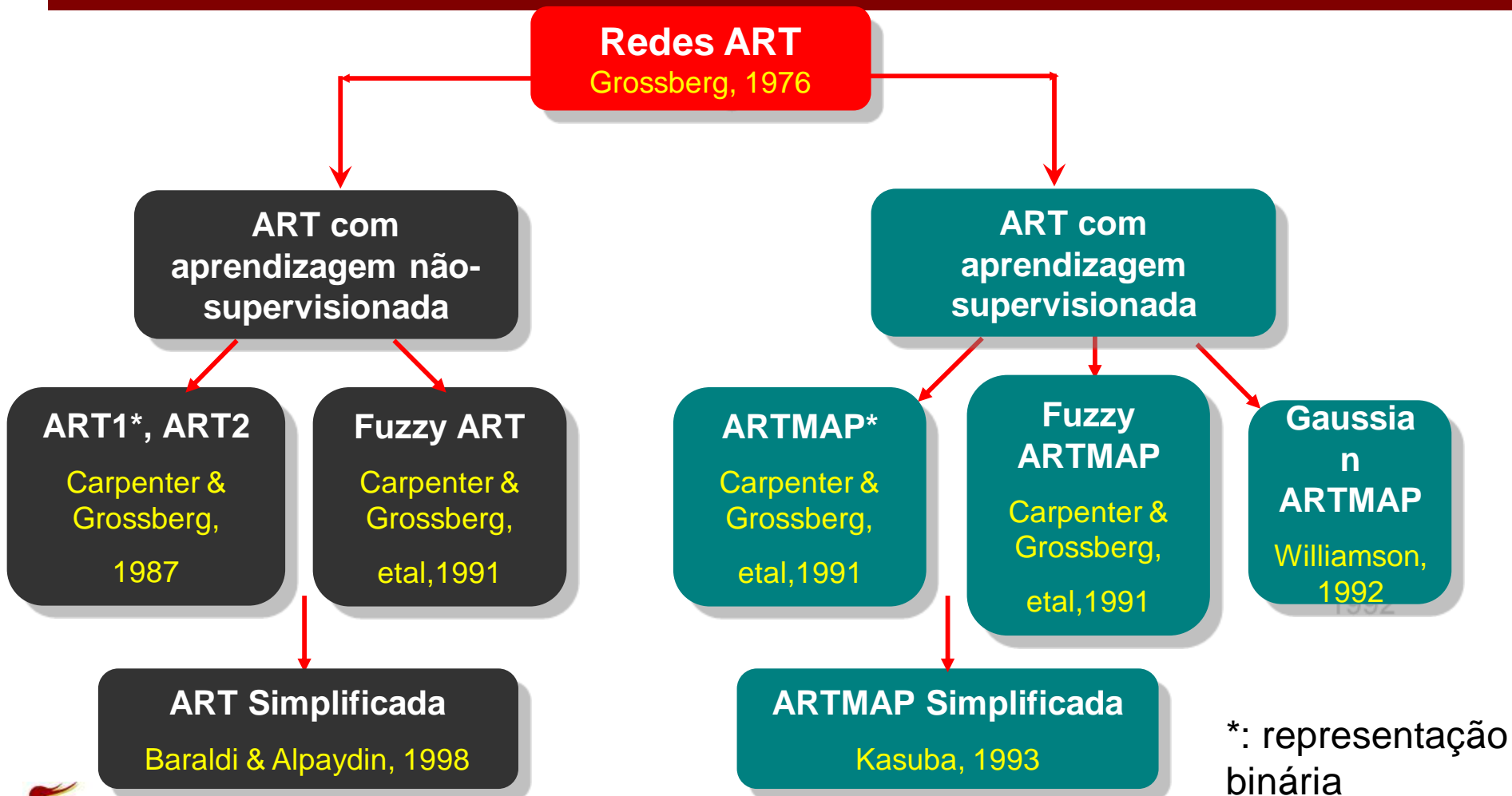
 Inicialize novo agrupamento

Introdução

- ART se constitui em uma família de redes neurais que trata do dilema estabilidade plasticidade. Isto é, a rede segue aprendendo sem esquecer conhecimento anteriormente aprendido. Além disto, a rede pode realizar aprendizagem online.
- A principal inovação da ART envolve a idéia de expectativa:
 - Uma dada entrada será considerada como de um grupo apenas se tiver grau de semelhança pré-estabelecida com o protótipo daquele grupo.
- A família ART pode realizar treinamento não-supervisionado ou supervisionado.

Introdução

Principais Redes ART



Introdução

Principais Redes ART

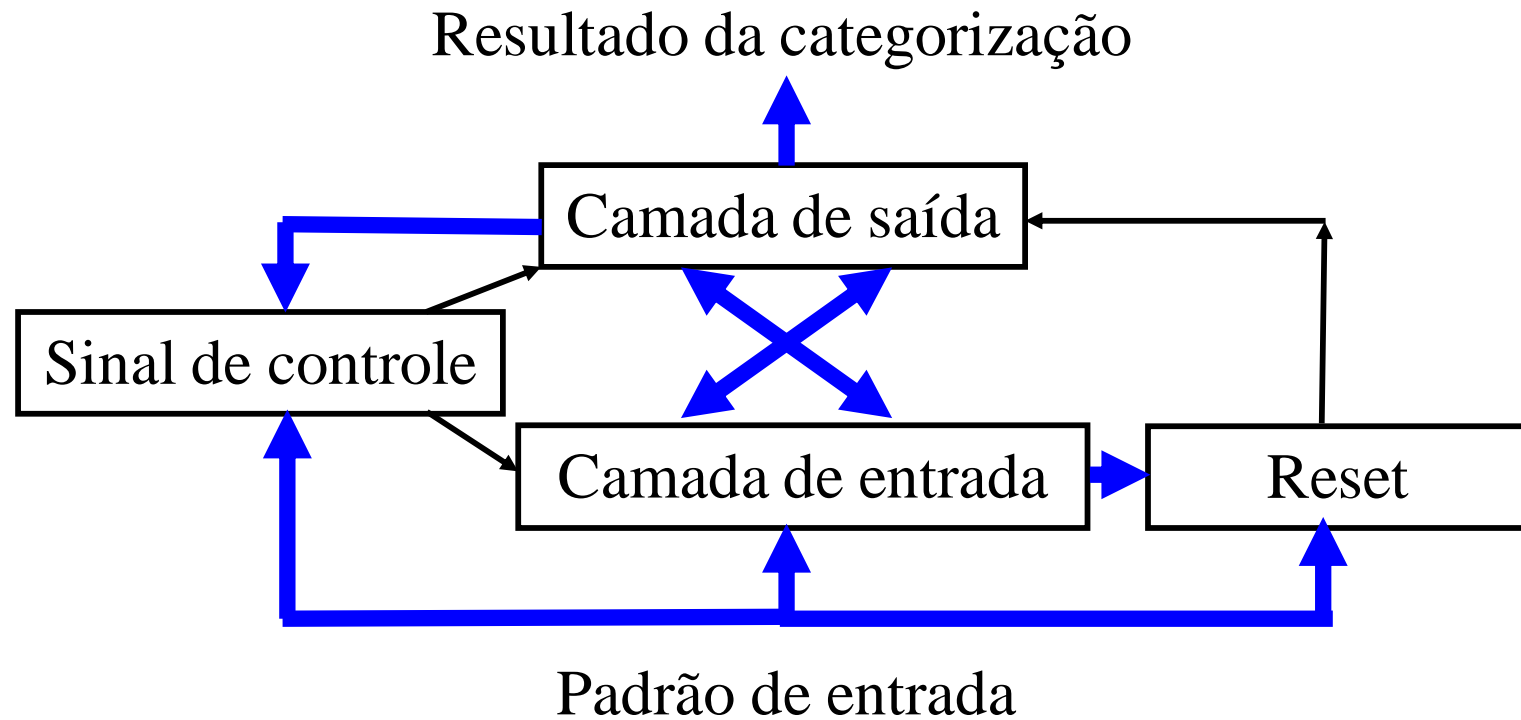
- ART1: Agrupamento de vetores binários;
- ART2: Agrupamento de vetores com variáveis reais;
- ARTMAP: Versão supervisionada de ART1, modelo original, ou ART2;
- Fuzzy ART: Síntese de ART e sistema *fuzzy*;
- Fuzzy ARTMAP: Art supervisionado *fuzzy*;
- ARTMAP gaussiano: diminui o número de categorias em ambientes ruidosos;

