UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Ilha Solteira

TEORIA DA RESSONÂNCIA ADAPTATIVA ATRAVÉS DA LINGUAGEM JAVA PARA DETECÇÃO E CLASSIFICAÇÃO DE E-MAILS INDESEJADOS

Carlos Roberto dos Santos Junior

Ilha Solteira – SP

2013

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Ilha Solteira

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

"Teoria da Ressonância Adaptativa Através da Linguagem JAVA para Detecção e Classificação de E-mails Indesejados"

CARLOS ROBERTO DOS SANTOS JUNIOR

Orientador (a): Prof. Dra. Anna Diva Plasencia Lotufo
Co-Orientador (a): Prof. Dra. Maria do Carmo Gomes da Silveira

Dissertação apresentada á Faculdade de Engenharia - UNESP - Campus de Ilha Solteira, como parte dos requisitos par obtenção do título de Mestre em Engenharia Elétrica.
Área de Conhecimento: Automação.

Ilha Solteira – SP
2013

FICHA CATALOGRÁFICA
Desenvolvido pelo Serviço Técnico de Biblioteca e Documentação

Santos Júnior, Carlos Roberto dos.
S237t Teoria da ressonância adaptativa através da linguagem Java para detecção e classificação de e-mails indesejados / Carlos Roberto dos Santos Júnior. -- Ilha Solteira: [s.n.], 2013
77 f. : il.

Dissertação (mestrado) - Universidade Estadual Paulista. Faculdade de Engenharia de Ilha Solteira. Área de conhecimento: Automação, 2013

Orientador: Anna Diva Plasencia Lotufo
Co-orientador: Maria do Carmo Gomes da Silveira
Inclui bibliografia

1. E-mails. 2. Spams. 3. Filtro de spams. 4. Redes neurais artificiais. 5. ARTMAP fuzzy.

unesp

UNIVERSIDADE ESTADUAL PAULISTA

CAMPUS DE ILHA SOLTEIRA

FACULDADE DE ENGENHARIA DE ILHA SOLTEIRA

CERTIFICADO DE APROVAÇÃO

TÍTULO: Teoria da Ressonância Adaptativa através da Linguagem JAVA para Detecção e Classificação de E-mails Indesejados

AUTOR: CARLOS ROBERTO DOS SANTOS JUNIOR

ORIENTADORA: Profa. Dra. ANNA DIVA PLASENCIA LOTUFO
CO-ORIENTADORA: Profa. Dra. MARIA DO CARMO G DA SILVEIRA

Aprovado como parte das exigências para obtenção do Título de Mestre em Engenharia Elétrica , Área: AUTOMAÇÃO, pela Comissão Examinadora:

Profa. Dra. ANNA DIVA PLASENCIA LOTUFO
Departamento de Engenharia Elétrica / Faculdade de Engenharia de Ilha Solteira

Profa. Dra. MARA LÚCIA MARTINS LOPES
Departamento de Matemática / Faculdade de Engenharia de Ilha Solteira

Prof. Dr. BENEDITO ISAIAS DE LIMA LOPES
Instituto de Engenharia de Sistemas e Tecnologias Da Informação / Universidade Federal de Itajubá

Data da realização: 28 de fevereiro de 2013.

Dedicatória

A Deus

e a

minha família

Agradecimentos

Primeiramente agradeço a Deus por me fortalecer nos momentos de dificuldade, estar sempre ao meu lado guiando meus passos, e me presentear com pessoas tão especiais que me ensinam a cada dia a tornar-me uma pessoa melhor.

Ao meu pai por me mostrar os reais valores de um homem, nem sempre com palavras, mas principalmente com atitudes, minha mãe por exercer seu papel de mãe de maneira impecável incentivando-me e cobrando-me nas horas certas; minha irmã, que tenho tanto orgulho e carinho, por ser minha companheira de todas as horas, com quem posso sempre contar. Sem vocês tudo seria mais difícil.

A minha namorada Thays, que sem dúvida foi uma das pessoas responsáveis pela realização deste sonho, que nunca hesitou em me ajudar, agradeço pelo amor, respeito e carinho.

Aos meus colegas Marlon Borges, Lucas Teles, Rogério dos Reis e a república Beleza pura, pela força e pelas vezes que de forma atenciosa me receberam em suas casas.

Aos meus colegas de laboratório, pelos conselhos, companheirismo e cafezinhos.

Em especial agradeço a minha orientadora Anna Diva, pelo companheirismo e confiança depositada em mim ao longo desses dois anos. A minha co-orientadora Carminha e a professora Mara pelas dicas e incentivo, e a FEIS por proporcionar toda estrutura para a concretização deste projeto.

A maravilhosa disposição e harmonia do universo só pode ter tido origem segundo o plano de um Ser que tudo sabe e tudo pode. Isto fica sendo a minha última e mais elevada descoberta.

Isaac Newton

Resumo

O problema de mensagens não solicitadas pelos usuários em meios de comunicação eletrônica, apesar de ter surgido antes mesmo da popularização da Internet, ainda é um assunto preocupante. Desperdício de largura de banda, perda de tempo, de produtividade e de dados, ou atraso na leitura de e-mails legítimos, são alguns dos problemas que as mensagens não solicitadas, ou Spams, podem causar. Diversas técnicas de filtragem automática de e-mails são apresentadas na literatura, porém muitas destas não oferecem a possibilidade de adaptação, já que o problema em sistemas reais tem como um de seus principais aspectos ser dinâmico, ou seja, mudar constantemente de características com intuito de evadir as técnicas de filtragem. Neste trabalho é desenvolvido um filtro anti-spam utilizando uma técnica de pré-processamento disponível na literatura, no qual os e-mails são submetidos à extração e seleção de características; e uma Rede Neural Artificial baseada na Teoria da Ressonância Adaptativa, para detecção e classificação de Spams. Tais redes neurais possuem grande capacidade de generalização e adaptabilidade, características importantes para um bom desempenho de filtros anti-spam. O modelo proposto neste trabalho é testado a fim de se validar a eficiência do filtro.

Palavras-chave: E-mails. Spams. Filtro de spams. Redes neurais artificiais. ARTMAP fuzzy.

Abstract

The problem in receiving non desired messages in electronic communication systems is a very hard task; even it has begun before the popularization of Internet. The problems that these kinds of messages can cause are among others: waste of time, waste of band width, productivity and data or delay in reading the real e-mails. Several e-mail automatic filtering techniques are presented in the literature, however many of them without capacity of adaptation, while the problem in real systems must be dynamical, i.e. avoid filtering techniques. This work develops a SPAM filtering using a pre processing technique available in the literature, where the e-mails are submitted to extract and select the characteristics; and a neural network based on the resonance adaptive theory to detect and classify the SPAMS. These neural networks have capacity in generalization and adaptation, important characteristics of good performance of SPAM filters. The proposed model is submitted to several tests to validate the efficiency of the filter.

Keywords: E-mails. Spam. Spam filter. Artificial neural networks. Fuzzy ARTMAP.

