

5 Redes Neurais Artificiais

5.1. Introdução

A motivação original desta metodologia¹ foi a tentativa de modelar a rede de neurônios humanos visando compreender o funcionamento do cérebro. Portanto, como o próprio nome da metodologia revela, sua motivação inicial foi a de realizar tarefas complexas que o cérebro executa com elevada efetividade (por exemplo: reconhecimento de padrões, percepção e controle motor) através da simulação de seu funcionamento.

Segundo Haykin (1998), uma rede neural artificial (RNA) é um sistema de processamento massivamente paralelo, composto por unidades simples com capacidade natural de armazenar conhecimento e disponibilizá-lo para uso futuro.

Do ponto de vista neurofisiológico, muito pouco se conhece sobre o funcionamento dos neurônios e suas conexões o que compromete a fidelidade destes modelos em fisiologia. As RNAs assemelham-se ao cérebro em dois aspectos:

- Elas extraem conhecimento do ambiente através de um processo de *aprendizagem* ou *treinamento*; e
- Os pesos das conexões entre os neurônios, conhecidos como *pesos sinápticos*, são utilizados para armazenar o conhecimento adquirido.

A figura 6 apresenta um modelo de neurônio biológico com a seqüência de propagação dos sinais pela célula. Os neurônios artificiais também são chamados de *nós*, ou *unidades*.

A natureza das RNAs faz com que seu estudo seja multidisciplinar, envolvendo pesquisadores de diversas áreas, como neurofisiologia, psicologia, física, computação, engenharia, estatística, entre outras.

¹ Haykin (1998).

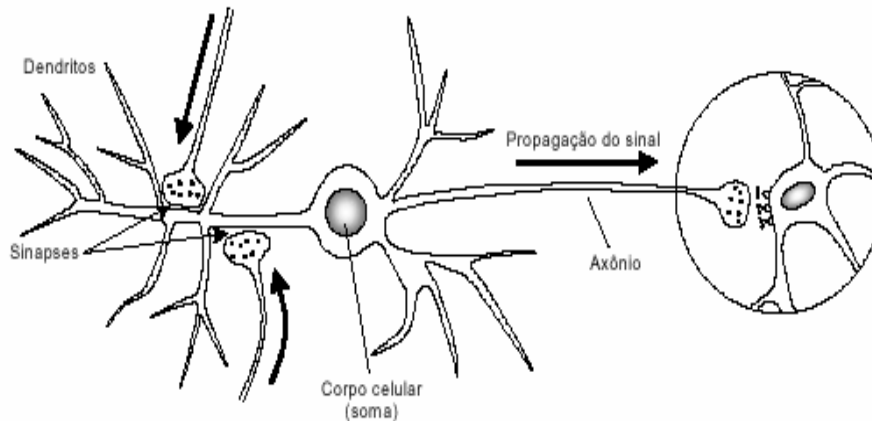


Figura 6. Célula neural biológica (as setas largas indicam a seqüência de propagação de sinais pelos neurônios)

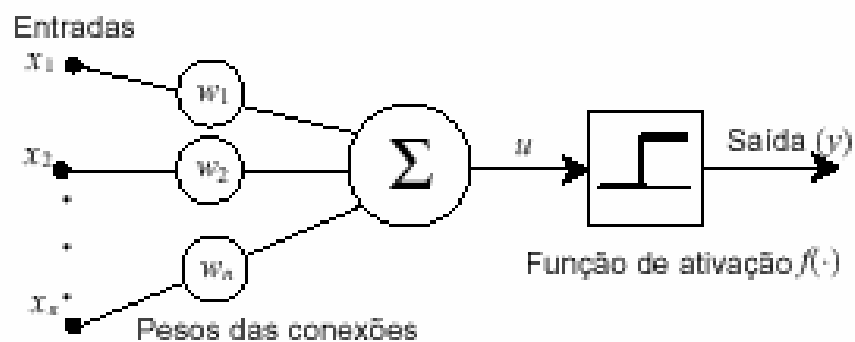


Figura 7. Representação funcional de um neurônio artificial

Neurofisiologistas e psicólogos estão particularmente interessados em compreender o funcionamento do sistema neural humano. As características de resposta a estímulos apresentada por neurônios individuais, bem como redes de neurônios, são alvos de estudo dos neurofisiologistas, enquanto os psicólogos estudam funções do cérebro tratando aspectos cognitivos e estão interessados na utilização de técnicas baseadas em redes neurais para criar modelos detalhados do comportamento humano.

Cientistas da área de computação têm em vista a construção de computadores dotados de processamento paralelo e distribuído, buscando superar as limitações impostas pelos computadores atuais, que realizam processamento serial simbólico.