ART1

- Características:
 - Um padrão integra um agrupamento se estiver próximo o suficiente do protótipo deste grupo;
 - Poder trocar modos de operação estável e plástico;
 - Sensibilidade a contexto, para diferenciar informações irrelevantes ou repetidas;
- Arquitetura:
 - Camadas com funções diferentes;
 - Partes externas às camadas controlam o fluxo de dados na rede.

ART1

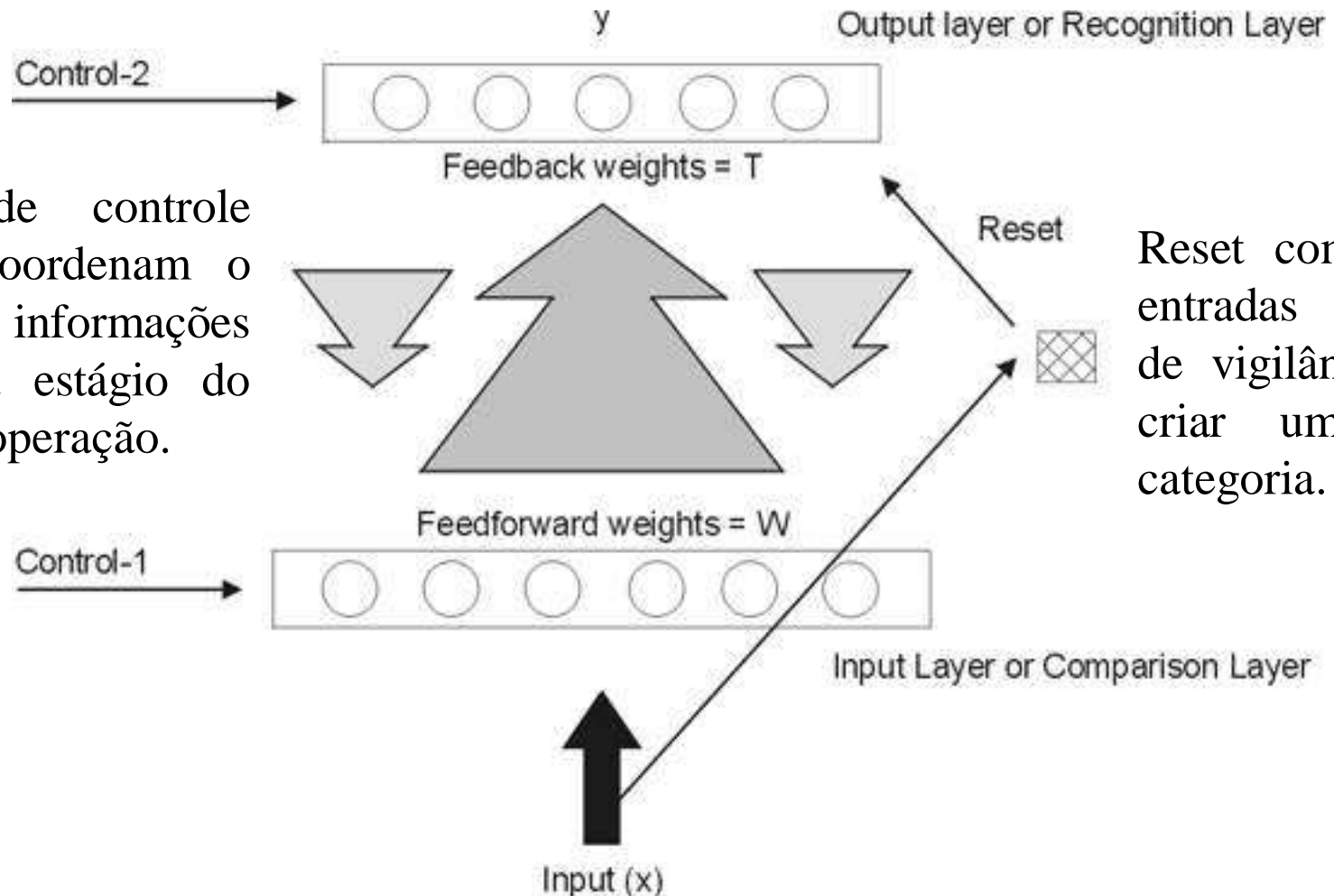
Diagrama de Blocos



ART1

Topologia

Sinais de controle lógico coordenam o fluxo de informações em cada estágio do ciclo de operação.



Reset compara as entradas a limiar de vigilância para criar uma nova categoria.

ART1

Topologia

- Módulo reset:
 - Conexões com valores fixos;
 - Realiza o teste de vigilância;
 - Desabilita cumulativamente nodo vencedor de saída caso seu protótipo não seja suficientemente próximo do estímulo de entrada;
 - O reset permanece ativo até novo padrão ser apresentado.
- Módulo de controle:
 - Conexões com valores fixos;
 - Controla o ciclo de ativação da camada de entrada;
 - Possui dois sinais de controle que trabalham sincronizados.

ART1

Operação

- Possui as seguintes fases:
 - Fase de Inicialização;
 - Fase de Reconhecimento;
 - Fase de Comparação;
 - Fase de Busca.
- A rede trabalha com o fluxo de informação para frente e para trás até atingir o estágio de ressonância.