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas e Siglas

RNA	Rede Neural Artificial
RNAs	Redes Neurais Artificiais
SVM	Support Vector Machines
MDL	Minimum Description Lenght
CF	Confidence Factors
TOE	Train On Error
WEKA	Waikato Environment for Knowledge Analysis
WTA	Winner-Take-All
SOM	Self-Organizing Map
LVQ	Learning Vector Quantization
FD	Frequency Distribution
CTSS	Compatible Time-Sharing System
MIT	Massachusetts Institute of Technology
ARPANET	Advanced Research Projects Agency
SMTP	Simple Mail Transfer Protocol
DNS	Domain Name System
POP3	Post Office Protocol
MUDs	Multi-User Dungeons
ART	Adaptive Resonance Theory
STM	Short-Term Memory
LTM	Long-Term Memory
FEC	Category Choice Function
HTML	Hyper Text Markup Language

MIME	Multipurpose Internet Mail Extensions

Lista de Símbolos

F_1	Camada de comparação
F_2	Camada de reconhecimento
F_0	Camada que representa o vetor de entradas
F_1^a	Camada de comparação do módulo ART_a
F_2^a	Camada de reconhecimento do módulo ART_a
F_0^a	Camada que representa o vetor de entradas de ART_a
F_1^b	Camada de comparação do módulo ART_b
F_2^b	Camada de reconhecimento do módulo ART_b
F_0^b	Camada que representa o vetor de entradas de ART_b
F^{ab}	Módulo de memória associativa Inter-ART
α	Parâmetro de escolha
β	Parâmetro taxa de treinamento
ρ	Parâmetro de vigilância
T_j	Função de escolha
I	Atividade em F_0
M	Número de neurônios na camada F_1
N	Número de neurônios ativos na camada F_2
M_a	Número de neurônios na camada F_1^a
M_b	Número de neurônios na camada F_1^b
N_a	Número de neurônios ativos na camada F_2^a
N_b	Número de neurônios ativos na camada F_2^b
J	Categoria ativa do módulo ART_a

K	Categoria ativa do módulo ART_b
a	Vetor de entrada do módulo ART_a
a^{c}	Complemento do vetor de entrada do módulo ART_a
b	Vetor de entrada do módulo ART_b
b^{c}	Complemento do vetor de entrada do módulo ART_b
A	Entrada para $\mathrm{F}_1^{\mathrm{a}}$ em forma de codificação de complemento
B	Entrada para $\mathrm{F}_2^{\mathrm{b}}$ em forma de codificação de complemento
ρ_a	Parâmetro de vigilância do módulo ART_a
ρ_b	Parâmetro de vigilância do módulo ART_b
ρ_{ab}	Parâmetro de vigilância do módulo Inter-ART
$\bar{\rho}_a$	Parâmetro de vigilância base
x	Vetor de atividade em F_1
y	Vetor de atividade em F_2
x^a	Vetor de atividade em $\mathrm{F}_1^{\mathrm{a}}$
y^a	Vetor de atividade em $\mathrm{F}_2^{\mathrm{a}}$
w_j^a	Vetor peso do módulo ART_a
x^b	Vetor de atividade em $\mathrm{F}_1^{\mathrm{b}}$
y^b	Vetor de atividade em $\mathrm{F}_2^{\mathrm{b}}$
w_k^b	Vetor peso do módulo ART_b
x^{ab}	Vetor de atividade do módulo Inter-ART
w_j^{ab}	Vetor peso do módulo Inter-ART
t	Termo de um conjunto
C	Conjunto
n	Número de ocorrências do termo t no conjunto C

T	Número total de termos no conjunto C

Sumário

1 Introdução

A necessidade de uma forma de comunicação rápida e econômica tornou-se imprescindível no mundo globalizado, necessidade essa suprida com o uso dos e-mails que permitem atingir vários destinatários com facilidade e sem aumento de custos. Essa facilidade também é observada pelo mercado publicitário e por Spammers que se utilizam desse recurso para atingir seus objetivos através do envio de e-mails não solicitados pelos destinatários, os chamados Spams, a principal desvantagem do recurso. O número de Spams se tornou absurdamente maior se comparado aos e-mails legítimos, como mostram estatísticas divulgadas por grandes corporações de segurança que indicaram no ano de 2012 o total de 72,1% de Spam em relação aos e-mails trafegados no mundo, sendo que 3,4% continham anexos maliciosos (GUDKOVA, 2013). Estima-se que, atualmente, 76,45% dos e-mails sejam Spams (BARRACUDA, 2013). Esses dados representam grande insatisfação dos usuários, já que causa inundação nas caixas de e-mails consumindo tempo, dinheiro, largura de banda, além de fraudes, roubos, etc.

Os filtros de Spams são utilizados para identificar e bloquear o maior número possível de e-mails não solicitados que chegam aos usuários. Na literatura encontram-se várias técnicas para filtragem de Spams, entre elas destacam-se o uso das Redes Neurais Artificiais (RNA), Sistema Imunológico Artificial, SVM (Support Vector Machines), filtros Bayesianos, Lógica Fuzzy, entre outros (UPASANA; CHAKRAVARTY, 2010).

No trabalho de Almeida (2010) foi proposto um novo método de classificação automática de Spams baseado no principio da descrição mais simples auxiliado por fatores de confiança (MDL-CF) com o objetivo de minimizar significativamente as principais deficiências encontradas em filtros de Spams: Alta dependência de aprendizagem inicial, alta sensibilidade a ruídos, queda de desempenho por dimensionalidade e alto custo de treinamento e/ou classificação. O treinamento utilizado é o treinamento por erro (Train On Error - TOE) que tem como principal característica iniciar com o modelo vazio. Em Ma, Q., et. al. (2010) a base de regras disponibilizada em (SPAMASSASIN, 2011) é utilizada para criar um tabela de palavras válidas e pré-processar as amostras de e-mails. Com o pré-processo é gerado o vetor de amostras que alimenta a rede neural do tipo Perceptron Multicamadas com o algoritmo de treinamento Backpropagation. O corpus de e-mails utilizado no trabalho é o SpamAssassin corpus (SPAMASSASIN, 2011). No trabalho de Manjusha, K., Kumar, R. (2010) é utilizado no pré-processamento o software WEKA StringToWord-Vector (WEKA, 2012) para a extra-

ção de características. Para classificação foram combinadas redes Bayesianas e Redes Neurais Artificiais com treinamento baseado em algoritmo genético, a primeira responsável pelos dados do cabeçalho e a segunda pelo corpo do e-mail. O corpus utilizado é uma coleção de 2.000 e-mails coletadas na caixa de e-mails do próprio autor do artigo. Já em Ruan e Tan (2009), na fase de pré-processamento é utilizado Concentração Imunológica Artificial com o objetivo de gerar uma biblioteca de genes e em seguida o vetor de características. A rede neural Perceptron Multicamadas de três camadas e treinamento por Backpropagation é a responsável pela classificação das amostras contidas nos corpus PU1 e Ling corpus utilizados no trabalho. Nos trabalhos de Carpinteiro, Lima e Assis (2006) e Silva (2009) é utilizado o mesmo modelo de extração de características desenvolvido por Carpinteiro, Lima e Assis, (2006) que propôs o uso intensivo do pré-processamento dos dados e métodos de seleção de características. As redes SOM-WTA e SOM-LVQ foram utilizadas em experimentos por Silva (2009) e a rede Perceptron Multicamadas em ambos os trabalhos. Além do corpus SpamAssassin utilizado para testes nos trabalhos, Silva (2009) também utilizou seu próprio corpus para adquirir amostras com características de e-mails originados no Brasil.

