

PS- 1070

INTRODUCTION TO ARTIFICIAL NEURAL NETWORKS

Alcantaro Lemes Rodrigues (Mestrando no curso de Programa Interunidades Pós Graduação de Energia (PIPGE) da USP, São Paulo. Brasil) - lemes@iee.usp.br
Paulo Lucas Filho Dantas (Mestrando no curso de Programa Interunidades Pós Graduação de Energia (PIPGE) da USP, São Paulo. Brasil) - dantas@iee.usp.br

Since 40 year the Artificial Neural Networks are investigated with of the intention to pretend the biological brain function. The scientists created various computation models that based in the knowledge of doctors, biologists and physicals about brain function. In this paper search find biography about this issue and than it creates one compendium about various exists net types. Although there are many works that treat ANN and its application, little are those that research explaining profound manner what are the ANN. This work finds to full this gap.

Keywords: artificial neural networks

INTRODUÇÃO ÀS REDES NEURAIIS ARTIFICIAIS

Desde os anos 40 as Redes Neurais Artificiais (RNA) são estudadas com o objetivo de imitar o funcionamento do cérebro biológico. Os cientistas criaram vários modelos computacionais com base no entendimento de médicos, biólogos e até psicólogos sobre o funcionamento do cérebro. Neste trabalho buscou se levantar a bibliografia sobre o assunto e assim criar um compêndio sobre os diversos tipos de redes existentes. Embora existam muitos trabalhos que tratam das RNA e sua aplicação, poucos são aqueles que buscam explicar de maneira aprofundada o que são as RNA. Este trabalho busca preencher esta lacuna.

Palavras-chave: redes neurais artificiais

Introdução

As Redes Neurais Artificiais (RNA) são empregadas em diversas áreas do conhecimento e em particular, no campo dos Sistemas de Informação é utilizada em Sistemas de Apoio a Decisão (LAUDON E LAUDON, 2001), porém pouco se conhece como tais sistemas são construídos. Muito se sabe de sua aplicação, mas pouco sobre sua estrutura, o que caracteriza o conceito de caixa preta, onde os mecanismos de funcionamento e construção são desconhecidos e somente uma interface de entradas e saídas são utilizadas e documentadas.

O objetivo principal deste trabalho é levantar bibliografia atual sobre as RNA e criar um compêndio sobre a história das RNA, como são estruturadas, como funcionam e quais caminhos podem ser trilhados para elaborar novos trabalhos.

Neste trabalho propõe se conhecer de forma mais aprofundada o que são as RNA, sua história, sua estrutura e como são operadas. O Capítulo 1 faz um paralelo entre as RNA e os sistemas computacionais convencionais. No Capítulo 2 discorre sobre os conceitos básicos para o entendimento sobre RNA. O Capítulo 3 aborda a história das RNA, como iniciaram, se desenvolveram, foram abandonadas e depois de mais estudos ressurgira. O Capítulo 4 apresenta algumas topologias existentes na atualidade, no capítulo 5 algoritmos e regras de operação de cada topologia são apresentados. No final as conclusões deste trabalho são apresentadas.

1 RNA

As RNA são sistemas computacionais, *hardware* ou *software*, que imitam as habilidades de raciocínio dos sistemas biológicos pelo uso de vários neurônios artificiais, os quais são simples emulações dos neurônios biológicos. Usam paralelismo lógico combinando várias operações (as informações da camada de um neurônio, localizadas em uma região ou camada do sistema, são transferidas para os neurônios de outra camada). As três características que descrevem uma rede neural artificial e contribuem para sua habilidade funcional são: estrutura, dinamismo e conhecimento (MAREN e HARSTON, 1990).

Na natureza a função de uma rede é determinada pelas conexões entre elementos. Pode-se treinar uma rede neural artificial para realizar uma função particular pelo ajuste dos valores (pesos) das conexões entre seus elementos ou unidades (DEMUTE e BEALE, 2002).

Algumas das vantagens que as RNAs têm com relação a sistemas especialistas são as características presentes no cérebro humano e não presentes nos computadores, tais como (JAIN *et al*, 1996):

- Paralelismos massivos;
- Representação e computação distribuídas;
- Habilidade em aprender;
- Habilidade em generalizar;
- Adaptabilidade;
- Processar informações dentro de um contexto;
- Tolerância à falhas; e
- Baixo consumo de energia.

Lau (LAU, 1992) observa que ao invés de executar um programa de instruções seqüencialmente como no computador convencional, o modelo de redes neurais explora muitas hipóteses que competem simultaneamente usando redes paralelas compostas de muitos elementos computacionais conectados por ligações com pesos variáveis. A Tabela 1-1 apresenta uma lista onde são comparados itens de estrutura e desempenho entre os sistemas neurais biológicos e o modelo de computador convencional idealizado por von Neumann (VON NEUMANN, 1945).

Tabela 1-1 - Computador convencional versus sistema neural biológico (JAIN *et al.*, 1996)

	Computador de von Neumann	Sistema neural biológico
Processador	Complexo Alta velocidade Um ou alguns	Simples Baixa velocidade Uma grande quantidade
Memória	Separada para um processador Localizada Nenhum conteúdo endereçável	Integrada no processador Distribuída Conteúdo endereçável
Computação	Centralizada Seqüencial Programas armazenados	Distribuída Paralela Auto-aprendizagem
Confiabilidade	Muito vulnerável	Robusto
Ambiente operacional	Bem definido, bem restrito	Definição pobre, não restrita

2 Definição de RNA

Uma razão para utilizar o sistema nervoso como modelo é que as simulações dos sistemas computacionais funcionam melhor se usam exemplos do cérebro. Muitas pesquisas sobre inteligência artificial só obtiveram melhores resultados quando as equipes de engenharia de *software* consultaram os neurocientistas e neurofisiologistas. Assim conforme novas descobertas sobre o funcionamento do cérebro biológico ocorriam, melhorias eram criadas nos sistemas computacionais.