Inspirados na habilidade apresentada pelos seres humanos e outros animais no desempenho de funções como o processamento de informação sensorial e a

capacidade de interação com ambientes pouco definidos, os engenheiros, por exemplo, estão preocupados em desenvolver sistemas artificiais capazes de desempenhar tarefas semelhantes. Habilidades como capacidade de processamento de informação incompleta ou imprecisa e generalização são propriedades desejadas em tais sistemas.

McCulloch & Pitts (1943) projetaram a estrutura que é conhecida como a unidade básica de uma rede neural. Estes pesquisadores propuseram um modelo de neurônio como uma unidade de processamento binária (Figura 7) e provaram que estas unidades são capazes de executar várias operações lógicas (OU, AND, etc.). Este modelo, apesar de muito simples, trouxe uma grande contribuição para as discussões sobre a construção dos primeiros computadores digitais, permitindo a criação dos primeiros modelos matemáticos de dispositivos artificiais que buscavam analogias biológicas. Matematicamente, o neurônio da Figura 7 pode ser expresso por:

$$y = f(u) = f(x_1w_1 + x_2w_2 + \dots + x_nw_n) = f(w^T x) \quad (5.1)$$

onde y é a saída do neurônio, u é a ativação do neurônio, $f(\cdot)$ sua função de ativação, x_i ($i = 1, \dots, n$) é o i -ésimo componente do vetor x de entradas, e w_i ($i = 1, \dots, n$) é o i -ésimo componente do vetor w de pesos do neurônio.

5.2. Características Principais

É importante ressaltar que a comparação com a neurofisiologia foi apenas uma motivação original da qual pouco sobrou além do nome da ferramenta. Desde o final da década de 80, as redes neurais artificiais são uma metodologia, na fronteira da estatística com a inteligência artificial, eficiente e capaz de resolver uma gama de problemas importantes. No exterior, em especial nos E.U.A, já se encontrou grande aplicabilidade fora dos muros acadêmicos sendo que, aqui no Brasil, começa-se a perceber seu grande potencial. Conceitualmente, uma rede neural artificial é um dispositivo tanto capaz de processar informação de forma distribuída quanto de incorporar conhecimento através de exemplos. Trata-se, portanto, de um processador capaz de extrair conhecimento experimental disponibilizando-o para uso prático (tomada de decisões por exemplo).

As redes neurais artificiais têm sido desenvolvidas como generalizações de modelos matemáticos de cognição humana ou neurobiologia, assumindo que:

- O processamento da informação ocorre com o auxílio de vários elementos chamados *neurônios*;
- Os sinais são propagados de um elemento a outro através de *conexões*;
- Cada conexão possui um *peso* associado, que, em uma rede neural típica, pondera o sinal transmitido; e
- Cada neurônio (ou unidade) possui uma *função de ativação* (geralmente não-linear), que tem como argumento a soma ponderada dos sinais de entrada, para determinar sua saída.

Uma grande vantagem de usar-se uma rede neural é a capacidade de resolver problemas sem a necessidade de definição de listas de regras ou de modelos explícitos. Isto possibilita tratar de situações onde é difícil criar modelos adequados da realidade ou situações com freqüentes mudanças no ambiente.

Atenta-se que grande parte desta sua adequabilidade funcional deve-se à sua capacidade em inferir relações não lineares complexas. Frente a estas suas propriedades, hoje, pode-se observar sua aplicabilidade principalmente nas áreas de classificação de padrões (em um sentido amplo) e de previsão.

Uma rede neural caracteriza-se pela capacidade de extrair conhecimento experimental e por disponibilizar este conhecimento para uso prático. Apesar da plausibilidade biológica ter sido apenas uma motivação original cabe aqui uma comparação. O cérebro desenvolve a função de, a partir da observação de dados (*input*), extrair informação disponibilizando-a para a tomada de decisões.

Sabe-se que o conhecimento é adquirido através de um processo de aprendizado. O mesmo acontece com as redes neurais artificiais. A informação é armazenada em “densidades de conexão” conhecidas como “pesos sinápticos” (ou simplesmente pesos). O processo de aprendizado de uma rede se dá através de um algoritmo que deve ser capaz de ajustar iterativamente os pesos de modo que se atinja o objetivo proposto.

A rede neural aprende, então, o ambiente através de um processo iterativo de modificação dos pesos de interligação, a partir de estímulos fornecidos pelo ambiente. O tipo de aprendizado é determinado pelo modo com que se promove a adaptação dos parâmetros e isso pode ser feito de dois modos:

1) Aprendizado Supervisionado – usa-se um conjunto de pares, entrada e saída, previamente conhecidos que representam a realidade;

2) Aprendizado Não Supervisionado – não se usa um conjunto de exemplos previamente conhecidos. Uma medida da qualidade da representação do ambiente pela rede é estabelecida e os parâmetros são modificados de modo a otimizar esta medida. Este tipo de aprendizado é muito utilizado na área de reconhecimento de padrões.