Uma Rede Neural Artificial é um processador maciçamente paralelamente distribuído de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para uso. Ela se assemelha ao cérebro por dois aspectos principais: O conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem e, as conexões sinápticas entre neurônios são utilizadas para armazenar o conhecimento adquirido (HAYKIN, 1999).

Neste trabalho é proposto um modelo de filtro de spams utilizando a extração e seleção de características proposto em Carpinteiro, Lima, e Assis, (2006) e a RNA ARTMAP Fuzzy para a detecção e classificação de Spams (CARPENTER, 1992).

A extração e seleção de características é realizada na fase de pré-processamento dos e-mails, e tem como objetivo simplificar a tarefa da RNA na identificação de e-mails legítimos (Hams) e Spams, já que prepara os dados antes de enviá-los à entrada da RNA, tornando-os mais simples, uniformes e sem informações desnecessárias. Na extração de características é identificado o tipo de e-mail, e a partir daí é separada as palavras e descartado informações desnecessárias. Já na seleção de características são avaliadas todas as palavras identificadas na extração e selecionadas as mais relevantes na caracterização dos e-mails. A seleção de características é feita de forma automática com auxílio do método estatístico Frequency Distribution (FD).

Para avaliar a eficiência do modelo proposto, utilizam-se as medidas taxa de erro, taxa de falso positivo e taxa de falso negativo.

1.1 **Organização**

O trabalho está dividido em sete capítulos, sendo organizados da seguinte forma: O e-mail, o spam, redes neurais artificiais ART, metodologia, resultados e finalizando com as conclusões.

O capítulo 2 apresenta a estrutura e o funcionamento dos e-mails, além de informar dados históricos e números atuais do uso de e-mails.

No capítulo 3 o Spam é descrito de uma forma ampla, apresentando as motivações para o seu aumento excessivo em relação aos e-mails legítimos e os principais problemas que causam aos usuários.

No capítulo 4 são abordadas as principais redes neurais artificiais com arquitetura baseada na teoria da ressonância adaptativa, arquitetura escolhida para o uso neste trabalho.

O capítulo 5 apresenta a metodologia proposta composta por um método de pré-processamento e uma rede neural artificial escolhida neste trabalho.

No capítulo 6 os resultados são apresentados para análise da eficiência do modelo.

No capitulo 7 são desenvolvidas algumas conclusões gerais e propostas futuras a respeito do trabalho.

Este texto, também, contém um apêndice, sendo que este apresenta os conceitos básicos das redes neurais artificiais.

2 O E-mail

O e-mail é uma ferramenta que permite compor, enviar e receber mensagens através de sistemas eletrônicos de comunicação dentro de uma rede de dados.

O primeiro sistema de troca de mensagens entre computadores de que se tem notícia foi criado em 1965 (Antes da Internet, que teve início na década de 90), e possibilitava a comunicação entre os múltiplos usuários de um computador do tipo mainframe. Acredita-se que os primeiros sistemas criados com tal funcionalidade foram o Q32 da System Development Corporation e o Compatible Time-Sharing System (CTSS) do MIT (A BRIEF HISTORY, 2012).

O sistema eletrônico de mensagens transformou-se rapidamente em um "e-mail em rede", permitindo que usuários situados em diferentes computadores trocassem mensagens. A rede de computadores ARPANET trouxe uma grande contribuição para a evolução do e-mail, aumentando significativamente a sua popularidade. Relatos indicam que houve transferência de mensagens eletrônicas entre diferentes sistemas situados nesta rede logo após a sua criação, em 1969 (VAN, 2012). Em 1971, o programador Ray Tomlinson introduziu o uso do sinal @ para separar os nomes do usuário e da máquina (domínio) no endereço de correio eletrônico (BELLIS, 2012).

O envio e recebimento de uma mensagem de e-mail são realizados através de um sistema de correio eletrônico. Esse sistema é composto de programas que suportam a funcionalidade de cliente de e-mail e de um ou mais servidores de e-mail que, através de um endereço, conseguem transferir uma mensagem de um usuário para outro. Estes sistemas utilizam protocolos de Internet que permitem o tráfego de mensagens de um remetente para um ou mais destinatários que possuem computadores conectados à Internet (ALMEIDA, 2010).

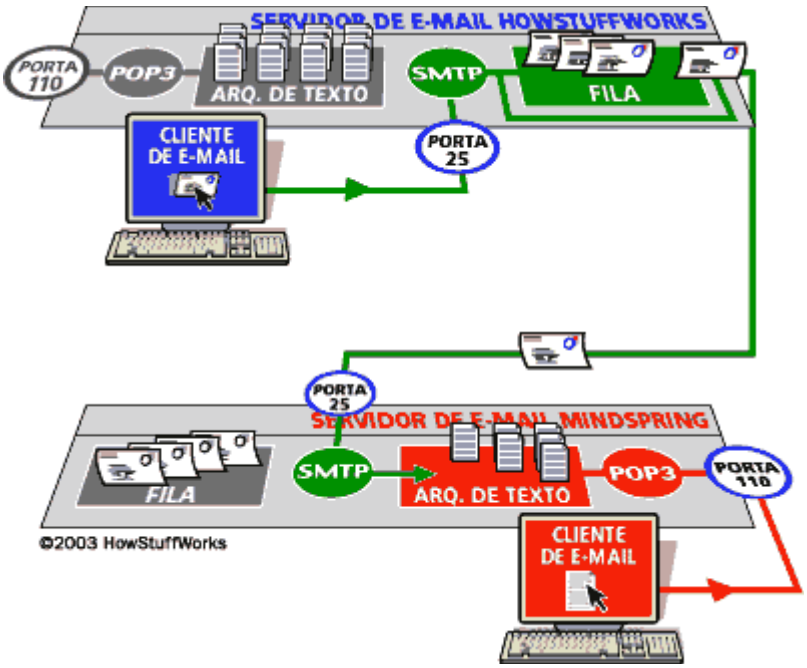
Mensagens de e-mail são compostas basicamente por duas partes principais (HUNT ; LOUKIDES, 1997):

- ✓ Cabeçalho (header) – é estruturado em campos que contém o remetente, destinatário e outras informações sobre a mensagem;
- ✓ Corpo (body) – contém o texto da mensagem separado do cabeçalho por uma linha em branco.

A Figura 1 apresenta a sequência típica de eventos que ocorre ao enviar um e-mail. Considere que um endereço de e-mail com domínio howstuffworks.com envie uma mensagem para um endereço de e-mail com domínio mindspring.com.

As seguintes etapas são realizadas:

Figura 1 - Sequência típica no envio de e-mails



Fonte: (BRAIN, 2012)

- ✓ O editor de e-mails formata a mensagem deixando-a no formato padrão de e-mails da Internet. Utilizando o protocolo SMTP (Simple Mail Transfer Protocol), ele envia a mensagem para o agente de transferência local de e-mail, neste caso smtp.howstuffworks.com;
- ✓ O agente de transferência analisa o endereço de destino fornecido pelo SMTP, e procura pelo domínio mindspring.com no Sistema de Nome de Domínio (DNS), para encontrar o servidor de troca de mensagens que aceite aquele domínio;
- ✓ O servidor DNS do domínio howstuffworks.com, responde com o endereço do domínio mindspring.com;
- ✓ O servidor smtp.howstuffworks.com envia a mensagem para o POP3 (Post Office Protocol) mindspring.com usando SMTP, que entrega a mensagem à caixa de mensagens do usuário;
- ✓ Finalmente, quando o usuário clicar em “Ler mensagem” no seu editor de e-mails, este carregará a mensagem utilizando o protocolo POP3.