Os pesquisadores constataram também que com os ensaios dos sistemas computacionais baseados nas idéias que surgiam sobre o funcionamento do cérebro, o funcionamento do próprio cérebro era entendido ao se desenvolver as simulações computacionais.

Experimentos e pesquisas sobre o comportamento em animais e humanos em meados do século XX sugeriram que o cérebro humano incorpora informação de uma forma paralela e distribuída com conexões adaptadas para aprender, reconhecer, generalizar e discriminar. A seguir duas das principais contribuições da época.

Lei de Hebbe: segundo Hebbe (HEBBE, 1949) a atividade sináptica facilita a comunicação entre os neurônios, portanto a memória do conhecimento se localiza na atividade sináptica. No processo de aprendizagem ocorre a mudança de peso das sinapses. Este processo é considerado ser a aprendizagem no campo das RNA.

Pesquisa de Lashley: Lashley (LASHLEY, 1950) realizou experiências com ratos retirando partes de seus cérebros e analisou o comportamento dos mesmos para sair de dentro de um labirinto. Ele constatou que mesmo tirando partes do cérebro dos ratos eles continuavam a achar a saída do labirinto. Assim formulou algumas leis:

Lei da equipotencialidade – várias áreas do cérebro têm potencial de aprendizado similar.

Lei da ação em massa – os erros que os ratos cometiam eram proporcionais à massa de cérebro removida.

2.1 ORGANIZAÇÃO ESTRUTURAL DOS SISTEMAS NEURAIS: CAMADAS E COLUNAS.

No cérebro os neurônios são maciçamente conectados, muito mais que as complexas e densas redes de telefonia. Cada neurônio possui de 10^3 a 10^4 ligações com outros neurônios. No total o cérebro humano possui cerca de 10^{14} a 10^{15} interconexões.

Como princípio o cérebro possui elementos de processamento organizados em camadas e também em colunas. Há tecidos que são especializados em sistemas específicos e não específicos.

As sensações que sentimos são passadas por nossos sentidos (visão, audição, paladar, olfato e tato) por múltiplos níveis de camadas. Nossa retina possui três ou quatro camadas de neurônios para o foto-receptor e seis camadas no córtex visual para processamento das imagens. A informação visual por sua vez é processada por mais camadas até chegar no entendimento ou consciência.

O córtex, a substância cinzenta do cérebro, é dividido em colunas que atuam como unidades funcionais. No córtex as entradas começam na parte inferior e terminam com neurônios na segunda e quarta camadas e há sinapses colaterais na terceira, quinta e sexta camadas. A informação é processada no córtex e a via de saída é seguida da terceira e quinta camadas para outras partes do cérebro e espinha dorsal.

2.2 NEURÔNIO BIOLÓGICO.

Os neurônios são formados pelo corpo celular ou soma e filamentos que são os dendritos e os axônios, e conectados uns aos outros formando uma grande teia ou rede neural. As sinapses formam as conexões entre os neurônios e compõe a saída de um neurônio por meio do axônio (*hillock*) e os dendritos. Elas são formadas por um espaço, entre o terminal de um axônio e o receptor de um dendrito, que equivale à cerca de 50 a 200 ângstrons. A resposta de um neurônio ocorre por meio da transferência de impulsos elétricos ou potenciais de ação. A propagação de um neurônio para outro é realizada por meio de substâncias chamadas neurotransmissores.

A geração do potencial de ação ocorre em uma região do neurônio chamada “*axon hillock*” localizada junto ao corpo do neurônio. A partir deste ponto o potencial de ação se propaga ao longo do axônio, até atingir os neurônios seguintes. A Figura 2-1 mostra a estrutura resumida de um neurônio biológico.

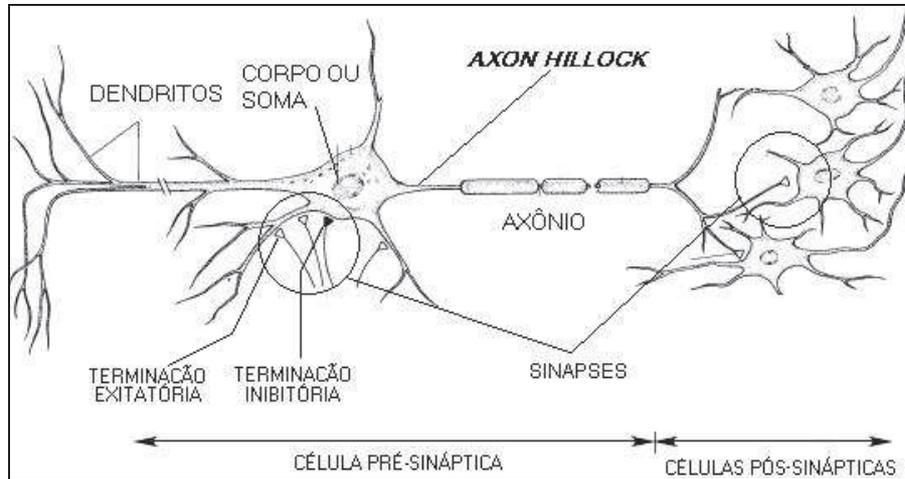


Figura 2-1 - Estrutura resumida neurônio biológico

2.3 NEURÔNIO ARTIFICIAL.

O primeiro modelo de neurônio artificial surgiu a partir dos estudos de McCulloch e Pitts (MCCULLOCH e PITTS, 1943) que formularam um modelo de neurônio computacional conforme mostra a Figura 2-2. Neste modelo o neurônio atua como uma unidade de decisão binária, que é ativada quando a soma ponderada de suas entradas x_N multiplicadas pelos pesos w_i e se o resultado excede um limiar f produz saídas conforme o tipo de função limiar ou função de transferência. Aqui existe uma clara analogia com os neurônios biológicos: o modelo de interconexões dos axônios e dendritos, as conexões por pesos são as sinapses e as funções de transferência representam a atividade do soma ou corpo celular. Entretanto o modelo apresenta diversas simplificações que não refletem o comportamento verdadeiro dos neurônios biológicos.