Salienta-se que, neste trabalho, usou-se o aprendizado não supervisionado, ou seja, escolheu-se as variáveis referentes às concessionárias de energia elétrica como uma medida da representação do ambiente em estudo.

Em síntese, uma rede neural pode ser caracterizada por três aspectos principais: (1) padrão de conexões entre as unidades (*arquitetura* ou *estrutura*), (2) método de determinação dos pesos das conexões (*algoritmo de treinamento* ou *aprendizagem*) e (3) *função de ativação*.

5.2.1. Arquitetura

A forma pela qual os neurônios de uma RNA estão estruturados (interconectados) está intimamente relacionada ao algoritmo de aprendizagem a ser utilizado para treiná-la. A classificação dos algoritmos de aprendizagem será apresentada na Seção 5.2.2. Nesta seção, será feita uma breve descrição das principais arquiteturas de redes neurais artificiais.

Em geral é possível distinguir três classes fundamentais de arquiteturas: *redes feedforward de uma única camada*, *redes feedforward de múltiplas camadas* e *redes recorrentes*.

5.2.1.1. Redes Feedforward de Uma Única Camada

No caso mais simples de redes em camadas (*layers*), tem-se uma camada de entrada com neurônios cujas saídas alimentam a última camada da rede. Geralmente, os neurônios de entrada são propagadores puros, ou seja, eles simplesmente repetem o sinal de entrada em sua saída distribuída. Por outro lado, as unidades de saída costumam ser unidades processadoras, como apresentado na

Figura 7. A propagação de sinais nesta rede é puramente unidirecional (*feedforward*): os sinais são propagados apenas da entrada para a saída, e nunca vice-versa. Esta arquitetura está ilustrada na Figura 8(a) e a direção de propagação dos sinais na Figura 8(b).

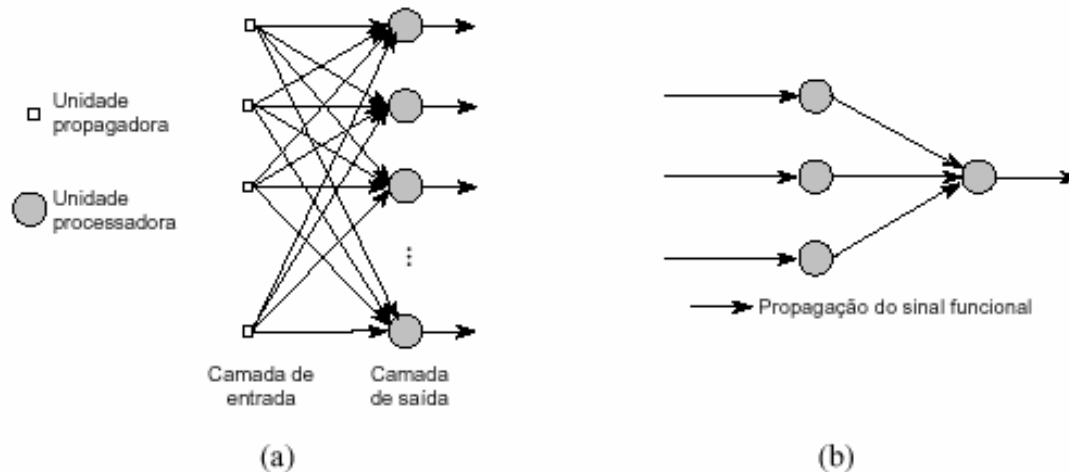


Figura 8. Redes neurais tipo *feedforward* com uma única camada de unidades processadoras. (a) Arquitetura. (b) Sentido de propagação do sinal funcional.

5.2.1.2. Redes Feedforward de Múltiplas Camadas

A segunda classe de rede feedforward se distingue pela presença de uma ou mais camadas intermediárias ou escondidas (camadas em que os neurônios são efetivamente unidades processadoras, mas não correspondem à camada de saída). Adicionando-se uma ou mais camadas intermediárias, aumenta-se o poder computacional de processamento não-linear e armazenagem da rede. O conjunto de saídas dos neurônios de cada camada da rede é utilizada como entrada para a camada seguinte. A Figura 9(a) ilustra uma rede feedforward de múltiplas (duas) camadas intermediárias.