As aplicações do e-mail normalmente oferecem ao usuário uma série de facilidades. A maior parte delas fornece um editor de textos embutido e a possibilidade do envio de arquivos anexados à correspondência. Além disso, a maioria das aplicações permite o envio de corres-

pondências para um único destinatário, para mais de uma pessoa ou para um grupo de pessoas (ALMEIDA, 2010).

Embora não tenha sido desenvolvida como uma ferramenta de trabalho cooperativo, os serviços de e-mail adaptaram-se muito bem ao ambiente de grupos de trabalho, tendo se tornado indispensáveis nas organizações, agilizando processos, democratizando o acesso às informações e diminuindo os custos (ALMEIDA, 2010).

De acordo com um estudo realizado pelo grupo de pesquisa Radicati Group, o volume de e-mails enviados diariamente em 2012 passou de 144,8 bilhões, e em 2016 devem ultrapassar 192 bilhões. O número de contas de e-mail em todo o mundo é esperado aumentar de 3300 milhões de contas em 2012 para mais de 4,3 bilhões até o final de 2016 (TEAM, 2013).

3 O Spam

O desenvolvimento e a popularização da Internet e do e-mail ocorreram de forma simultânea ao crescimento de um fenômeno que, desde o seu surgimento, tornou-se um dos principais problemas da comunicação eletrônica em geral: o envio em massa de mensagens não solicitadas. Esse fenômeno ficou conhecido como Spamming, as mensagens em si como Spam e os seus autores como Spammers (ALMEIDA, 2010).

Apesar da existência de mensagens não eletrônicas que podem ser comparadas ao spam, como por exemplo, correspondências, ligações telefônicas e folhetos promocionais não solicitados, o termo é reservado aos meios eletrônicos devido às motivações que os tornam muito mais propícios ao crescimento do fenômeno (ALMEIDA, 2010).

O termo Spam, abreviação em inglês de “spiced ham” (presunto condimentado), diz respeito a uma mensagem eletrônica não solicitada enviada em massa. Consequentemente, as mensagens legítimas passaram a ser chamadas de Hams (ALMEIDA, 2010).

Na sua forma mais popular, um Spam consiste numa mensagem de e-mail com fins publicitários ou mal intencionados. O termo Spam, no entanto, pode ser aplicado a mensagens enviadas por outros meios e com outras finalidades. Geralmente, os Spams têm caráter apelativo e, na grande maioria das vezes, são incômodos e inconvenientes (ALMEIDA, 2010).

O primeiro registro oficial de uma mensagem eletrônica não solicitada enviada em massa ocorreu no CTSS do MIT. Pouco tempo depois da sua criação, Tom Van Vleck e Noel Morris implementaram o programa CTSS MAIL que permitia que usuários se comunicassem através de mensagens. Em 1971, um administrador do CTSS chamado Peter Bos utilizou o CTSS MAIL para enviar a mensagem pacifista intitulada “THERE IS NO WAY TO PEACE. PEACE IS THE WAY”, ou “Não há caminho para a paz. A paz é o caminho”. Quando Tom Van Vleck considerou tal comportamento como inadequado, Bos se defendeu dizendo que aquilo era importante (VAN, 2012).

A propaganda de hardwares da chamada família DECSYSTEM, em 1978, foi o conteúdo do primeiro registro de uma mensagem comercial não solicitada enviada em massa utilizando mensagem eletrônica.

Apesar do número considerável de reações indignadas e de subsequentes discussões a respeito, em parte devido ao fato de que a ARPANET era considerada de uso exclusivo para assuntos do governo norte-americano, a questão não foi considerada por todos de maior importância na época (TEMPLETON'S, 2013).

Embora os dois registros possam ser considerados como o início do Spamming, o termo Spam não foi associado ao envio de mensagens não solicitadas até a década de 80. O início exato do uso da palavra é incerto e alvo de muita especulação, mas existe um consenso de que ele provavelmente originou-se nos Multi-User Dungeons (MUDs), ambientes virtuais onde múltiplos usuários conectados por uma rede podem interagir e conversar (TEMPLETON'S, 2013).

Alguns relatos descrevem que o ato de prejudicar o sistema através do envio excessivo de dados, e o ato de enviar múltiplas mensagens com o objetivo de deslocar as mensagens de outros usuários para fora da tela, passou a ser conhecidos como Spamming, quando alguns usuários começaram a comparar esse comportamento ao dos vikings presentes em um quadro do grupo humorístico britânico Monty Python. Nesse episódio, um casal vai a um restaurante onde todos os pratos são servidos com Spam, uma marca americana de carne enlatada da empresa Hormel Foods Corporation (Figura 2). A mulher não gosta do alimento, mas não consegue nenhuma opção sem ele. Ao longo do diálogo, a palavra é repetida insistentemente pelos protagonistas, principalmente por um grupo de vikings presentes no local, que começa a cantar: “Spam, spam, spam, spam. Lovely spam! Wonderful spam!” atrapalhando a conversa no restaurante da mesma maneira que as mensagens não solicitadas atrapalhavam a comunicação nos MUDs (TEMPLETON'S, 2013).

O quadro foi escrito para ironizar o racionamento de comida ocorrido na Inglaterra durante a Segunda Guerra Mundial. Spam foi um dos poucos alimentos excluídos desse racionamento, o que eventualmente levou as pessoas a enjoarem do produto e motivou a criação do quadro (TEMPLETON'S, 2013).

Figura 2 - Latas de carne suína da marca “SPAM”

Fonte: (MARDWARE, 2013).

Na década de 90, o termo spam tornou-se popular quando passou a ser usado na rede USENET, o maior sistema de grupos de notícias e listas de discussão da época. Entretanto, outras mensagens não solicitadas já haviam sido enviadas anteriormente na USENET. Em

1988, Rob Noha enviou para diversas listas de discussão um pedido de doações para seu fundo de faculdade e David Rhodes iniciou no mesmo período a circulação de uma corrente eletrônica conhecida como “Make Money Fast”. Contudo, o primeiro uso conhecido da palavra spam na USENET para designar esse tipo de comportamento foi feito por Joel Furr após um episódio em 1993 que ficou conhecido como “ARMM Incident”. Nesse incidente, um software experimental chamado ARMM, criado por Furr para moderar mensagens inadequadas em listas de discussão, acidentalmente enviou recursivamente dezenas de mensagens para a lista news.admin.policy devido a uma falha de implementação. Joel Furr lamentou o ocorrido e declarou que “não era sua intenção enviar Spam para as listas” (TEMPLETON'S, 2013).