O modelo de McCulloch e Pitts generaliza de muitas maneiras. Uma generalização óbvia é usar funções de ativação no lugar de funções de limiar, como linear, sigmóide ou gaussiana, mostradas na Figura 2-2.

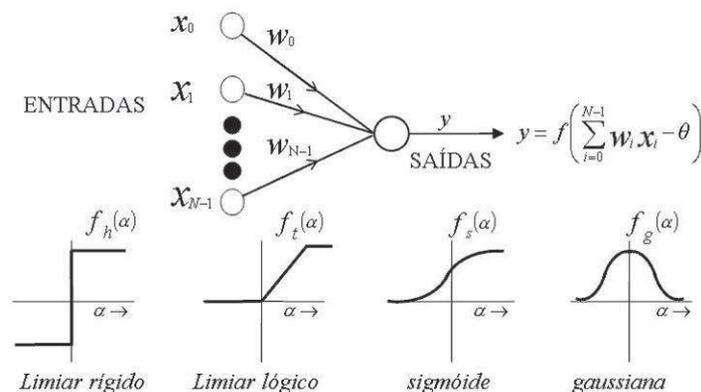


Figura 2-2 - Elemento computacional ou nó e funções de transferência

A função sigmóide é frequentemente usada em RNA. É uma função estritamente crescente que mostra suavidade e tem uma propriedade assintótica. O padrão de função sigmóide é logística, definida por:

$$g(x) = \frac{1}{1 + e^{-\beta x}}, \quad (2-1)$$

onde β é o parâmetro de declive da função.

3 Histórico da Evolução da RNA

A idéia de criação da RNA surgiu antes dos computadores eletrônicos, há quase 60 anos, evoluiu passando por um período de entusiasmo no início dos anos 1960, e depois sobreviveu a um período de frustração e ruptura no final da mesma década. Esta sobrevivência foi em parte devida aos avanços dos computadores eletrônicos, que se tornaram mais baratos e mais rápidos, e às novas soluções propostas por alguns pioneiros que ousaram quebrar o silêncio do período de frustração e defender este importante campo da ciência, isto a partir dos anos de 1980.

3.1 PERÍODO DOS FUNDAMENTOS DAS RNA.

Na época em que surgiram os fundamentos das RNA, estes na verdade sintetizam pesquisas anteriores feitas por diversos profissionais como psicólogos, neurofisiologistas, matemáticos, cientistas da computação e outros colaboradores pertencentes ao chamado grupo dos conexionistas que defendiam a criação de sistemas computacionais baseados nos sistemas biológicos. Tal período compreendeu:

- 1943 – Surge o primeiro modelo de neurônio a partir de estudos sobre a representação de eventos do sistema nervoso (WINDROW, 1990). McCulloch (neurofisiologista) e Pitts (matemático) (MCCULLOCH e PITTS, 1943) formularam o primeiro modelo computacional de neurônio.
- 1949 – Donald Hebb (HEBB, 1949), com o livro *The Organization of Behavior*, relatou pela primeira vez uma regra de aprendizado para um conjunto de neurônios interligados, a partir da modificação da intensidade da conexão das sinapses: a regra de Hebb. Esta intensidade de conexão da sinapse do neurônio biológico é conhecida como peso, uma função matemática que controla entradas e saídas do mesmo neurônio artificial. A regra de treinamento de Hebb foi elaborada em função da hipótese das conexões internas do cérebro estarem continuamente sendo alteradas na medida em que diferentes tarefas vão sendo aprendidas pelos organismos. Esta regra determina que a conexão entre dois neurônios é reforçada, na medida em que o neurônio vai ativando o segundo repetidamente por meio desta conexão.
- 1952, 1954, 1956 – Ashby (ASHBY, 1952), Minsky (MINSKY, 1954) e Uteley (UTLEY, 1956) desenvolveram as idéias de Hebb.
- 1958 – Rosenblatt (ROSENBLATT, 1959) cria o modelo *perceptron* o qual é capaz de classificar padrões.

3.2 PERÍODO DE TECNOLOGIAS PROMISSORAS NO DESENVOLVIMENTO DE RNA.

Neste período importantes melhorias foram feitas em RNA com a contribuição de cientistas das ciências biológicas, ciências exatas e humanas durante o treinamento e validação de novas soluções propostas.

- 1960 – Widrow e Hoff (WINDROW, 1960), na Universidade de Stanford, desenvolveram uma importante variação do algoritmo de aprendizado do *perceptron*. Por meio da aplicação do conceito de erro médio quadrático (*Least Mean Square algorithm - LMS*) também conhecido como regra delta de aprendizado, formularam o *Adaline (Adaptive Linear Element)* ou elemento linear adaptativo.
- 1962 – Com a idéia da *Adaline*, Widrow definiu uma das primeiras redes construída como elementos adaptativos, o *Madaline (Multiple-adaline)*
- Década de 60 – Foram desenvolvidos vários estudos de aplicação do *perceptron*, indicando que as RNA poderiam solucionar qualquer tipo de problema

3.3 PERÍODO DO DESAPONTAMENTO COM RNA.

- 1969 – Minsky e Papert (MINSKY, 1969) lançaram um polêmico livro intitulado *Perceptrons*, onde foi demonstrado matematicamente que o teorema de convergência do *perceptron* desenvolvido por Rosenblatt só era aplicável à classe de problemas linearmente separáveis e não era capaz de resolver problemas elementares como o Ou-exclusivo (XOR). Também mencionaram que não havia motivo aparente para acreditar que redes formadas por múltiplas camadas de *perceptrons* pudessem herdar as boas qualidades do *perceptron* de camada única.