ART1

Operação: Fase de Inicialização

- Cinco componentes são inicializados:
 - Sinal de controle 1 (c_1): Determina se a camada funciona como entrada (propagação) ou como comparação;
 - Sinal de Controle 2 (c_2): Habilita a operação na camada de reconhecimento;
 - Conexões entrada-saída (**W**): $w_{ji} = (1/(1 + N_{ent})) + \Delta$, onde N_{ent} é o número de nodos de entrada e Δ é um incremento aleatório.
 - Conexões saída-entrada: (**T**): $t_{ji} = 1, \forall j, i$;
 - Limiar de vigilância (l): valor no intervalo $0 < l < 1$.

ART1

Operação: Fase de Reconhecimento

- Transmissão para frente via pesos entrada-saída (\mathbf{W});
- Cálculo do produto interno entre a entrada e cada vetor componente da matriz \mathbf{W} ;
- Determinação do nodo vencedor na camada de saída;
- Nodo vencedor na camada de saída responde;
- Vetor de pesos associado ao vetor vencedor possui protótipo do grupo;
- Parte dos bits do protótipo correspondem àqueles no padrão de entrada.

ART1

Operação: Fase de Reconhecimento

- Cada vetor de conexões, w_j , da unidade j da camada de saída pode ser visualizado como aproximação de um protótipo de uma categoria;
- O vetor de entrada é comparado aos protótipos associados a cada nodo de saída. O melhor casamento (*match*) é o maior valor de produto interno entre o vetor de entrada e o vetor de peso do nodo. A atividade no nodo j é calculada:

$$a_j = \sum_{i=1}^{Nent} w_{ji} o_i, \quad Nent \text{ é o número de unidades de entrada.}$$

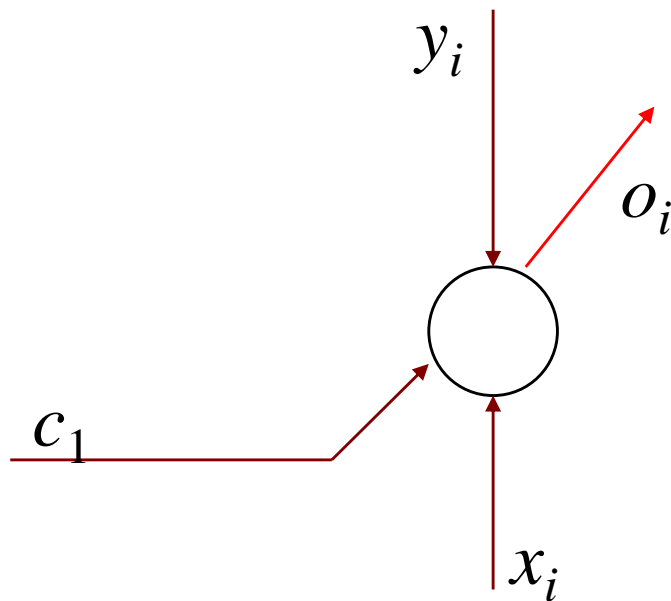
- Apenas um nodo de saída, o vencedor (v), terá saída 1 enquanto os demais nodos terão saída zero:

$$o_v = 1, \quad \text{para } v = \max_j [a_j]$$

ART1

Operação: Fase de Reconhecimento

- Saída de cada nodo de entrada i é calculada pela regra dos 2/3:



- Três sinais de Entrada:
 - A componente x_i do vetor de entrada;
 - O sinal de realimentação da camada de saída y_i ;
 - O sinal de controle c_1 .

$$o_i = \begin{cases} 1 & \text{se } y_i \cdot c_1 = 1 \text{ ou } y_i \cdot x_i = 1 \text{ ou } x_i \cdot c_1 = 1 \\ 0 & \text{de outra forma} \end{cases}$$

ART1

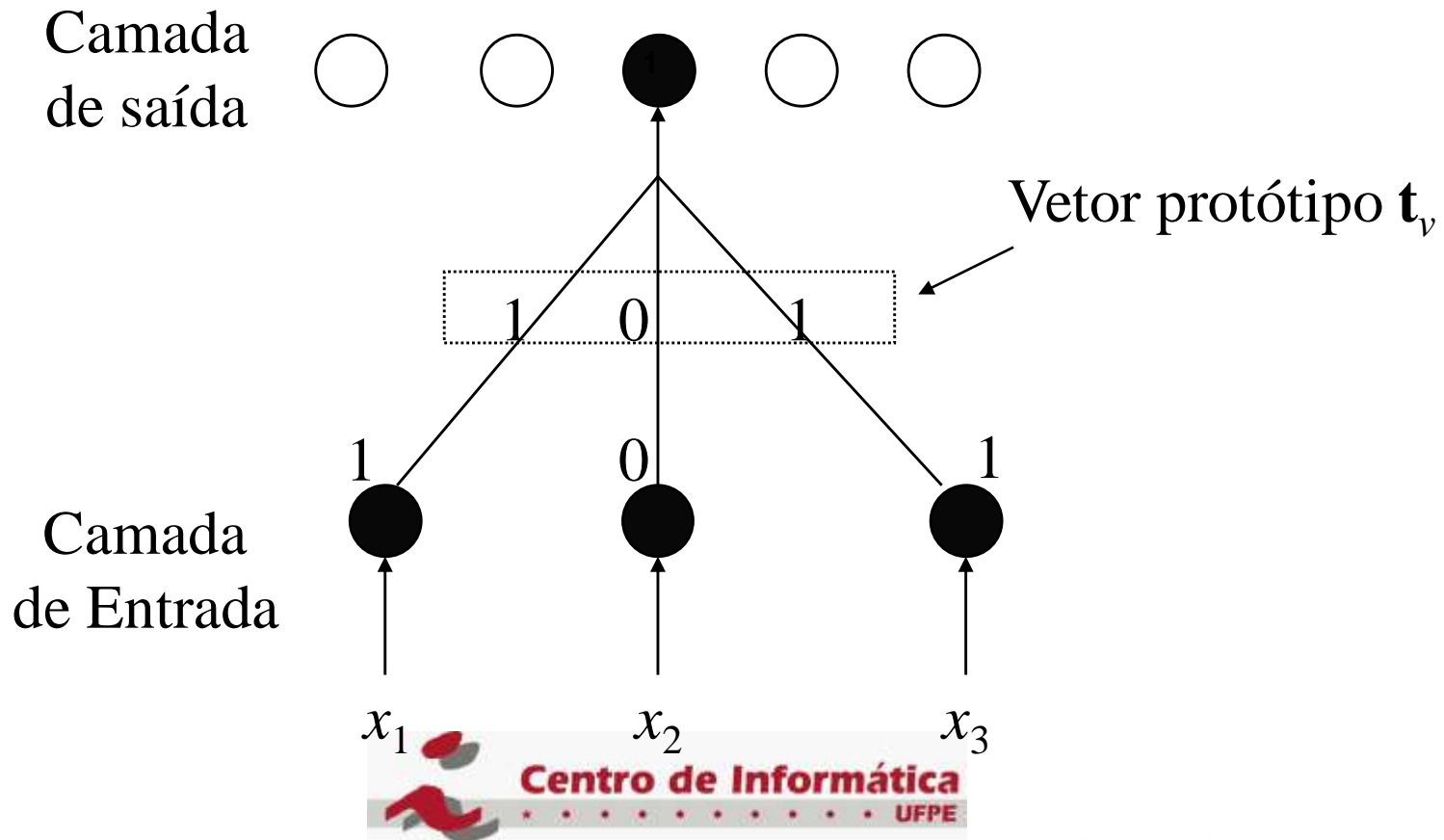
Operação: Fase de Comparação

- Transmissão para trás via pesos saída-entrada (**T**);
- Teste de vigilância para avaliar quanto o protótipo se casa com o vetor de entrada;
- A categorização alcança sucesso e a rede entra em ressonância se o protótipo e o vetor de entrada forem suficientemente próximos;
- Se não houver proximidade suficiente, deve buscar outro nodo para representar a categoria do vetor de entrada, na fase de busca.