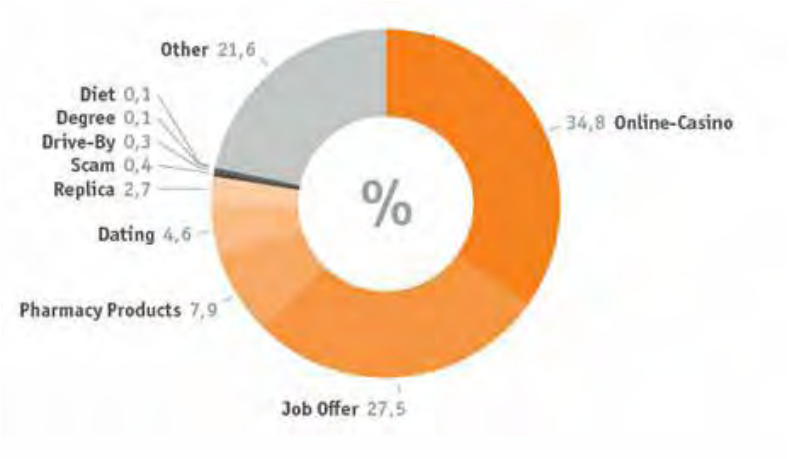
Em 18 de Janeiro de 1994 foi enviado o primeiro e mais importante spam. Clarence Thomas, administrador do sistema da Andrews University, enviou a todos os grupos de notícias uma mensagem religiosa intitulada “Global Alert for All: Jesus is Coming Soon”. E quatro meses depois, em 12 de Abril de 1994, Laurence Canter e Martha Siegel, dois advogados da cidade norte-americana de Phoenix, que trabalhavam em casos de imigração, enviaram uma mensagem anunciando serviços que teoricamente ajudavam as pessoas a ganharem vistos de permanência (Green Card) nos Estados Unidos. Por causa disso, a mensagem ficou conhecida como Green Card Spam e, já na época, imediatamente gerou as mesmas reações do spam atual, com questionamentos sobre ética e legalidade da prática. Apesar de não se tratar de uma mensagem inédita, eles usaram uma tática inovadora: contrataram um programador para criar um script simples e enviar o anúncio da dupla para todos os milhares de grupos de notícias da USENET. O esquema deu certo e todos receberam o primeiro spam comercial em larga escala da história (TEMPLETON'S, 2013).

Diversas pessoas referiram-se à mensagem como Spam, e o termo passou a ser utilizado em qualquer outra instância de comportamento análogo. Em poucos anos, com a popularização cada vez maior da Internet, o envio de mensagens não solicitadas passou a crescer no ambiente do correio eletrônico, em parte estimulado pela existência dos programas para envio automático de mensagens, e se expandiu rapidamente para os outros meios disponíveis (TEMPLETON'S, 2013).

Dentre os meios de propagação de spam podem ser destacados: e-mails, mensagens instantâneas, grupos de discussão e fóruns, mensagens de texto em celulares, jogos on-line, sites de busca, blogs, e sites de compartilhamento de arquivos e vídeos.

No ano 2012, os assuntos mais abordados por Spams foram: Cassinos online (34,8%), ofertas de emprego (27,5%), produtos farmacêuticos (7,9%), namoro (4,6%) entre outros assuntos apresentados na Figura 3.

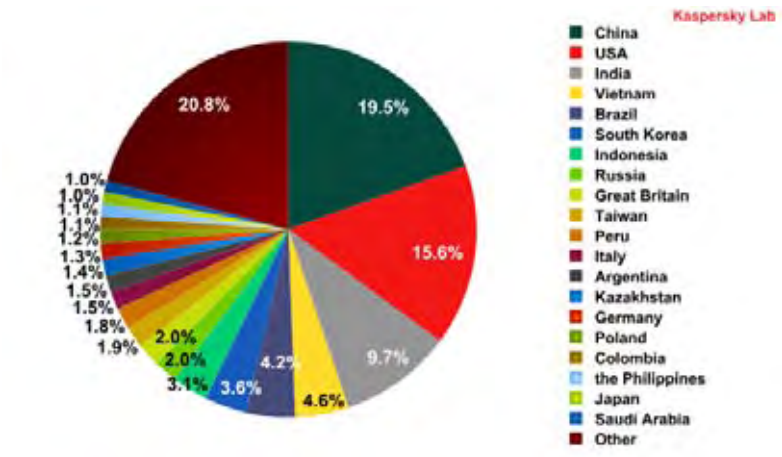
Figura 3 - Assuntos mais abordados nos SPAMS



Fonte: (ELEVEN, 2012).

Segundo o boletim de segurança divulgado em 2013 pela Kaspersky Security, 4,2% dos spams enviados diariamente partem do Brasil. A Figura 4 apresenta a porcentagem de envio dos principais países de origem de spams (GUDKOVA, 2013).

Figura 4 - Spams por país de origem



Fonte: (GUDKOVA, 2013).

A consequência direta do Spam é o consumo de recursos de máquina e de rede dos provedores de Internet e dos usuários, além do custo do tempo e da atenção humana gastos para selecionar e apagar as mensagens não solicitadas. Indiretamente, há o custo gerado pelas vítimas dos crimes que geralmente acompanham os spams, tais como roubo financeiro, roubo de identidade, roubo de dados e propriedade intelectual, vírus e outras infecções por malwares (códigos maliciosos), pornografia infantil, fraudes e propagandas enganosas (LEVINE, 2013).

Os Spammers frequentemente utilizam falsos nomes, endereços, números de telefones, e outras informações para contato, necessários para criar e configurar contas de e-mail na maioria dos provedores de Internet. Em muitos casos, eles usam cartões de crédito falsificados

ou roubados para pagar por essas contas. Dessa forma, quando o provedor detecta e remove uma conta eles podem reagir rapidamente criando outras novas (ALMEIDA, 2010).

O custo associado aos Spams também inclui gastos colaterais causados pelo conflito constante entre os Spammers, e os administradores e usuários do meio atacado pelo Spamming. Além disso, o conteúdo do Spam frequentemente é ofensivo, já que muitos produtos anunciados envolvem pornografia. Uma vez que os Spammers enviam seus Spams indiscriminadamente e em larga escala, anúncios pornográficos podem ser expostos em locais de trabalho ou até mesmo a crianças (ALMEIDA, 2010).

Os Spams exemplificam uma "tragédia dos comuns": Spammers usam recursos (físicos e humanos) sem se preocuparem com o custo total. Com isso, o custo acaba sendo dividido com cada usuário. Em alguns aspectos o Spam pode ser visto como uma grande ameaça ao sistema de e-mail atual (ALMEIDA, 2010).

Uma vez que o custo de envio de e-mail é extremamente baixo, um grupo reduzido de Spammers pode saturar a Internet com mensagens não solicitadas. Apesar de somente uma pequena porcentagem dos seus alvos ser motivada a comprar os seus produtos, o baixo custo pode determinar um ganho suficiente para manter o Spamming ativo (ALMEIDA, 2010).

Os Spammers conseguem obter lucros, mesmo com um índice de resposta de um para cada 12,5 milhões de e-mails enviados. Os autores conseguiram se infiltrar em uma rede que envia Spams e estudar a sua estrutura econômica. A análise sugere que um baixo índice de respostas é suficiente para produzir milhões de dólares de lucro por ano e pessoas que enviam Spams estão mais suscetíveis a ataques de hackers. Os autores se infiltraram na rede Storm, que utiliza computadores de usuários domésticos "sequestrados" para enviar Spam – ou seja, o computador do usuário envia Spams sem que o seu dono perceba. No seu ápice, acredita-se que o Storm controlou até 1 milhão de computadores pelo mundo. Os autores acompanharam os Spams enviados por 75869 computadores sequestrados – apenas uma pequena fração da rede Storm. Um dos Spams promovia o site de uma farmácia que oferecia remédios para aumento da libido. O site era falso e foi criado pelos próprios autores. A página da farmácia falsa retornava uma mensagem de erro cada vez que alguém tentava comprar remédios com cartão de crédito. Foram enviadas 469 milhões de mensagens de Spam, a grande maioria promovendo a falsa farmácia. Depois de 26 dias, apenas 28 vendas foram realizadas. O índice de resposta da campanha foi inferior a 0,00001% – muito abaixo do índice de 2,15% prometido por serviços privados e legalizados de envio de mensagens. O lucro gerado foi de US$ 2.731,88 – um pouco mais de US$ 100 por dia, no período computado. Ao expandir esse nú-

mero para a totalidade da rede Storm, os autores indicam que apenas essa rede de spams produz cerca de US$ 7.000 por dia, ou US$ 3,5 milhões por ano (ALMEIDA, 2010).