3.4 PERÍODO DAS INOVAÇÕES.

Este período compreendeu basicamente a década de 70 (Séc. XX) onde ocorreram os seguintes fatos significativos:

- 1973 – von der Malrsburg (MARLSBURG, 1973) foi o primeiro pesquisador a demonstrar o princípio da auto-organização que é a capacidade da rede neural artificial de se auto-organizar independente de qualquer mecanismo arbitrário.
- 1976 – Willshaw e von der Malrsburg (WILLSHAW e VON DER MALRSBURG, 1976) publicaram um artigo pioneiro sobre a formação das redes auto-organizáveis, motivado pelo mapeamento topológico do cérebro.
- 1977 – Kohonen (KOHONEN, 1977) desenvolveu seus modelos de memórias associativas que procuram imitar os mecanismos de memória realizados pelo cérebro .
- 1980 – Grossberg (GROSSBERG, 1980) com base em seus trabalhos preliminares sobre aprendizado competitivo, estabeleceu os princípios para uma nova classe de redes neurais denominadas ART (*Adaptive Resonance Theory*).

3.5 PERÍODO DO RESSURGIMENTO DE PESQUISAS COM RNA.

Este período compreendeu basicamente a década de 80 (Séc. XX) onde ocorreu o retorno aos estudos sobre RNA por meio dos seguintes fatos:

- 1982 – Hopfield (HOPFIELD, 1982) utilizou a idéia de minimização da função de energia que é uma função matemática (embora referida como uma energia não representa uma energia real de um sistema físico e sim uma quantidade matemática que atua na dinâmica das redes neurais tal como a energia potencial atua na mecânica de Newton) com enfoque ao princípio de funcionamento das redes recorrentes com conexões sinápticas simétricas, dando origem às *Hopfield Networks*.
- 1982 – Kohonen (KOHONEN, 1982) publicou um artigo sobre mapas auto-organizáveis, utilizando estruturas do tipo *lattice*, diferindo ligeiramente do trabalho de Willshaw e Von der Malrsburg (WILLSHAW e VON DER MALRSBURG, 1976).
- 1983 – Kirkpatrick descreveu um novo procedimento para a otimização de redes (KIRKPATRICK, 1983) baseado nos princípios da termodinâmica estatística, denominado *Simulated Annealing*.
- 1983 – Cohen e Grossberg (COHEN, 1983) estabeleceram um princípio geral para as memórias de conteúdo endereçável. No mesmo ano Barto (BARTO, 1983) publicou um artigo sobre o controle de pêndulo invertido, baseado no aprendizado por reforço (*reinforcement learning*), inicialmente considerado por Minsky em 1954 (MINSKY, 1954).
- 1986 – Um grande marco para o desenvolvimento das RNA foi a criação do algoritmo de retro-propagação (*backpropagation*) por Rumelhart, Hinton e Williams (RUMELHART *et al.*, 1986). Este procedimento havia sido proposto antes por Werbos em 1974 em sua tese de doutorado – (WERBOS, 1974) pelo livro, do mesmo ano, denominado *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* por Rumelhart e McClelland (RUMELHART *et al.*, 1974) onde, pela primeira vez, foi apresentada de forma organizada todo o progresso das redes neurais até então.
- 1988 – Broomhead e Lowe (BROOMHEAD, 1988) descreveram o uso das funções de base radial (*radial basis functions*), muito usadas em problemas práticos.

4 Principais topologias de RNA

As RNA podem ser como grafos dirigidos sobrecarregados onde os neurônios artificiais são as unidades e as arestas dirigidas são conexões (com pesos) entre os neurônios de saída e os de entrada.

As topologias ou arquiteturas, padrões de conexão das RNA, podem ser agrupadas em duas categorias:

- redes alimentadas à diante, em que os grafos não têm laços, e
- redes recorrentes (ou redes realimentadas) onde o laço ocorre devido às conexões de realimentação.

Dentro da família de redes alimentadas à diante, o chamado *Perceptron* de múltiplas camadas, os neurônios são organizados em camadas que possuem conexões unidirecionais entre eles. A Figura 4-1 mostra como são as redes em cada categoria.

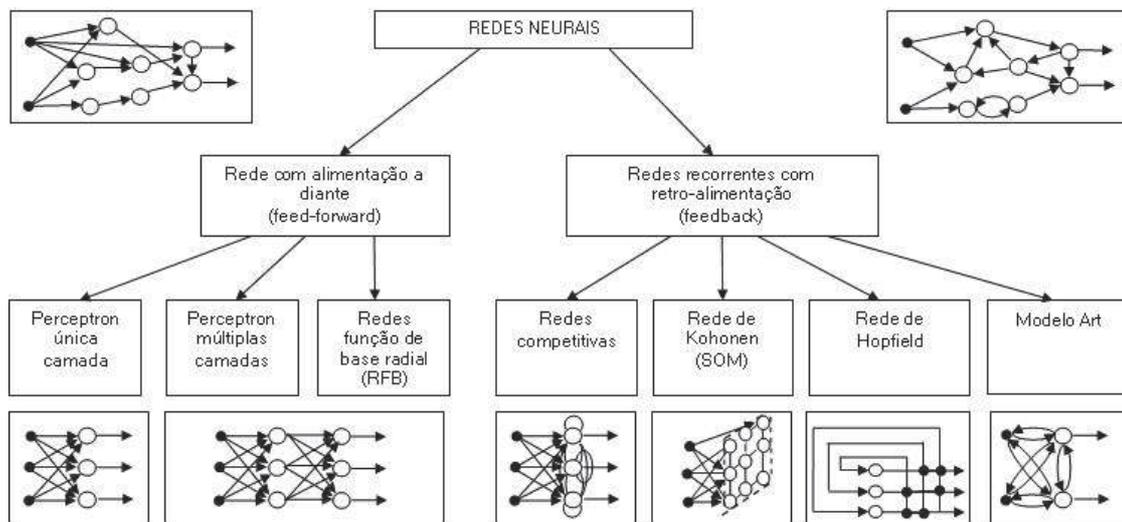


Figura 4-1 - Uma taxonomia das RNA (JAIN *et al.*, 1995)