ART1

Operação: Fase de Comparação

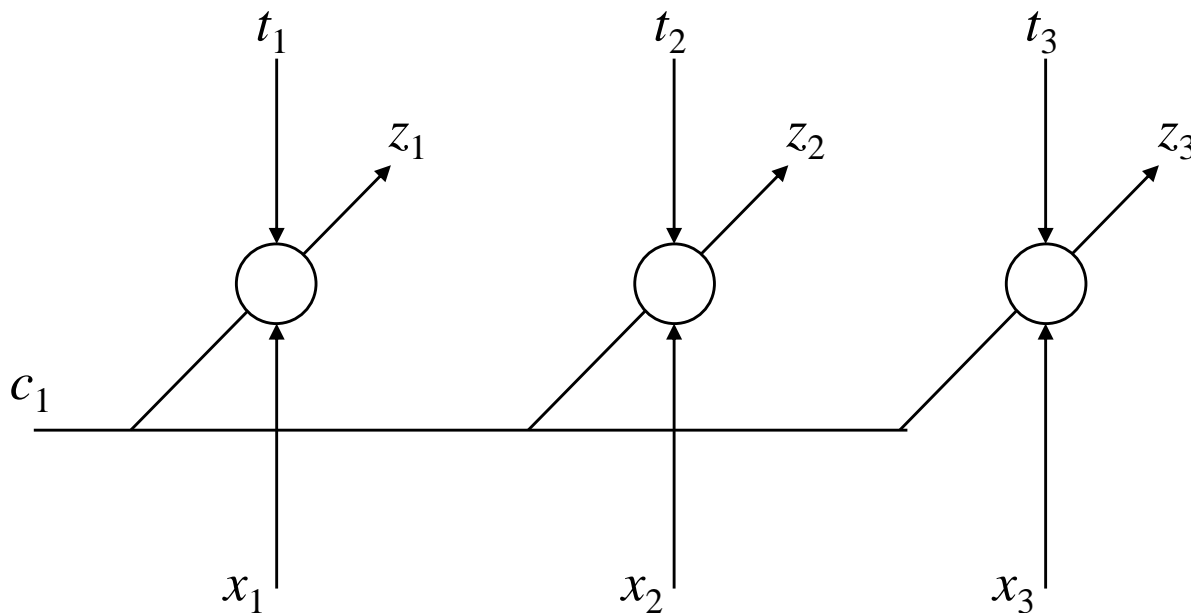
- Como na propagação para frente, na realimentação da unidade de saída ativa v é propagada multiplicando-a pelo vetor \mathbf{t}_v :



ART1

Operação: Fase de Comparação

- Para sinal de controle $c_1=0$, apresentam-se à camada de entrada (no modo de comparação) o sinal re-alimentado (protótipo do nodo vencedor) e o sinal de entrada. Aplica-se a regra dos 2/3:



Vetor de entrada

O vetor comparação (Z) é entregue ao circuito reset junto com o vetor atual de entrada (X).

ART1

Operação: Fase de Comparação

- O circuito reset testa a similaridade entre os vetores comparação e de entrada através da razão entre estes valores (S):

$$\text{Se } S = \frac{\sum_{i=1}^{N_{ent}} z_i x_i}{\sum_{i=1}^{N_{ent}} x_i} > l,$$

então categorização realizada

de outro modo, inicie fase de busca

ART1

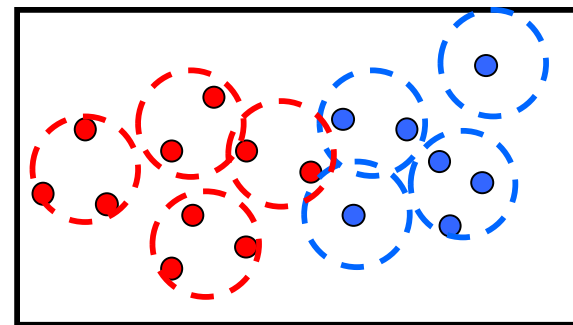
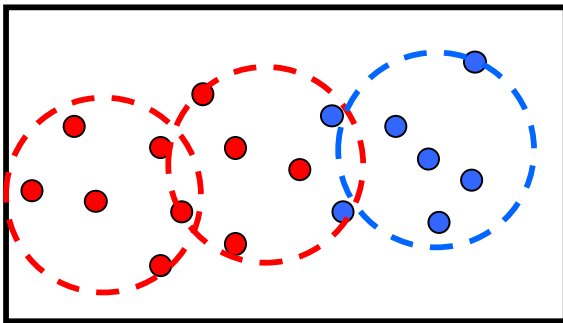
Operação: Fase de Busca

- O objetivo é achar um novo casamento na camada de reconhecimento para a entrada atual. Isto ocorre como se segue:
 - A camada de saída é “resetada” (encerra o ciclo de comparação e força o sinal de controle c_2 para um);
 - A entrada é reapresentada à camada de reconhecimento e a determinação do melhor casamento é refeita;
 - A rede entra outra vez na fase de comparação;
 - O processo é repetido desabilitando os nodos na camada de saída, até achar nodo na camada de reconhecimento que se case com a entrada dentro do limiar de vigilância.

ART1

Papel do Limiar de Vigilância

- O limiar determina a granularidade do agrupamento.
- O limiar estabelece a bacia de atração de cada protótipo.
- Limiar baixo leva a criação de poucos agrupamentos, maior probabilidade de errar na categorização.
- Limiar alto provoca criação de muitos grupos, maior precisão na categorização.