A quantidade de Spams em circulação continua aumentando de forma preocupante e, por enquanto, não apresenta sinais de enfraquecimento. O número de Spams que os usuários veem em suas caixas de mensagens é apenas a ponta do iceberg, uma vez que as listas de e-mails dos Spammers contém uma enorme porcentagem de endereços inválidos e muitos filtros simplesmente recusam ou apagam os Spams óbvios.

Em relação ao custo gerado pelo Spam, a Comissão de Comércio Interno da União Europeia estimou em 2001 que o Spam custou, somente naquele ano, cerca de 10 bilhões de euros aos usuários da Internet (ALMEIDA, 2010).

Segundo o US Technology Readiness Survey, produzido pelo Centro de Excelência em Serviço da Universidade de Maryland, o custo gerado pela perda de produtividade por causa do Spam, em escala global, foi estimado em US$ 50 bilhões, somente em 2004. Em 2009, estimou-se que mais de US$ 130 bilhões foram perdidos, em escala global, por causa do Spam (ALMEIDA, 2010).

Contudo, apesar do conhecimento e das calorosas discussões sobre o mal ocasionado pelo Spam, nenhuma solução realmente eficiente é adotada e, ao que tudo indica, não será, pelo menos em breve. Acredita-se que, a menos que o sistema convencional de e-mail passe por profundas reestruturações, o problema continuará oferecendo sérios riscos à existência e manutenção do próprio sistema (ALMEIDA, 2010).

4 Redes Neurais Artificiais ART

4.1 Introdução

As RNAs são modelos computacionais inspirados no sistema nervoso dos seres vivos. Possuem a capacidade de aquisição e manutenção de conhecimento e podem ser definidas como um conjunto de unidades de processamento, caracterizadas por neurônios artificiais separados entre camadas (entrada, oculta e saída) (SILVA, 2012).

As principais características das RNAs são: aprendizado e generalização, processamento paralelo, não linearidade, robustez e flexibilidade (HAYKIN, 1999). Os tipos de aprendizado mais utilizados de uma rede neural são realizados da forma supervisionado ou não supervisionado. No treinamento supervisionado para cada amostra dos dados de entrada obtém-se a respectiva saída desejada. Já a rede neural que possui treinamento não supervisionado é auto organizada, sendo capaz de descobrir estaticamente, padrões relevantes aos dados de entrada. O apêndice A, descreve de forma mais detalhada os principais fundamentos das redes neurais.

Carpenter e Grossberg foram os responsáveis pela construção do grupo de modelos de redes neurais com capacidade de dar a sistemas de inteligência artificial a possibilidade de simular com mais exatidão comportamentos humanos. Esses novos modelos, conhecidos como redes neurais artificiais da família ART, foram frutos de mais de quatro décadas em teorias sobre o funcionamento do cérebro e inteligência artificial (CAPUANO, 2009; AMORIM, 2006).

Grossberg defende que uma rede neural artificial verdadeiramente semelhante à humana deveria ser “autônoma, de aprendizagem rápida e adaptável”; isso significa uma rede capaz de aprender rapidamente como organizar e manejar um mundo pleno de surpresas (AMORIM, 2006). Dessa maneira surge o dilema plasticidade/estabilidade, o qual Carpenter e Grossberg se propuseram a resolver com seus novos modelos e melhorar a eficiência das redes neurais artificiais.

4.2 Dilema Estabilidade/Plasticidade

A estabilidade está relacionada com a garantia de agrupamento de todos os elementos nas categorias criadas pelo modelo, tendo em vista que os pesos da rede possuem a característica somente de decrescimento, ou seja, à medida que as adaptações dos pesos são realizadas, os

novos valores tendem sempre a diminuir até a estabilização. A plasticidade é a característica que a rede possui de aprender um novo padrão, em qualquer tempo de sua operação, sem perder o aprendizado adquirido anteriormente (GROSSBERG, 1987).

O dilema da plasticidade-estabilidade tem como suas principais questões:

I. Como um sistema de aprendizagem pode manter sua plasticidade (ser adaptável) em resposta às informações novas, ainda não conhecidas, e manter sua estabilidade diante da apresentação de informação irrelevante?
II. Como pode um sistema preservar seu conhecimento adquirido e ao mesmo tempo ser suficientemente flexível para armazenar uma nova informação?
III. Como o sistema pode decidir quando alternar do estado de estabilidade à plasticidade e vice-versa?

A resposta a este dilema está na Teoria da Ressonância Adaptativa (Adaptive Resonance Theory, ART) que será apresentada na próxima seção.

4.3 **Teoria da Ressonância Adaptativa**

A Teoria da Ressonância Adaptativa (ART) é um paradigma de rede neural desenvolvido no Centro de Sistemas Adaptativos da Universidade de Boston (EUA) por Carpenter e Grossberg, tendo como principal característica sua similaridade com os processos de aprendizado humano. Esse paradigma se contrapõe ao tradicional, baseado na lógica de primeira ordem e na heurística, pois busca na própria natureza os processos de aprendizado, entendendo que os seres vivos sobrevivem porque aprendem a se adaptar continuamente ao ambiente mutante (CAPUANO, 2009).

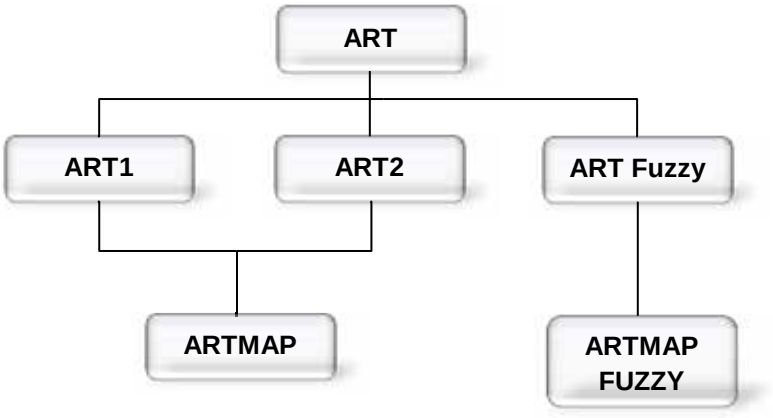
O conceito de ressonância corresponde a “oscilações” que ocorrem em um sistema físico quando o mesmo é submetido a uma entrada de frequência específica (frequência de ressonância), usualmente definida pelas propriedades físicas do sistema (materiais, dimensões, disposição geométrica, etc.). Em uma rede ART, a aprendizagem ocorre quando as informações, na forma de saída das unidades de processamento (neurônios), “oscilam” entre as camadas de unidades da rede, desenvolvendo um estado de ativação de equilíbrio. Essas “oscilações” seriam o equivalente neural à ressonância em um sistema físico. (AMORIM, 2006).