4.1 REDES ALIMENTADAS À DIANTE.

Diferentes conectividades produzem diferentes comportamentos da rede. De forma genérica as redes alimentadas à diante são “estáticas”, isto é, elas produzem um conjunto de valores de saída no lugar de uma seqüência de valores de uma dada entrada. As redes alimentadas à diante não possuem memória no sentido que sua resposta para uma entrada é independente do estado anterior da rede.

4.2 REDES RECORRENTES.

As redes recorrentes ou redes de realimentação (*feedback*) são sistemas dinâmicos. Quando um novo padrão de entrada é apresentado, as saídas dos neurônios são calculadas. Por causa dos caminhos de realimentação, as entradas são modificadas para cada neurônio, o que leva a rede a entrar em um novo estado.

Diferentes arquiteturas de rede necessitam de algoritmos de aprendizagem adequados. A seguir tem-se um resumo dos processos de aprendizagem.

5 Operação das RNA

5.1 REGRAS DE APRENDIZADO.

O processo de aprendizagem nas RNA ocorre por meio do ajuste dos pesos das conexões e da atualização de sua arquitetura com o objetivo de executar uma função específica de forma eficiente. Este processo de aprendizagem ocorre pelo treinamento da rede por meio de iterações e apresentações de exemplos ou padrões que servem como parâmetros para a atualização dos pesos na rede.

Para entender um processo de aprendizagem, primeiro deve-se conhecer qual é o paradigma de aprendizado (HAYKIN, 1994), ou seja, qual informação da rede está disponível; segundo devem-se entender as regras de aprendizado, ou melhor, como os pesos da rede são atualizados e finalmente como é o algoritmo de aprendizado, isto é, qual é o procedimento em que as regras de aprendizagem são usadas para ajustar os pesos.

Há três paradigmas de aprendizagem: supervisionado, não supervisionado e híbrido. No aprendizado supervisionado os pesos são determinados para a rede produzir respostas tão próximas quanto possíveis às respostas corretas. No aprendizado não supervisionado não há necessidade de uma resposta correta. A estrutura básica da rede é explorada por meio da correlação entre padrões e organizada em categorias de padrões. O aprendizado híbrido por sua vez combina aprendizado supervisionado e não supervisionado onde parte dos pesos é determinada por meio de um aprendizado supervisionado, enquanto outra parte é obtida por meio de um aprendizado não supervisionado.

Existem quatro tipos básicos de regras de aprendizagem: correção de erros, Boltzmann, Hebbian e aprendizado competitivo.

5.1.1 Regra de aprendizado de Correção por erros

No paradigma de aprendizagem supervisionado a rede fornece uma saída desejada para cada padrão de entrada. Durante o processo de aprendizado uma saída y gerada pela rede não pode ser igual a uma saída desejada d . O princípio básico da regra de aprendizagem de correção de erros é usar o sinal de erro $(d - y)$ para modificar os pesos da conexão para gradualmente reduzir este erro.

A regra de aprendizagem do *Perceptron* é baseada no princípio de correção de erros. Um *Perceptron* consiste de um único neurônio com ajustes de pesos, w_j , $j = 1, 2, \dots, n$, e o limiar (*bias*) μ , como mostra a Figura 2-2. Dado um vetor de entrada $x = (x_1, x_2, \dots, x_n)^t$, a entrada da rede para o neurônio é dada pela equação (3-2).

$$y = \sum_{j=1}^n w_j x_j - \mu \quad (5-1)$$

A saída dos *perceptron* é $+1$ se $v > 0$, ou 0 caso contrário. Em um problema para classificar duas classes, o *perceptron* atribui um padrão de entradas para uma classe se $y = 1$, e para outra classes se $y = 0$. A equação linear (3-3) define a fronteira de decisão (um hiperplano em um espaço de dimensão n) que divide o espaço.

$$\sum_{j=1}^n w_j x_j - \mu = 0 \quad (5-2)$$

Rosenblatt (ROSENBLATT, 1962) desenvolveu um procedimento de aprendizagem para determinar os pesos e o limiar em um *perceptron*, dado um conjunto de padrões de treinamento. Nota-se que o aprendizado ocorre somente quando o *perceptron* comete um erro, pois os pesos e limiares se ajustam, Rosenblatt provou que quando padrões para treinamento são extraídos de duas classes linearmente separáveis, o procedimento de aprendizado do *perceptron* converge depois de um número finito de iterações. Este é o **Teorema de convergência do *perceptron***. Na prática, não se sabe quando os padrões são linearmente separáveis. Outras funções de ativação que conduzem para diferentes características de aprendizagem podem ser usadas. Entretanto, uma simples camada do *perceptron* pode separar padrões linearmente separáveis contanto que uma função de ativação monotônica¹ seja usada.

O algoritmo de aprendizagem *backpropagation* é também baseado no princípio da correção de erros.