ART1 Algoritmo

- Passo 0 : Inicialização de parâmetros e dos pesos

$$L > 1$$

$$0 < l \leq 1$$

$$0 < w_{ji}(0) < \frac{L}{L-1+Nent}$$

onde L é uma constante (usualmente, L=2), tipicamente,

$$w_{ji} = (1/(1+Nent)) + \Delta,$$

$$t_{ji}(0) = 1$$

ART1

Algoritmo

- Passo 1: Enquanto o critério de parada for falso faça passos 2 a 13
- Passo 2: Para cada padrão de entrada faça os passos 3 a 13
- Passo 3: Atribua zero para todas unidades da camada de saída e o valor do vetor de entrada para as unidades de entrada
- Passo 4: **Normalize o vetor de entrada considerando $\|\mathbf{x}\|$**
- Passo 5: Para cada nodo da camada de saída que não esteja inibido, calcule

$$a_j = \sum_{i=1}^{N_{ent}} w_{ji} x_i$$

ART1

Algoritmo

- Passo 6: Enquanto reset for verdadeiro, faça passos 7 a 11
- Passo 7: Determine o vencedor: $v = \max_j [a_j]$
- Passo 8: Atribua $y_v = 1$ para o nodo com ativação a_v
- Passo 8: Calcule a variável de comparação $z_i = x_i \cdot t_i$
- Passo 9: Calcule
$$S = \left(\frac{\sum_{i=1}^{N_{ent}} z_i x_i}{\sum_{i=1}^{N_{ent}} x_i} \right)$$
- Passo 10: Realize o teste do parâmetro de vigilância:
Se $S < l$ então resete o nodo vencedor
vá para o passo 6
- Passo 11: Libere todos nodos na camada de saída

ART1

Algoritmo

- Passo 12: Atualize os pesos do vencedor

$$t_{vi}(t+1) = t_{vi}(t).x_i$$

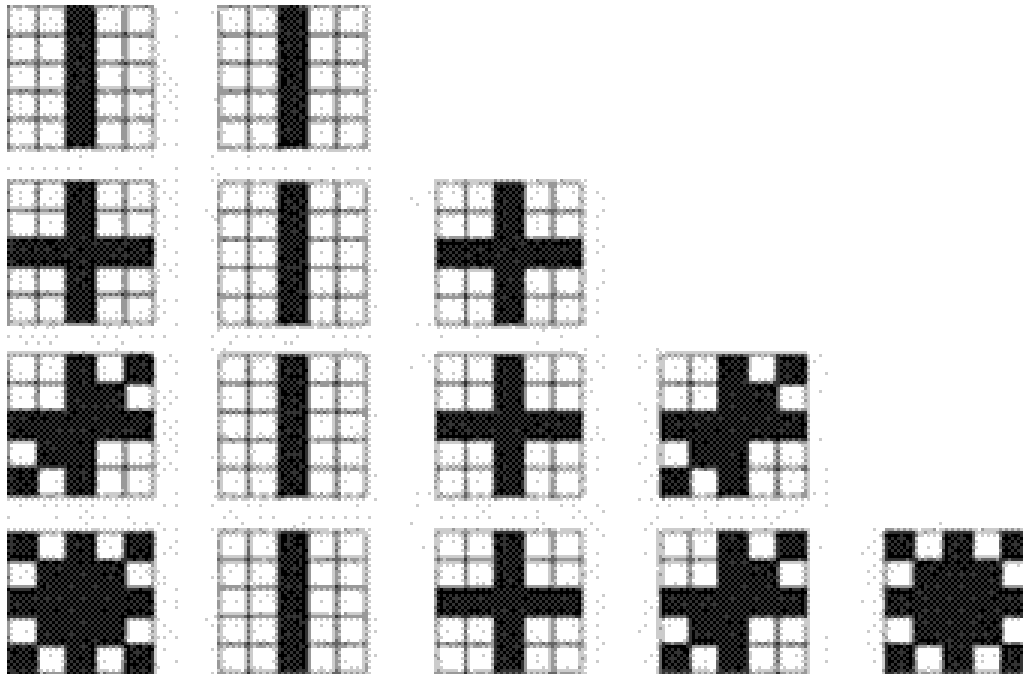
$$w_{vi}(t+1) = \frac{t_{vi}(t+1).x_i}{0,5 + \sum_{i=1}^{N_{ent}} t_{vi}(t).x_i}$$

- Passo 13: Teste a condição de parada

ART1

Exemplo

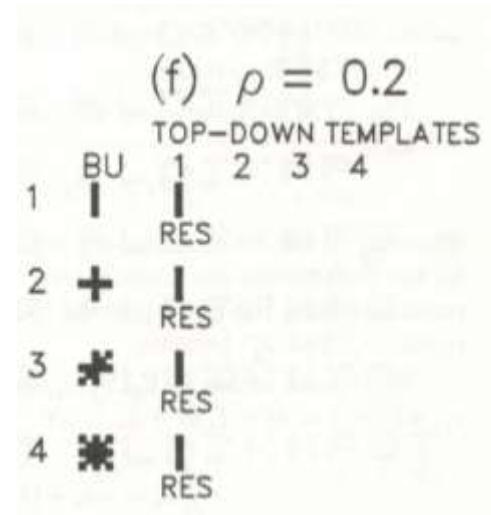
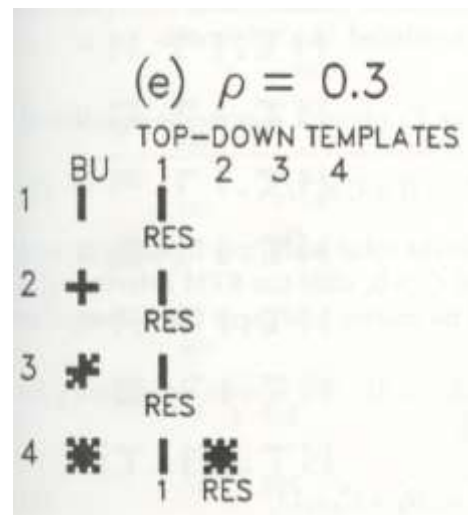
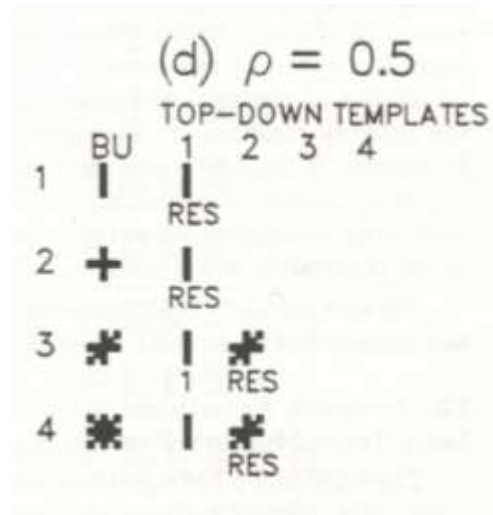
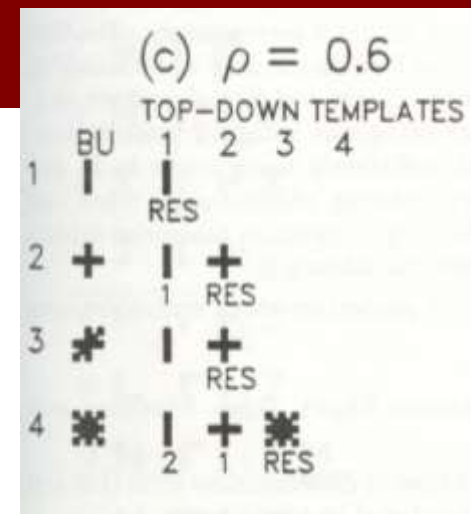
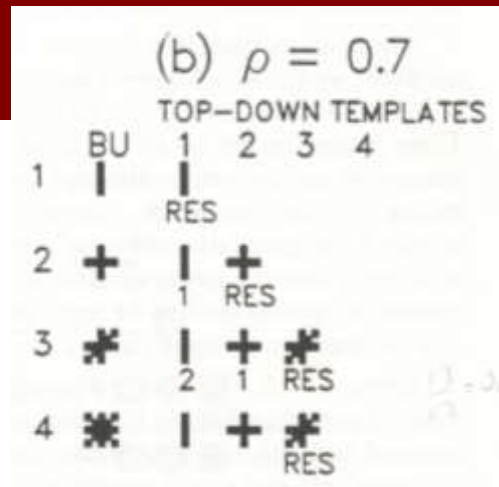
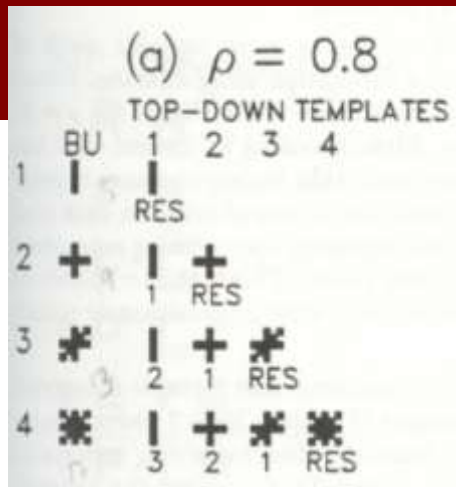
- ART1 apresenta alta sensibilidade à variação de parâmetros em especial quando se trata do parâmetro de vigilância.