As redes feedforward de múltiplas camadas, são geralmente treinadas usando o algoritmo de retro-propagação do erro (*error backpropagation*), embora existam outros algoritmos de treinamento. Este algoritmo requer a propagação direta (*feedforward*) do sinal de entrada através da rede, e a retro-propagação (propagação reversa, ou *backpropagation*) do sinal de erro, como ilustrado na Figura 9(b).

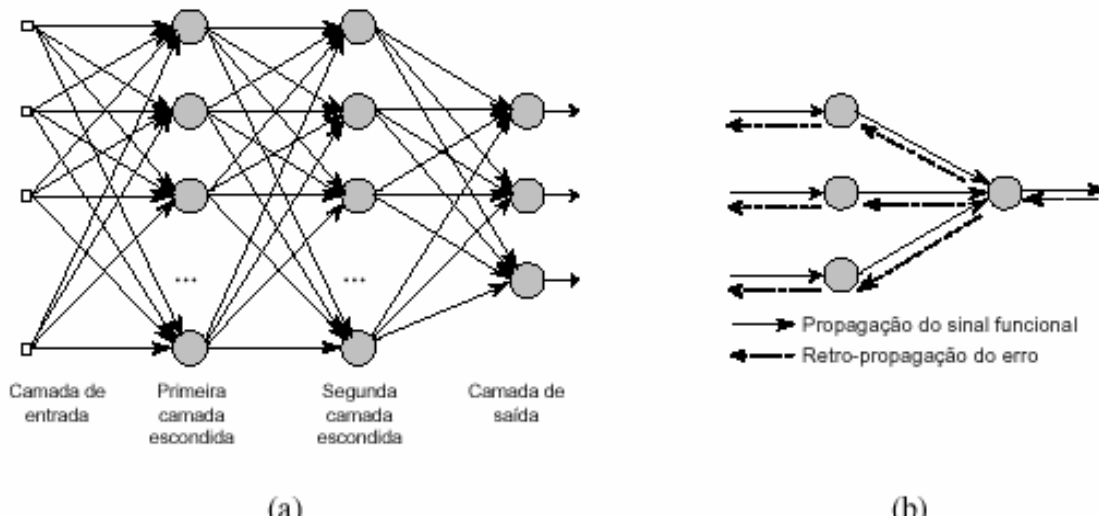


Figura 9. Redes neurais tipo *feedforward* com múltiplas camadas. (a) Arquitetura. (b) Sentido de propagação do sinal funcional e do sinal de erro.

5.2.1.3. Redes Recorrentes

As redes recorrentes distinguem-se das redes feedforward pela existência de pelo menos um laço (*loop*) de recorrência (*feedback*). Por exemplo, uma rede recorrente pode consistir de uma única camada de neurônios com cada neurônio alimentando seu sinal de saída de volta para a entrada de todos os outros neurônios, como ilustrado na Figura 10. O laço de recorrência possui um grande impacto na capacidade de aprendizagem e no desempenho da rede (dos Santos & Von Zuben, 2000). Além disso, este laço envolve a utilização de ramos particulares compostos de unidades de retardo (z^{-1}), resultando em um comportamento dinâmico não-linear, assumindo que a rede possui componentes não-lineares.

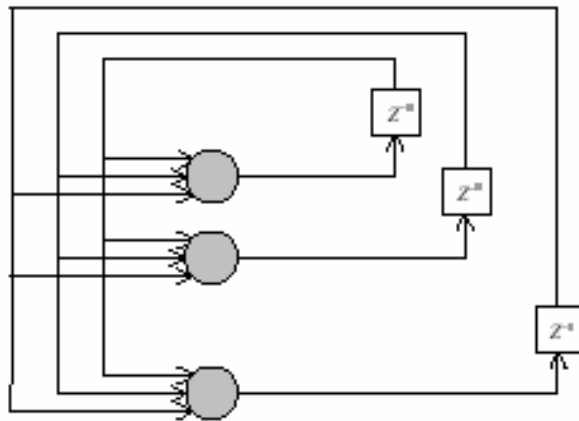


Figura 10. Arquitetura recorrente de rede sem nenhuma camada intermediária.

5.2.2. Métodos de Aprendizagem

A capacidade de *aprendizagem* é uma das características marcantes das RNAs. Uma rede neural aprende, basicamente, através de um processo iterativo de ajuste de pesos e limiares (bias). Atualmente, existem processos mais sofisticados de aprendizagem (ou *treinamento*), que são capazes de ajustar não apenas os pesos da rede, mas também sua arquitetura e as funções de ativação dos neurônios (Von Zuben, 1996, Kwok & Yeung, 1997, de Castro *et al.*, 1999a,b; de Castro & Von Zuben, 1999c).