5.2 ALGORITMOS DE APRENDIZADO.

A Tabela 5-1 resume vários algoritmos de aprendizagem e sua arquitetura associada. Os paradigmas de aprendizagem supervisionados e não supervisionados empregam regras de aprendizagem baseadas na correção de erros, Hebbian e aprendizado competitivo. As regras de aprendizagem baseadas em correção de erros podem ser usadas para treinar redes de alimentação à frente, enquanto regras Hebbian de aprendizagem são usadas para todos os tipos de arquiteturas de redes. Entretanto, cada algoritmo de aprendizagem é projetado para treinar uma arquitetura específica. Então, quando um algoritmo de aprendizagem é discutido, uma arquitetura de rede em particular está implícita. Cada algoritmo pode executar poucas tarefas de modo adequado. A última coluna da Tabela 5-1 lista as tarefas que cada algoritmo pode realizar devido à limitação de espaço.

¹ Uma função monotônica pode ser crescente ou decrescente e pode ter vários pontos de descontinuidade.

Tabela 5-1 - Características de algoritmos de aprendizado mais comuns em RNA (adaptado JAIN *et al.*, 1995).

Algoritmos de aprendizagem conhecidos				
Paradigma	Regra	Arquitetura	Algoritmo	Tarefas
Supervisionado	Correção de erros	Perceptron simples ou múltiplas camadas	Perceptron Backpropagation Adaline Madaline	Classificação de padrões Aproximação de funções Previsão Controle
	Boltzmann	Recorrentes	Boltzmann	Classificação de padrões
	Hebbian	Multicamadas alimentação adiante (<i>feedforward</i>)	Análise discriminação linear	Análise de dados Classificação de padrões
	Competitivo	Competitivo	Vetor de quantificação	Categorização em classes Compressão de dados
Não supervisionado	Correção de erros	Multicamadas alimentação adiante (<i>feedforward</i>)	Projeto de Sammon	Análise de dados
	Hebbian	Alimentação a diante ou competitivo	Análise dos componentes principais	Análise de dados Compressão de dados
		Rede Hopfield	Memória associativa	Memória associativa
	Competitivo	Competitivo	Vetor de quantização	Categorização Compressão de dados
Híbrida	Correção de erros e competitivo	Rede RBF	RBF	Classificação de padrões Aproximação de funções Previsão; Controle

5.2.1 Algoritmo para Redes multicamadas Alimentadas à Diante

A classe de rede multicamadas alimentadas à diante é o *perceptron* multicamadas em que cada unidade emprega também a função de limiar ou função sigmóide. O *perceptron* de multicamadas pode formar limiares de decisão complexa e representar qualquer função booleana (MINSKY, 1969). O desenvolvimento de algoritmos de aprendizagem *backpropagation*, e os pesos em um *perceptron* multicamadas fazem estas redes muito populares entre pesquisadores e usuários de RNA.

É denotado $w_{ij}^{(l)}$ como o peso da conexão entre a i ésima unidade da camada ($l-1$) para j ésima unidade na camada l .

Dado que $\{(x^{(1)}, d^{(1)}), (x^{(2)}, d^{(2)}), \dots, (x^{(\rho)}, d^{(\rho)})\}$ é um conjunto de ρ padrões treinamento (pares de entrada e saída), onde $x^{(l)} \in R^n$ é a entrada do vetor no espaço dimensional de padrões, e $d^{(i)} \in [0,1]^m$, um hipercubo m -dimensional para propósitos de classificação, m é o número de classes, a função custo dos erros quadrados mais freqüente usada em RNA na literatura é definida como:

$$E = \frac{1}{2} \sum_{i=1}^p \|y^{(i)} - d^{(i)}\|^2 \quad (5-3)$$

O algoritmo *backpropagation* (RUMELHART, 1989) é um método de gradiente-descendente para minimizar função custo dos erros quadrados na Equação (3-4) conforme Figura 5-1.