As redes ART, portanto, são inspiradas na inteligência artificial baseada na natureza, que é pensada em termos de três princípios elementares:

I. a melhor maneira de se compreender como funciona a inteligência humana é estudar, antes, como funciona um modelo de inteligência mais elementar, como a inteligência animal;
II. a inteligência pode ser emergente, uma propriedade da interação complexa de elementos mais simples;
III. a inteligência é demasiadamente complexa para que possa ser projetada a partir do zero.

O trabalho de Carpenter e Grossberg a princípio gerou três classes de arquiteturas ART: ART1, ART2, e ART Fuzzy. Apresenta-se na Figura 5, um organograma da família das principais redes neurais ART e descendentes (MALANGE, 2010).

Figura 5 - Algumas das principais redes neurais da família ART



Fonte: Próprio autor

4.4 **Principais Características e Fundamentos**

A rede ART é um sistema que auto-organiza padrões de entrada em categorias de reconhecimento, mantendo um equilíbrio entre as propriedades de plasticidade e de estabilidade (MALANGE, 2010).

Possui um mecanismo de vigilância que administra a inclusão de novas entradas em cada grupo. Uma regra de similaridade, que define onde agrupar um padrão, é determinada por um grau de semelhança entre um padrão previamente armazenado (MALANGE, 2010).

A rede ART é baseada em um sistema de aprendizagem competitivo não-supervisionado que auto-organiza categorias em resposta a sequências arbitrárias de padrões de entrada, em tempo real, para reconhecimento de padrões, tendo como característica importante a combinação de padrões, onde o padrão de entrada atual é comparado com uma repre-

sentação de categoria selecionada. Usualmente no reconhecimento é utilizado as representações de código WTA (Winner Take All- O vencedor leva tudo) (AMORIM, 2006).

No treinamento, um padrão é inserido na rede através da camada de entrada, codificado e passado para a camada de saída com as conexões de pesos da rede de filtragem adaptativa. Nela inicia-se a dinâmica do sistema com o treinamento competitivo (MALANGE, 2010).

O centro computacional neural para ambas as análises, científicas e tecnológicas, é a "regra de coincidência ART", que representa a interação entre a expectativa aprendida descendente (top-down) e a entrada sensorial ascendente (bottom-up). Esta interação gera um ponto de atenção, o qual, sucessivamente, determina a natureza da memória armazenada (AMORIM, 2006).

As redes ART são, então, redes neurais desenvolvidas para codificar reconhecimentos estáveis, em tempo real, através da auto-organização, em resposta a sequências arbitrárias de padrões de entrada (AMORIM, 2006).

As redes neurais da família ART apresentam algumas dificuldades operacionais: razoável sensibilidade aos parâmetros da rede (parâmetro de vigilância, etc.) e a precisão das análises. Os efeitos da sensibilidade paramétrica têm sido resolvidos, ou atenuados, através do emprego de novas concepções de treinamento e de arquitetura das redes neurais ART. Igualmente, várias propostas têm sido apresentadas na literatura especializada visando sanar o problema da imprecisão. Grande parte da imprecisão decorre em consequência, principalmente, do mecanismo de escolha de categorias e do teste de vigilância que necessita de aperfeiçoamento (MALANGE, 2010).

4.5 **Arquitetura da Rede Neural ART**

A arquitetura da rede ART é constituída por dois subsistemas principais: o subsistema de atenção e o subsistema de orientação, que juntos implementam a verificação de similaridade entre o padrão apresentado e um padrão já representado por algum neurônio já treinado, e habilita ou não o treinamento ou retreinamento de um determinado neurônio. Além delas, existe uma camada intermediária que funciona como uma rede de filtragem adaptativa entre as camadas de entrada e de saída. Para cada camada existem sinais de controle visando manipular o fluxo de dados.

O subsistema de atenção é estruturado por dois campos de neurônios: o campo F_1, cuja função é processar os dados de entrada, e o campo F_2, que classifica os padrões de treinamento em categorias de reconhecimento, em que cada campo pode ser constituído por vários neu-

rônios. Tais campos são conectados com pesos de conexão do tipo feedforward (w_i) e feedback (w_j) que são responsáveis pelo armazenamento das informações através de um processo que envolve a escolha da categoria, critério de equalização e treinamento. Os neurônios de F_2 são de dois tipos: os comprometidos – armazenam as informações dos padrões já apresentados à rede, representando categorias reais – e os descomprometidos (ou livres) – representam os nós ainda não utilizados, no processamento, ou seja, a memória "em branco" da rede (CARPENTER ; GROSSBERG ; REYNOLDS, 1991).

O subsistema de orientação é controlado por um parâmetro de vigilância (ρ), que determina se um padrão de entrada pode ser incluído em uma das categorias existentes. (CARPENTER ; GROSSBERG; REYNOLDS, 1991).

A Figura 6 apresenta, com maiores detalhes, a formação dos pesos de conexão alocados entre os campos F_1 e F_2. Trata-se de uma conexão entre M componentes de atividade correspondentes aos componentes do vetor de entrada a (em F_1) e N neurônios da atividade F_2. O número N pode ser escolhido de forma arbitrária, porém, denota-se que este deverá ter um valor suficiente para abrigar todas as categorias classificadas em F_2. Para que não haja a possibilidade de N ser insuficiente, costuma-se adotá-lo igual ao número de padrões de entrada a serem armazenados, ou seja, uma para cada padrão. Tal ação, mesmo sendo eficiente, do ponto de vista teórico, pode-se tornar um problema no que se refere ao armazenamento de informações: mesmo que algumas categorias fiquem vazias, estas ocuparão espaços na memória do sistema (CARPENTER ; GROSSBERG; REYNOLDS, 1991).

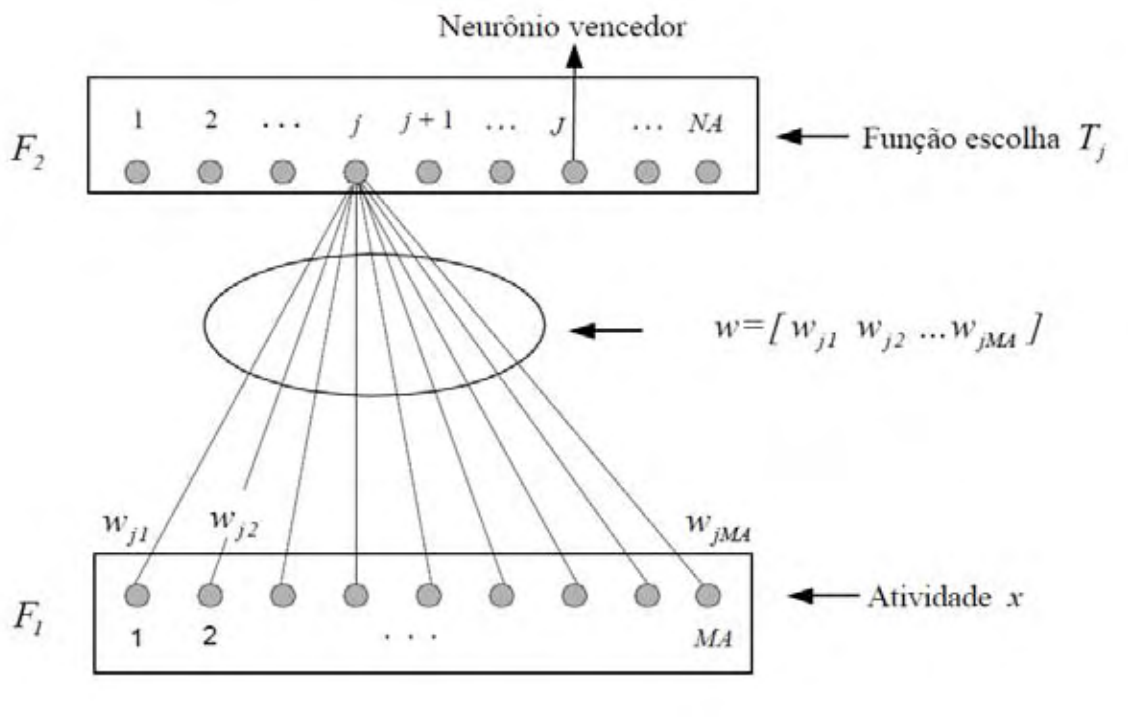
Assim, outra alternativa, de algoritmo mais extenso, porém, em casos de grande número de padrões de entrada, de maior eficiência, é a criação contínua de novos neurônios à medida que novas categorias são formadas. Nota-se que esta segunda forma de operação poderá ser bem mais rápida do que a primeira, pois ocupará menos espaço de memória. As duas formas poderão ser aplicadas em algoritmos ART, dependendo do tipo de função a que a rede será submetida (MARCHIORI, 2006).