Segundo Haykin (1998), *Aprendizagem* (ou *treinamento*) é o processo pelo qual os parâmetros livres de uma rede neural são adaptados através de um mecanismo de apresentação de estímulos fornecidos pelo ambiente no qual a rede está inserida. O tipo de treinamento é definido pela forma na qual os parâmetros são modificados.

Esta definição de aprendizagem implica na seguinte seqüência de eventos:

- Apresentação de estímulos à rede neural;
- Alteração dos parâmetros livres da rede; e
- Novo padrão de resposta ao ambiente.

Os principais paradigmas de aprendizagem são: (1) supervisionada, (2) não-supervisionada, e (3) por reforço.

5.2.2.1. Aprendizagem Supervisionada

Trata-se de um paradigma de aprendizagem, no qual um *supervisor* possui conhecimento sobre o ambiente em que a rede está inserida. Este conhecimento está representado sob a forma de um conjunto de amostras de *entrada-saída*. O ambiente, por sua vez, é desconhecido. A Figura 11 ilustra esta abordagem. Os parâmetros da rede são ajustados pela combinação do sinal de entrada com um sinal de erro, que é a diferença entre a saída desejada e a fornecida pela rede.

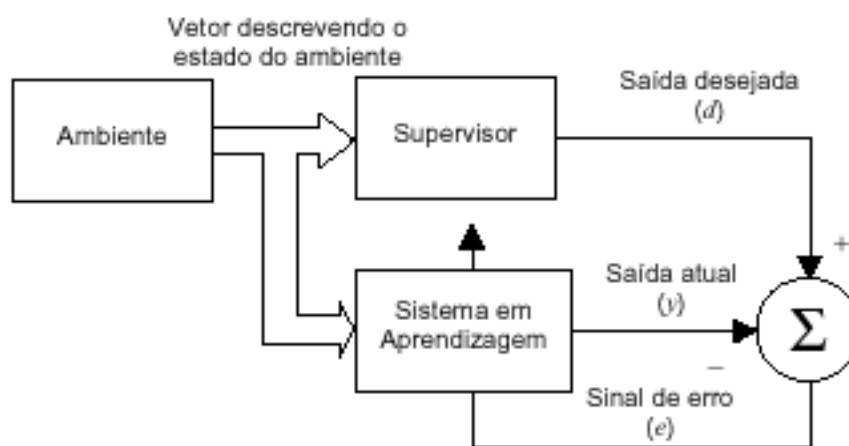


Figura 11. Diagrama de blocos do processo de aprendizagem supervisionada.

Seja t o índice que denota tempo discreto ou, mais precisamente, o intervalo de tempo do processo iterativo responsável pelo ajuste de pesos do neurônio k . O único sinal de saída $y_k(t)$, do neurônio k , é comparado com uma *saída desejada*, denominada $d_k(t)$.

Consequentemente, um sinal de erro $e_k(t)$ é produzido:

$$e_k(t) = d_k(t) - y_k(t) \quad (5.2)$$

5.2.2.2. Aprendizagem Não-Supervisionada

No processo de aprendizagem não-supervisionada ou auto-organizada, não existe um supervisor para avaliar o desempenho da rede em relação ao conjunto de treinamento (Figura 12), ou seja, os dados são não-rotulados. A rede se adapta a regularidades estatísticas dos dados de entrada, desenvolvendo a habilidade de

criar representações internas para codificar características da entrada e, assim, gerar novas classes automaticamente. Geralmente os algoritmos auto-organizados utilizam aprendizagem competitiva.

Na aprendizagem competitiva, os neurônios de saída da rede competem entre si para se tornarem ativos. Um único neurônio de saída é ativado a cada iteração. Esta característica torna o algoritmo apropriado para descobrir características estatísticas salientes, que podem ser utilizadas para classificar um conjunto de padrões de entrada.



Figura 12. Diagrama de blocos dos processos auto-organizados.

Existem três elementos básicos para uma regra de aprendizagem competitiva:

- Um conjunto de neurônios iguais, exceto pelos pesos das conexões;
- Um limite imposto ao peso de cada neurônio; e
- Um mecanismo de competição entre os neurônios. Aquele que vencer a competição é chamado de vencedor (*winner-takes-all*).

Neurônios individuais aprendem a se especializar em grupos (ou clusters) de padrões similares, tornando-se *detectores de características* para diferentes classes de padrões de entrada.

Em sua forma mais simples, uma rede competitiva possui uma única camada de neurônios de saída, totalmente interconectados. Também existem conexões laterais entre os neurônios, como indicado na Figura 13, capazes de efetuar uma inibição lateral entre os neurônios vizinhos.