$$l=0.8$$

Mais próximo de 1,
considera detalhes mais
atentamente

ART1



[illegible]

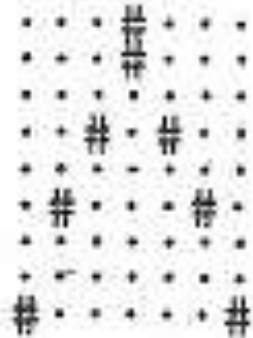
ART1

Exemplo

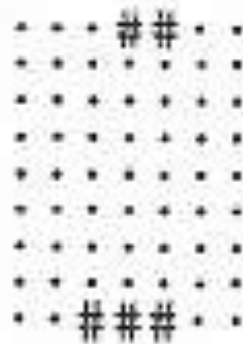
- Valores iniciais dos parâmetros: $l=0,3$; $L=2$; $m=10$.
- Ordem de apresentação dos padrões: A1, A2, A3, B1, B2, B3,..., K1, K2, K3.
- Agrupamento padrões
- 1 A1, A2
- 2 A3
- 3 C1, C2, C3, D2
- 4 J1, J2, J3
- 5 K2
- 6 B1, D1, E1, K1, B3, D3, E3, K3
- 7 B2, E2

ART1

Exemplo



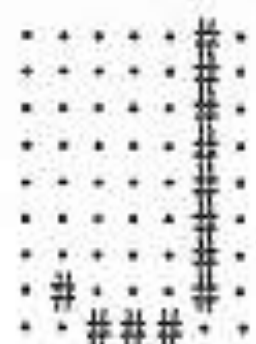
Cluster 1



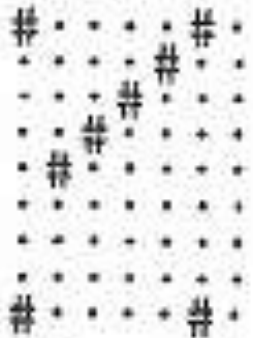
Cluster 2



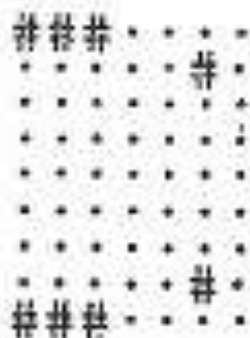
Cluster 3



Cluster 4



Cluster 5



Cluster 6



Cluster 7

ART1

Discussão

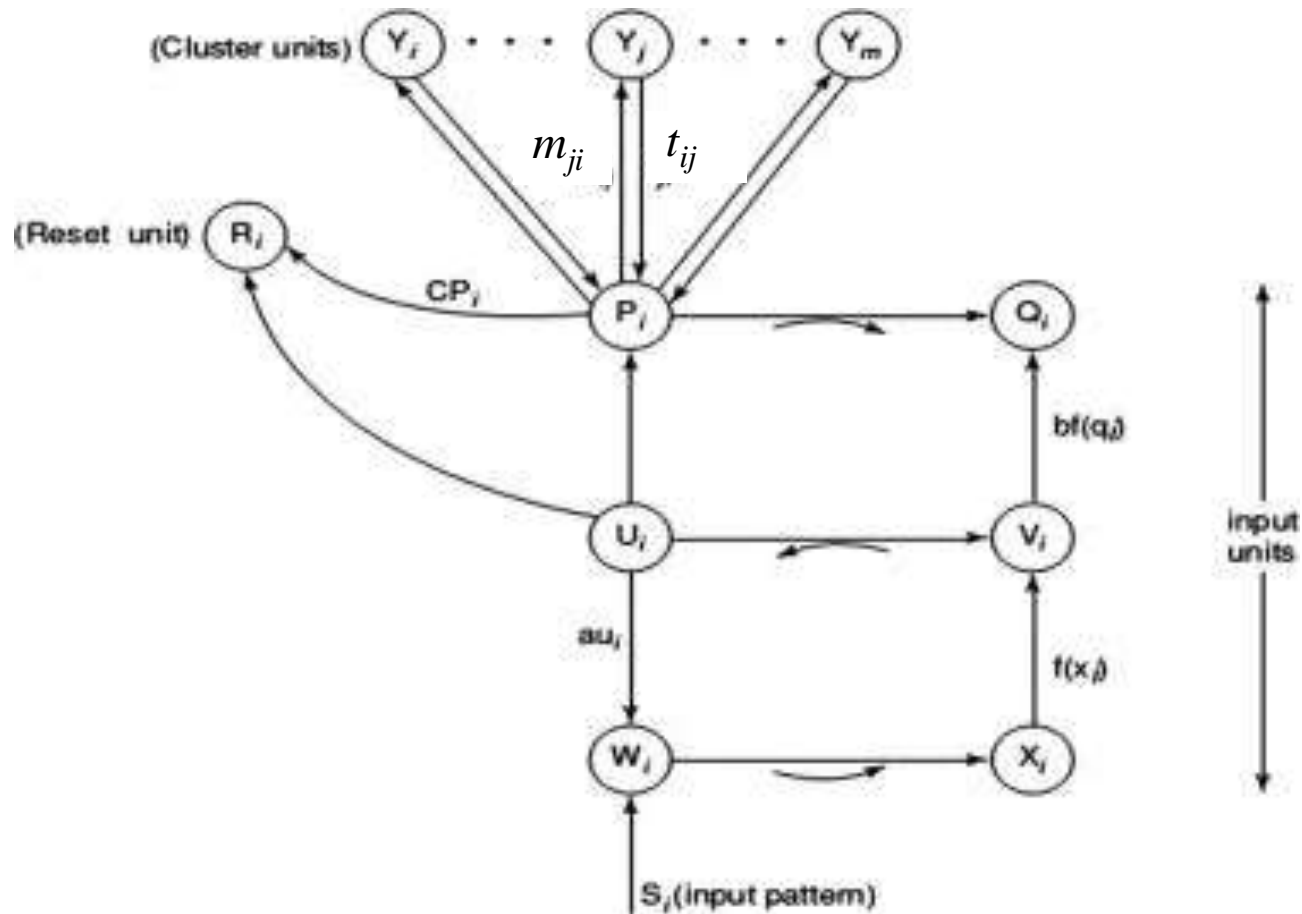
- O algoritmo de aprendizagem é rápido;
- Conhecimento adquirido pode ser recuperado;
- Dificuldade de ajustar o limiar de vigilância adequado;
- Baixa robustez a ruído;
- ART1 é sensível à ordem de apresentação dos dados;
- A precisão pode não ser boa;
- Só atualiza o vencedor, construindo mapa mais ponto a ponto que o SOM.