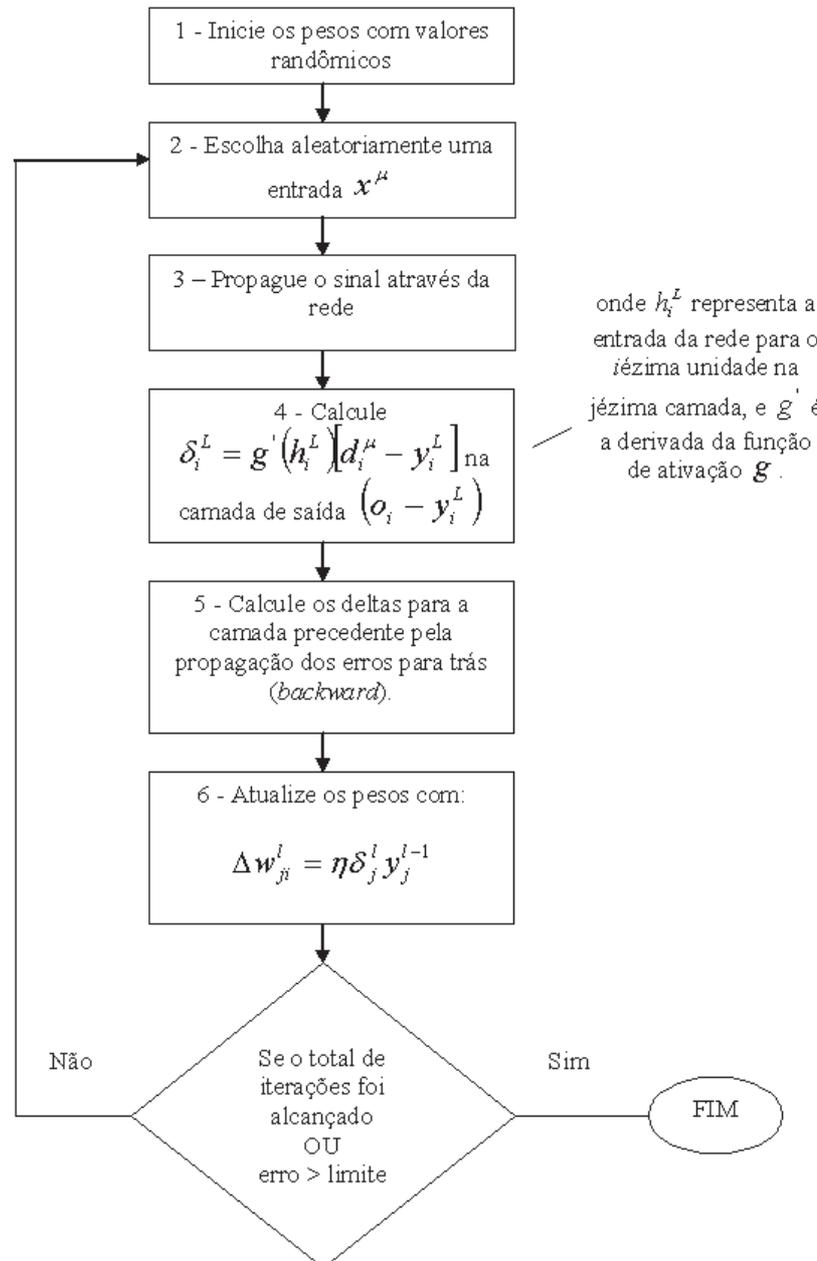


Figura 5-1 - Algoritmo de aprendizado *backpropagation*.

Cada unidade na primeira camada oculta forma um hiperplano no espaço de padrões. As fronteiras entre classes padrão podem ser aproximadas por hiperplanos. Uma unidade na segunda camada oculta forma uma hiper-região das saídas das unidades da 1ª camada; uma região de decisão é obtida pela execução de uma operação AND nos

hiperplanos. As unidades da camada de saída combinam as regiões de decisão feitas pelas unidades da segunda camada oculta pela execução de operações lógicas OR.

Uma rede, com duas camadas, forma várias fronteiras de decisão e, além disso, as MLP com função de ativação sigmóide podem formar fronteiras de decisão suaves no lugar de fronteiras lineares de variáveis.

5.2.2 Problemas com redes alimentadas à diante

Há muitos problemas no projeto deste tipo de rede:

- quantas camadas são necessárias para uma dada tarefa;
- quantas unidades são necessárias por camada;
- como a rede atuará com dados não incluídos no conjunto de treinamento; (habilidade de generalização); e
- dimensão necessária do conjunto de treinamento para obter uma boa generalização.

Embora redes alimentadas à diante usando *backpropagation* sejam muito empregadas para classificação e funções de aproximação (HERTZ, 1991), vários parâmetros de projeto ainda devem ser determinados por tentativas e erros. Resultados teóricos existentes apenas fornecem poucas linhas gerais para escolher estes parâmetros na prática.

Conclusão

As RNA são estruturas computacionais que são aplicadas em diversas áreas do conhecimento como engenharia, tomada de decisões, biologia, processamento de sinais e demais áreas que necessitam de sistemas computacionais robustos e eficientes em executar tarefas específicas.

Neste trabalho foram apresentadas as RNAs, sua história, principais topologias ou arquiteturas e algumas formas de operação como as regras de aprendizado, os algoritmos de aprendizado e possíveis tarefas de aplicação. Assim os objetivos deste trabalho em apresentar as RNAs de maneira mais aprofundada se mostra como atingido.

Um trabalho que pode ser elaborado é o uso de exemplos simplificados para explicar cada uma das várias RNAs citadas neste trabalho. Embora haja trabalhos que utilizam as RNAs e poderiam ser utilizados como exemplos, falta um trabalho que exemplifique de maneira mais didática como cada arquitetura deveria ser empregada e exemplos poderiam auxiliar no entendimento das técnicas empregadas nas RNAs.

Apesar de serem muito empregadas no campo da engenharia e das ciências as RNAs podem ser empregadas em sistemas para apoio a decisão (SAD), porém é necessário que exista conhecimento o qual é utilizado no processo de aprendizagem da rede. Uma vez que tal conhecimento nem sempre é disponível, pois na maioria das vezes pertence a uma pessoa que muitas vezes não pertence ao quadro de funcionários de uma empresa e, portanto fica complicado resgatar tal conhecimento, se faz necessário estudar RNAs que tratam com poucas informações, ou base de conhecimento restrita.

Os algoritmos de RNA recorrentes devem ser experimentados para que se possa aperfeiçoar este tipo de arquitetura. Poucas são as aplicações que se encontram na literatura a respeito deste tipo de RNA, assim mais pesquisa com este tipo de RNA se faz necessária.