Para um neurônio k ser o vencedor, seu campo induzido local v_k em relação a um determinado padrão x deve ser o maior de toda a rede. O sinal de saída y_k do neurônio vencedor k é setado em 1, e o sinal de saída de todos os outros neurônios que perderam a competição é setado em 0.

$$y_k(t) = \begin{cases} 1 & \text{se } v_k > v_j \quad \forall j, j \neq k \\ 0 & \text{demais casos} \end{cases} \quad (5.3)$$

onde o campo induzido local v_k representa a ação combinada das entradas positivas elaterais do neurônio.

Se um neurônio não responde a um determinado padrão de entrada, nenhuma aprendizagem ocorre. Por outro lado, se um neurônio ganha a competição, um ajuste $\nabla w_{k,j}$ é aplicado ao vetor de pesos $w_{k,j}$ deste neurônio vencedor.

$$w_{kj} = \begin{cases} \alpha (x_j - w_{kj}) & \text{se } k \text{ vence a competição} \\ 0 & \text{se } k \text{ perde a competição} \end{cases} \quad (5.4)$$

onde α é a taxa de aprendizagem. Esta regra possui o efeito geral de mover o vetor de pesos $w_{k,j}$ do neurônio vencedor k na direção do correspondente padrão de entrada x .

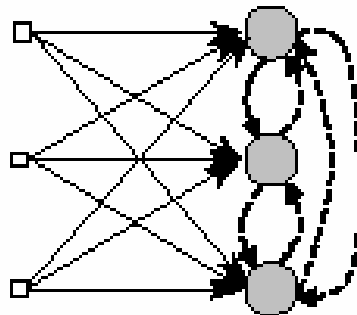


Figura 13. Arquitetura de uma rede competitiva simples com conexões diretas (feedforward) excitatórias da entrada para a saída e conexões laterais inibitórias (setas tracejadas).

5.2.2.3. Aprendizagem Por Reforço

A *aprendizagem por reforço* enfatiza a aprendizagem do indivíduo através da interação direta com o ambiente, sem se basear em uma supervisão ou um modelo completo deste ambiente, visando minimizar um índice escalar de desempenho. A Figura 14 ilustra um tipo de aprendizagem por reforço (Haykin, 1998) baseado em um crítico que converte um *signal primário de reforço* recebido do ambiente em um sinal de reforço de maior qualidade chamado *signal de reforço*

heurístico. O objetivo da aprendizagem é *minimizar uma esperança do erro acumulado*, ao invés de simplesmente tratar o erro atual.

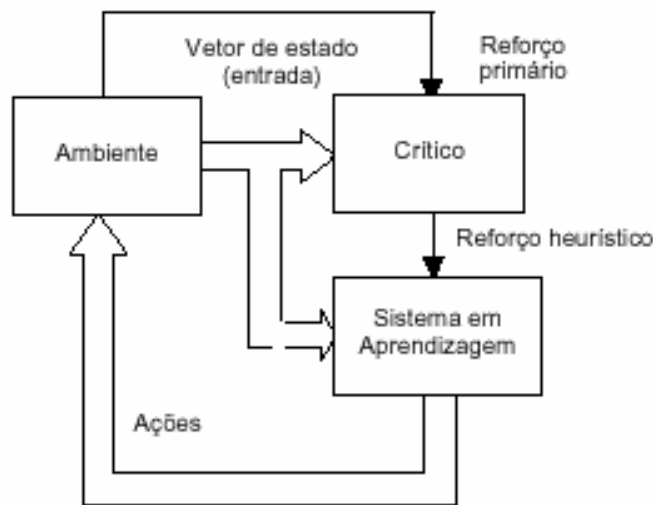


Figura 14. Diagrama de blocos do processo de aprendizagem por reforço.

5.3. Mapas Auto-Organizáveis de Kohonen

Existem diversas arquiteturas e algoritmos de treinamento distintos para as redes neurais artificiais². Esta seção tem por objetivo descrever apenas as redes de Kohonen (auto-organizadas), que foram utilizadas nesta tese.

No presente trabalho de pesquisa, optou-se pela implementação de um tipo de rede neural, os Mapas Auto-Organizáveis de Kohonen³, visto que este método, dentre outros objetivos, pode ser utilizado como uma valiosa ferramenta num problema clássico em estatística que diz respeito à análise de agrupamentos.

As redes que possuem aprendizado competitivo ou auto-organizado (Seção 5.2.2.2) são chamadas *redes competitivas* ou *auto-organizadas*, onde os neurônios de saída competem entre si para estarem ativos ou não. Apenas um neurônio de saída, ou um por grupo, é ativado a cada iteração. As redes auto-organizadas de Kohonen (2000) são caracterizadas pela formação de um mapa topográfico dos

² Como, por exemplo, o perceptron de uma ou mais camadas, as redes neurais de função de base radial (RBF – *radial basis function*) e as redes de Hopfield que podem ser verificadas em Haykin (1998).