ART2

- Características:
 - Categorização não-supervisionada para vetores de entrada com valores reais ou vetores binários com ruído;
 - A atualização dos protótipos considera sua adaptação na direção do novo exemplo por ele categorizado.
 - Considera combinação entre normalização e supressão de ruídos (pré-processamento):
 - Dados inicialmente normalizados e valores abaixo de um dado limiar são desconsiderados.
- Arquitetura:
 - Pré-processamento de entrada;
 - Três camadas de entrada (denominadas F1);
 - Uma camada de saída (F2);
 - Partes externas às camadas controlam o fluxo de dados na rede.

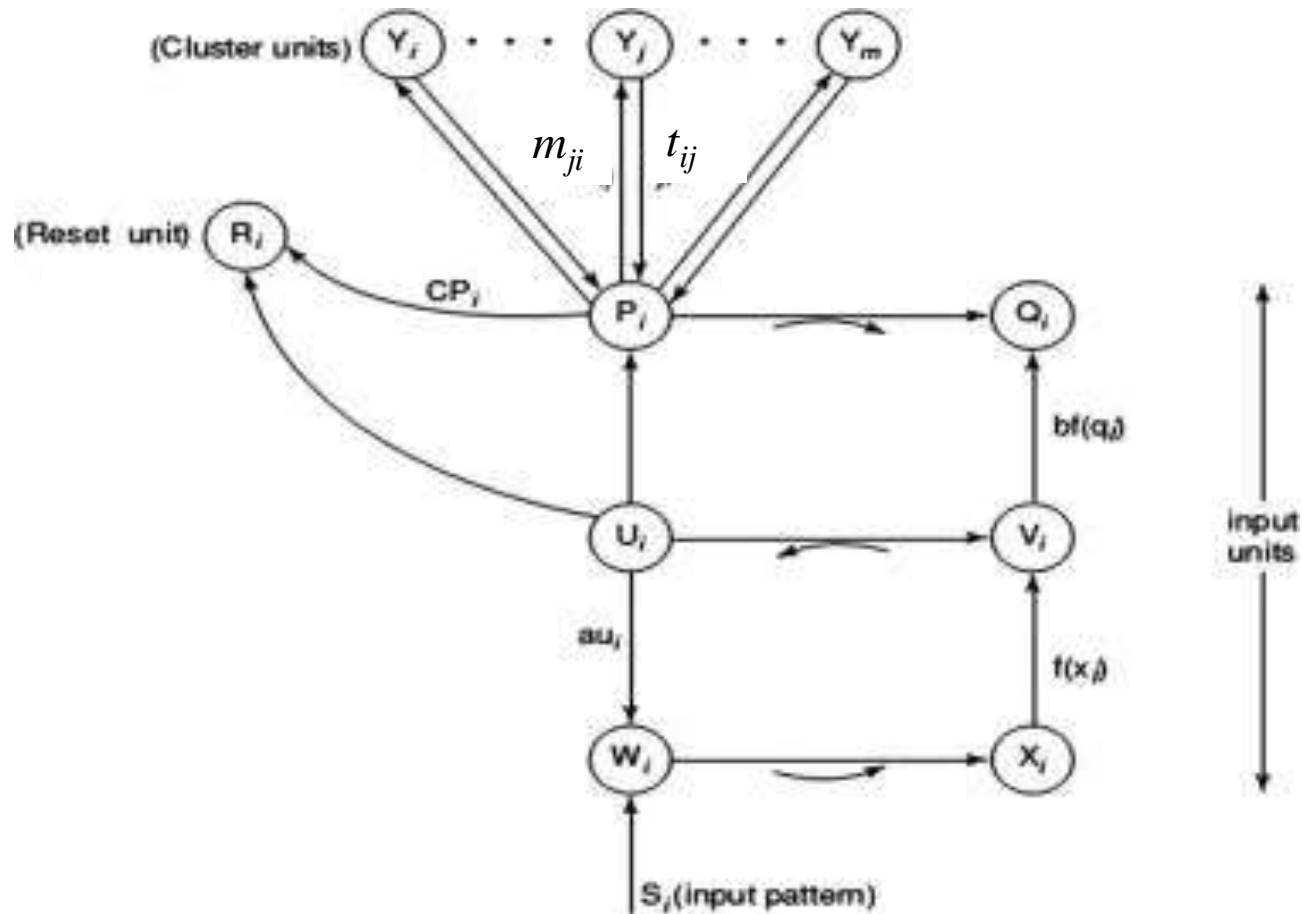
ART2

Topologia



ART2

Topologia



ART2

Algoritmo

- Passo 0 : Inicialização de parâmetros e pesos:

a, b, c, d : valores reais finitos ;

e : valor real pequeno;

$0 < \theta$: Limiar para ativação;

N_{ep}, N_{ita} : números de épocas e número de iterações de aprendizado ;

$0 < \alpha < 1$: taxa de aprendizagem;

$0 < l < 1$: limiar de vigilância;

$-1 < m_{ji}, t_{ij} < 1$: pesos das matrizes.

- Passo 1: Faça passo 2 ao 12 enquanto o número de épocas for menor que N_{ep} .