Bibliografia

- ANDERSON, J.A.; ROSENFELD, E. Neurocomputing: Foundations of Research. MIT Press, Cambridge, Mass., 1988.
- ASHBY, W.R.; Design for a brain. New York, John Wiley. (1952).
- BARTO, A.G.; SUTTON, R.S.; ANDERSON, C.W.; Neuronlike adaptive elements that can solve difficult learning control problems. IEEE Transactions on Systems, Man and Cybernetics. v.13, p.34-846, 1983.
- BROOMHEAD, D.S.; LOWE, D.; Multi-variable function interpolation and adaptive networks. Complex Systems, v.2, p. 269-303, 1988.
- CARPENTER, G.A; GROSSBERG, S. Pattern Recognition by Self-Organizing Neural Networks. MIT Press, Cambridge, Mass., 1991.
- COHEN, M.A.; GROSSBERG, S.; Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. IEEE Transaction on Systems, Man, and Cybernetics, b13, p.815-26, 1983.
- DEMUTE, H.; BEALE, M.; Neural Network Toolbox For Use with Matlab - User's Guide. Version 4, Release 13, The Math Works Inc., 2002.
- DILLON, T.S.; NIEBUR, D.; Neural Networks Applications in Power Systems. CRL Publishing, London, UK, 1995.
- GROSSBERG, S.; How does a brain build a cognitive code? Psychological Review. n.87, p.1-51, 1980.
- HAGAN, M. T.; MENHAJ, M.; Training feedforward networks with the Marquardt algorithm. IEEE Transactions Neural Networks, vol. 5, nov. 1994.
- HAYKIN, S.S.; Redes Neurais: Princípios e Práticas. Artmed Editoras, 2002.
- HAYKIN, S.S.; Neural Networks: A Comprehensive Foundation. MacMillan College Publishing Co., New York, 1994.
- HEBB, D.O.; The organization of behavior: a neuropsychological theory, John Wiley, New York, 1949.
- HERTZ, J.; KROGH, A.; PALMER, R.G. Introduction to the Theory of Neural Computation. Addison-Wesley, Reading, Mass. 1991
- HOPFIELD, J.J.; Neural networks and Physical Systems with Emergent Collective Computational Abilities. Proceedings of the national Academy of Sciences of the U.S.A., n.79, p.2554-8, 1982.
- JAIN, A.K.; MAO, J.; MOHIUDDIN, K.M.; Artificial Neural Networks: A tutorial. IEEE Computer. (Vol. 29, No. 3) pp. 31-44. March 1996.
- KOHONEN, T.; Associative memory: a system-theoretical approach. New York, Springer, 1977.
- KOHONEN, T. Self Organization and Associative Memory. Third Edition, Springer-Verlag, New York, 1989.
- KOHONEN, T.; Self-organized formation of topologically correct feature maps. Biological Cybernetics. n.43, p.59-69, 1982.
- LAU, C.; Neural Networks. Theoretical Foundations and Analysis. "IEEE Neural Networks Council, sponsor". Includes papers published in two IEEE proceedings special issues and in other publications, 1987-1990. "A Selected reprint volume." 1992.
- LAUDON, Kenneth C. e LAUDON, Jane P. Gerenciamento de sistemas de informação. 3ª ed. Rio de Janeiro. Editora LTC, (2001).
- MAREN, A.; HARSTON, G.; PAP, R.; Handbook of Neural Computing Applications. California. Academic Press, Inc. 1990.

- MCCULLOCH, W.S.; PITTS, W. "A Logical Calculus of Ideas Immanent in Nervous Activity" . Bull. Mathematical Biophysics, Vol. 5, 1943, pp. 115-133.
- MINSKY, M.L.; Theory of neuro-analog reinforcement systems and its application to the brain-model problem. Princeton, Thesis (PhD) – Princeton University, 1954.
- MINSKY, M.L.; PAPERT, S.A.; Perceptrons. Cambridge, MIT Press, 1969.
- OLESKOVICZ, M.; AGGARWAL, R. K.; COURRY, D. V.; O Emprego de Redes Neurais Artificiais na Detecção, Classificação e Localização de Falhas em Linhas de Transmissão, Revista Controle & Automação, Vol. 14, no. 2, Abri/Maio/Junho, 2003.
- ROSENBLATT, F.; The Perceptron: a probabilistic model for information storage and organization in the brain. Psychological Review. N. 65, p. 386-408, 1958.
- ROSENBLATT, R. Principles of Neurodynamics. Spartan Books, New York, 1962.
- RUMELHART, D.; HINTON, G.; WILLIAMS, R.J.; Learning Internal Representation by Error Propagation in Parallel Distributed Processing. The MIT Press. 1986.
- RUMELHART, D.E.; LOWE, D.; Parallel distributed processing. San Diego, MIT Press, 1986.
- STEVENSON JR., W. D.; Elementos de Análise de Sistemas de Potência, Editora McGraw Hill do Brasil, São Paulo, 1978.
- UTLEY, A.M.; A theory of the mechanism of learning based on the computation of condition probabilities. In: International Conference on Cybernetics, 1º, Namur, 1956. Proceedings. Paris, Gauthier-Vellair, 1956.
- VON DER MARLSBURG, C.; Self-organization of orientation sensitive cells in the striate cortex. Kybernetik, n14, p.85-100, 1973.
- VON NEUMANN, J.; First Draft of a Report on the EDVAC. Moore School of Electrical Engineering University of Pennsylvania, 1945, in: IEEE CS Press book. "The Anatomy of a Microprocessor: A Systems Perspective", by Shriver & Smith. http://www.math.nmsu.edu/hist_projects/EDVAC.pdf, acessado em 12/12/2005.
- WERBOS, P.J.; Beyond regression: new tools for prediction and analysis in the behavioral sciences. Cambridge. Thesis (PhD) – Harvard University. 1974.
- WILLSHAW, D.J.; VON DER MARLSBURG, C.; How patterned neural connections can be set up by self-organization. Proceedings of the Royal Society of London. Series B, 1976.
- WINDROW, B.; LEHR, M.A.; 30 years of adaptive neural networks: perceptron, madaline, and back propagation. Proceedings of the IEEE, v.78, n.9, p.1415-41, Sept. 1990.
- WINDROW, B; HOFF JR, M.E.; Adaptive switching circuits. IRE WESCON. Convection Record, v.4, p. 96-104, 1960.
- YANG, C.L.; OKAMOTO, H.; YOKOYAMA, A.; SEKINE, Y.; Expert system for fault section estimation of power system using time sequence information. Elect. Power Energy Syst., vol 14, no. 2, pp. 225-232. 1992.