³ Kohonen (2000) e Kaski & Kohonen (1996).

padrões de entrada e, por isso, são denominadas de *mapas auto-organizáveis de Kohonen* (SOM – *self-organizing maps*).

No SOM, a localização espacial dos neurônios auto-organizados vai indicar características intrínsecas aos padrões de entrada. Seu principal objetivo é a transformação adaptativa e ordenada de um conjunto de dados de entrada em um mapa (*grid*) uni- ou bidimensional de saída. Os dados de entrada que apresentam semelhanças entre si são agrupados em regiões do mapa de saída, formando classes ou agrupamentos denominados *clusters*.

Durante o processo de auto-organização do mapa, a unidade do *cluster* cujo vetor de pesos mais se aproxima do vetor dos padrões de entrada é escolhida como sendo a *vencedora*. A unidade vencedora e suas unidades vizinhas têm seus pesos atualizados. Além disso, estas redes possuem parâmetros variáveis com o tempo, como a taxa de aprendizagem a , e um raio de vizinhança N_R que indica quais unidades serão atualizadas simultaneamente.

A Figura 15 apresenta arquiteturas típicas de um SOM, considerando configurações de vizinhança unidimensional e bidimensional, embora dimensões mais elevadas possam ser consideradas (Costa, 1999). Além disso, dada a dimensão, a quantidade de unidades ou neurônios de saída pode ser arbitrada e mantida fixa, ou então definida automaticamente pelo algoritmo de treinamento (Fritzke, 1994; Cho, 1997; de Castro & Von Zuben, 1999b).

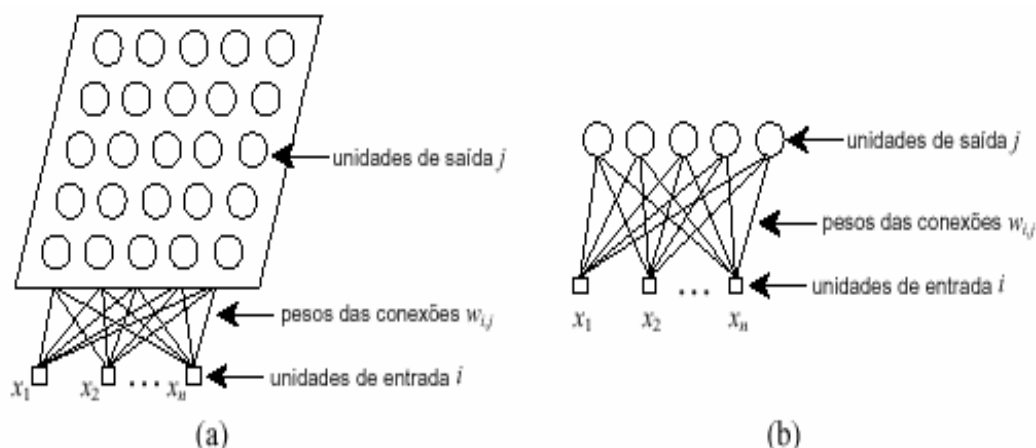


Figura 15. Arquiteturas típicas de um SOM. (a) Bidimensional. (b) Unidimensional.

O SOM foi escolhido pois para uma melhor compreensão dos dados, muitas vezes faz-se mister utilizar técnicas de redução de dimensionalidade, ou seja, de diminuição do número de variáveis e/ou observações, ao mesmo tempo

em que se procura perder o mínimo de informação possível. Desta forma, podemos utilizar, por exemplo, uma Análise de Fatores para identificar um número reduzido de fatores latentes que expliquem a maior parte da variação dos dados. Podemos aplicar também uma Análise de Agrupamentos a fim de determinar e caracterizar grupos de observações semelhantes.

Nesta fase exploratória dos dados, os Mapas de Kohonen, enquanto projeções de um espaço de entrada (uma matriz de dados multivariados por exemplo) em um espaço de menor dimensão (normalmente bidimensional), fornece-nos meios de extrair conhecimento sobre o problema em estudo.

O Mapa Auto-Organizável produz um ajuste seletivo dos neurônios criando um mapa topográfico dos padrões de entrada. Ainda que esta técnica não tenha sido originalmente concebida para classificação de padrões ou segmentação, é possível tirar-se grande proveito das suas propriedades de auto-organização para tais aplicações.