ART2

Algoritmo

- Passo 2: Para cada vetor de entrada s , faça os passos 3 ao 11.
- Passo 3: Atualize as ativações das unidades na camada de entrada (F1), para pré-processamento (normalização e desprezar valores pequenos de ativação):

$$u_i = 0; \quad p_i = 0; \quad q_i = 0;$$

$$w_i = s_i;$$

$$x_i = \frac{s_i}{e + \|s\|};$$

$$v_i = f(x_i) = \begin{cases} x_i, & x_i > 0 \\ 0, & \text{de outra forma.} \end{cases}$$

ART2

Algoritmo

- Passo 3 (continuação): Atualize as ativações das unidades na camada de entrada (F1) outra vez:

$$u_i = \frac{v_i}{e + \|\mathbf{v}\|};$$

$$p_i = u_i;$$

$$q_i = \frac{p_i}{e + \|\mathbf{p}\|};$$

$$w_i = s_i + a u_i;$$

$$x_i = \frac{w_i}{e + \|\mathbf{w}\|};$$

$$v_i = f(x_i) + bf(q_i).$$

ART2

Algoritmo

- Passo 4: Calcule as saídas das unidades da camada de saída (F2):

$$y_j = g(a_j) = g\left(\sum_{i=1}^{N_{ent}} m_{ji} p_i\right)$$

- Passo 5: Enquanto reset for verdadeiro, faça passos 6 a 7.
- Passo 6: Determina o vencedor: $v = \max_j [y_j]$
- Passo 7: Teste o parâmetro de vigilância (l):

$$u_i = \frac{v_i}{e + \|\mathbf{v}\|}; \quad p_i = u_i + d t_{iv}; \quad \text{logo} \quad r_i = \frac{u_i + c p_i}{e + \|\mathbf{u}\| + c \|\mathbf{p}\|};$$

ART2

Algoritmo

- Passo 7 (continuação): Teste o parâmetro de vigilância (l):

Se $\|\mathbf{r}\| < l - e$

então $y_v = -1$ (inibição da unidade v)

vá para o passo 5;

de outra forma

$$w_i = s_i + a u_i; \quad x_i = \frac{w_i}{e + \|\mathbf{w}\|}; \quad q_i = \frac{p_i}{e + \|\mathbf{p}\|};$$

$$v_i = f(x_i) + b f(q_i);$$

Desabilite o reset;

Vá para passo 8.

ART2

Algoritmo

- Passo 8: Faça os passos 9 a 11 enquanto o número de iterações de aprendizado for menor que N_{ite} :
- Passo 9: Atualize o peso do vencedor (v):

$$t_{iv} = \alpha d u_i + [1 + \alpha d (d - 1)] t_{iv}$$

$$m_{iv} = \alpha d u_i + [1 + \alpha d (d - 1)] m_{iv}$$

- Passo 10: Atualize as ativações das unidades de F1:

$$u_i = \frac{v_i}{e + \|\mathbf{v}\|}; \quad w_i = s_i + a u_i; \quad p_i = u_i + d t_{iv};$$

$$x_i = \frac{w_i}{e + \|\mathbf{w}\|}; \quad q_i = \frac{p_i}{e + \|\mathbf{p}\|}; \quad v_i = f(x_i) + b f(q_i).$$

ART2

Algoritmo

- Passo 11: Fim do laço para o número de iterações de aprendizado.
- Passo 12: Fim do laço para o número de épocas.

ART2

Exemplo

Input from
Font 1

D1

A1

E1

B1

J1

C1

K1

Input from
Font 2

D2

A2

E2

B2

J2

C2

K2

Input from
Font 3

D3

A3

E3

B3

J3

C3

K3

ART2

Exemplo

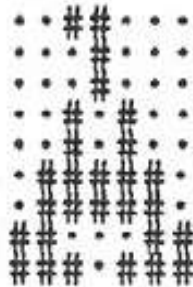
- Valores iniciais dos parâmetros: $a=10$; $b=10$; $c=0,1$; $d=0,9$; $\theta=0,126$; $l=0,8$.
- Ordem de apresentação dos padrões: A1, A2, A3, B1, B2, B3,..., K1, K2, K3.

Agrupamento	padrões
1	A1, A2;
2	A3;
3	C1, C2, C3, D2;
4	B1, D1, E1, K1, B3, D3, E3, K3;
5	K2;
6	J1, J2, J3;
7	B2, E2.

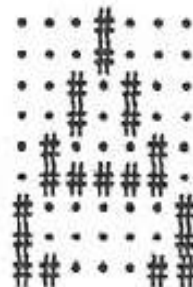
ART2

Exemplo

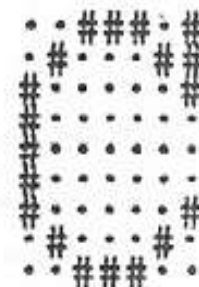
Cluster 1



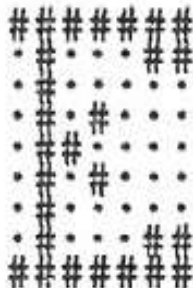
Cluster 2



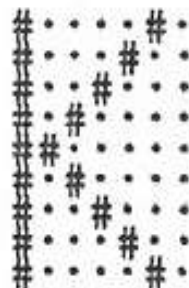
Cluster 3



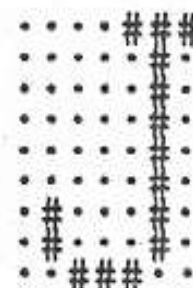
Cluster 4



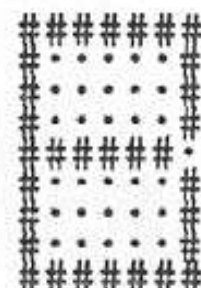
Cluster 5



Cluster 6



Cluster 7



ART

Aplicações

- Associação condição-ação;
- Mineração de dados:
 - Particionamento de conjunto de dados; detecção de agrupamentos emergentes; mercado de ações;
- Processamento de linguagem natural:
 - Agrupamento de documentos; recuperação (*retrieval*) de documentos; questionamento automático;
- Processamento de sinais:
 - Segmentação de imagens; produção de discurso;
- Reconhecimento de padrões:
 - Classificação de solo; reconhecimento de caracteres; reconhecimento de faces; verificação de assinaturas; diagnóstico médico;
- Robótica:
 - Navegação e controle;

Referências

- Beale, R. & Jackson, T. (1991). *Neural Computing: An Introduction*, IOP.
- da Silva, L; E. B., Elnabarawy, I., & Wunsch II, D. C. (2019). A survey of adaptive resonance theory neural network models for engineering applications. *Neural Networks*, 120: 167-203.
- Carpenter, G.A. & Grossberg, S. (1987). ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, 26 (23) 4919-4930.
- Carpenter, G.A. & Grossberg, S. (1990). ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks*, 3: 129-152
- Carpenter, G.A., Grossberg, S., & Rosen, D.B. (1991a). ART 2-A: An adaptive resonance algorithm for rapid category learning and recognition. *Neural Networks*, 4: 493-504.
- Carpenter, G.A., Grossberg, S., & Rosen, D.B. (1991b). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, 4: 759-771.

Referencias

- Carpenter, G.A., Grossberg, S., & Reynolds, J.H. (1991). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, 4: 565-588.
- Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., & Rosen, D.B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, 3: 698-713.
- Carpenter, G.A. & Grossberg, S. (2003). Adaptive resonance theory. In Michael A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks*, Second Edition, 87-90, Cambridge, MA: MIT Press.
- Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science*, 11, 23-63.
- Tauritz, D. R. & Hein, L. G. (1995). *Adaptive Resonance Theory (ART): An Introduction*. In Technical Report, 95-35, Leiden University.