O objetivo básico dos Mapas de Kohonen é agrupar m elementos de um conjunto de padrões de entrada, X , em J neurônios, ou seja, projeta-se o espaço de entrada em um espaço de menor dimensão (normalmente bidimensional). A “comunicação” entre os ambientes de entrada (subespaço natural dos padrões de entrada) e o de saída do algoritmo (malha) é feita por protótipos, ou seja, a cada neurônio no espaço de saída corresponde um vetor-peso que pode ser visto como um protótipo de características do espaço original.

O algoritmo proposto por Kohonen pode ser dividido em quatro etapas básicas: (i) Cálculo das distâncias aos J protótipos de um elemento sorteado X_k do conjunto de padrões de entrada; (ii) Comparação dos valores das J distâncias e reconhecimento do menor, ou seja, aponta-se o protótipo mais próximo de X_k no subespaço de saída, o que determina o neurônio vencedor; (iii) Ativação, por uma rede interativa, simultaneamente, do neurônio vencedor e da sua vizinhança; (iv) Diminuição gradativa, através de processo adaptativo, da distância do neurônio vencedor e da respectiva vizinhança a X_k .

A necessidade de mensurar distância impõe a escolha de duas métricas: uma para o ambiente de entrada (d) e outra para a malha (d^*). Define-se vizinhança (V) de raio R_v do neurônio j como o conjunto de neurônios cuja distância a j é inferior ou igual a R_v na malha:

$$V = V(R_v) = \{N: d^*(C_j, C_i) \leq R_v\} \quad (5.5)$$

O conjunto de neurônios $V(R_v) = r$ é dito r -ésima vizinhança. Embora métricas d^* diferentes possam gerar vizinhanças diferentes, resultados empíricos mostram que o resultado final do Mapa não é afetado de forma significativa pela escolha da métrica da malha.

As tarefas de acionar o neurônio vencedor e sua vizinhança e de fixar as parcelas para a diminuição da distância destes ao elemento apresentado cabem à função de vizinhança centrada no neurônio vencedor, \mathfrak{V}_{C_v} : o protótipo vencedor é atualizado pelo fator unitário e os demais protótipos também podem ser atualizados por fatores diferentes de zero, com valores proporcionais às distâncias, medidas por d^* , dos respectivos neurônios ao neurônio vencedor.

Estabelecida a forma da função de vizinhança, define-se o raio de atualização (R_a) como o parâmetro que determina até qual vizinhança, em relação ao neurônio vencedor, será efetuada a atualização dos protótipos. O raio R_a pode ser variante no tempo. Deste modo, a função de vizinhança centrada no protótipo vencedor v em t , contador do número de interações, é dada pela seguinte relação:

$$\mathfrak{V}_{C_v} = \mathfrak{V}_{C_v}(X(t), R_a(t)) \quad (5.6)$$

Os protótipos são agrupados na matriz U , onde na coluna j está o protótipo associado ao neurônio j . o treinamento então, dá-se em quatro passos:

- 1) Sortear aleatoriamente uma entrada X_k , tal que $X(t) = X_k$;
- 2) Encontrar o neurônio vencedor v tal que:

$$v = \arg_j \min d(X(t), U_j(t)), j = 1, 2, \dots, J \quad (5.7)$$

- 3) Atualizar a matriz de protótipos U através de:

$$U(t+1) = U(t) + \gamma(t) \cdot \mathfrak{V}_{C_v}(X_k, R_a(t)) \cdot [X_k - U_v(t)] \quad (5.8)$$

Onde $\gamma(t)$ é a taxa de aprendizado em t .

- 4) Interromper o processo quando não forem detectadas alterações significativas nos protótipos.

O Mapa Auto Organizável apresenta duas propriedades importantes. A primeira é que ele aproxima a função densidade de probabilidade dos padrões de entrada, $p(x)$. A transformação que ele realiza é capaz de capturar variações nas estatísticas da distribuição dos elementos do conjunto de entrada. Devido ao sorteio aleatório – passo 1 do treinamento – maior número de neurônios é deslocado para cobrir regiões de alta densidade de probabilidade. Em outras palavras, regiões do domínio dos dados de entrada com maiores probabilidades associadas ocupam mais neurônios na malha. Portanto, a resolução para estas

regiões é maior, ou seja, a malha tem maior habilidade em diferenciar elementos de entrada que estão em zonas de maior probabilidade $p(x)$. A segunda propriedade é a manutenção da topologia do sub-espço de saída.

Em síntese, salienta-se que através do aprendizado competitivo (neurônios competem entre si para serem ativados) os neurônios se especializam em responder a estímulos pertencentes a uma mesma classe, implicando no fato de que a topologia no espaço de saída reflete características semelhantes no espaço original, ou seja, entradas com características parecidas serão mapeadas em regiões próximas no espaço de saída.