O processo de classificação de padrões da rede ART consiste de quatro etapas principais: reconhecimento, comparação, busca e treinamento (CARPENTER; GROSSBERG; REYNOLDS, 1991), como a seguir:

Reconhecimento: Cada neurônio do campo F_1 (camada de entrada) recebe sinais de baixo para cima do vetor de entrada e de cima para baixo do campo F_2 (camada de saída, que representa a categoria ativa). A atividade é calculada e o vetor resultante é comparado com todos os vetores de pesos já armazenados na rede (memória) para encontrar o que mais se

assemelha ao padrão de entrada atual. O neurônio do campo F_2, com maior valor de ativação, é selecionado como provável categoria para armazenar o novo padrão. O neurônio selecionado envia ao campo F_1 seu protótipo;

Figura 6 - Formação do vetor peso da rede ART



Fonte: (SILVEIRA, 2003).

Comparação: Nesta fase acontece um mecanismo de reajuste, que é responsável por testar a similaridade entre o vetor de entrada e o vetor de comparação, w, que é o vetor resultante da atividade calculada no campo F_1 mediante o vetor protótipo dado pelo campo F_2 após a fase de reconhecimento. A maneira pela qual se define de que modo a comparação irá ocorrer é definida por um parâmetro chamado vigilância (ρ), que determina se um padrão de entrada pode ser incluído em uma das categorias existentes. Se o valor da comparação for maior do que ρ, então o padrão de entrada é incluído na categoria ativa, caso contrário, a rede entra em fase de busca;

Busca : Durante esta fase, a rede procura um novo neurônio do campo F_2 para representar o vetor de entrada atual. O neurônio da camada de saída rejeitado na fase de comparação anterior é desabilitado nesta seleção. O vetor de entrada é reapresentado, e a rede entra novamente na fase de comparação, que termina com o teste no mecanismo de reajuste para determinar a similaridade entre o novo protótipo escolhido e o vetor de entrada atual. Este processo é repetido, desabilitando neurônios da camada de saída, até encontrar um neurônio de saída que melhor se assemelhe ao vetor de entrada corrente, dentro dos limites do parâme-

tro de vigilância ρ. Se nenhum neurônio de saída for encontrado, o vetor de entrada é então considerado de uma classe desconhecida, sendo alocado um neurônio de saída que não esteja associado ainda a nenhuma categoria para representá-lo;

Treinamento : O algoritmo de aprendizado da rede ART é não-supervisionado e pode ser ativado a qualquer momento, permitindo que a rede aprenda novos padrões continuamente.

Há dois tipos de treinamento para a rede neural ART: aprendizado rápido e aprendizado lento. No aprendizado rápido, os pesos de conexão são ajustados para seus valores ótimos em poucos ciclos, geralmente em apenas um ciclo de treinamento. No aprendizado lento, os pesos são ajustados lentamente em vários ciclos de treinamento, possibilitando um ajuste melhor dos pesos da rede aos padrões de treinamento (GROSSBERG, 1976).

As redes ART são muito sensíveis a variações em seus parâmetros durante o treinamento. O parâmetro mais crítico é o parâmetro de vigilância (ρ) que controla a resolução do processo de classificação. Se ρ assume um valor baixo, a rede permite que padrões não muitos semelhantes sejam agrupados na mesma categoria de reconhecimento, criando poucas classes, e se for atribuído um valor alto (muito próximo a um), pequenas variações nos padrões de entrada levarão à criação de novas classes (MALANGE, 2010).

Tais etapas podem ser enxergadas de modo generalizado, já que este é, praticamente, o procedimento completo da maior parte das redes pertencentes à família ART, como a rede ART1 – a mais conhecida e utilizada da família (GROSSBERG, 1987). Porém, há algumas variações, em que os resultados poderão ser melhorados ou, também, utilizados na necessita de aplicações mais específicas para um determinado sistema. Dentre alguns modelos que fazem parte da família ART, destacam-se:

1. Rede neural ART1: rede que possui treinamento não-supervisionado. Possui a capacidade de reconhecer padrões de entrada binários de forma arbitrária (GROSSBERG, 1987);

2. Rede neural ART2: rede que também possui treinamento não-supervisionado e emprega tanto padrões de entrada binários como padrões de entrada analógicos (CARPENTER; GROSSBERG, 1987);

3. Rede neural ART Fuzzy: rede baseada no treinamento não-supervisionado e que engloba em sua arquitetura cálculos baseados na lógica fuzzy (CARPENTER; GROSSBERG; ROSEN, 1991).

4. Rede neural ARTMAP: rede que possui treinamento supervisionado. É composta por dois módulos ART interconectados, através do campo inter-ART. Esta rede também pode

identificar padrões de entradas binários ou analógicos (CARPENTER; GROSSBERG; REYNOLDS, 1991).

5. Rede neural ARTMAP Fuzzy: esta rede possui treinamento supervisionado como a rede neural ARTMAP, porém, todos os cálculos são fundamentados na lógica nebulosa (CARPENTER, 1992).

4.6 Evolução das Redes Neurais Artificiais Pertencentes a Família ART

Ao longo dos anos foram propostas distintas arquiteturas ART pertencentes aos dois principais paradigmas de aprendizagem neural: supervisionado e não supervisionado.

As primeiras redes ART usavam aprendizagem não supervisionada ou auto-organizativa, e entre elas tem que se mencionar ART1 (GROSSBERG, 1987) que foi desenvolvida para executar agrupamentos de padrões de valores binários, e ART2 (CARPENTER; GROSSBERG, 1987) desenvolvida para detectar regularidades nas sequências randômicas analógicas, emprega uma arquitetura computacionalmente onerosa que apresenta dificuldades na seleção de parâmetros.

Para superar estas dificuldades, Fuzzy ART (CARPENTER ; GROSSBERG; ROSEN, 1991) foi desenvolvido como uma generalização de ART1, também capaz de agrupar padrões de valores reais.

Nos anos posteriores foram propostos os modelos que usavam aprendizagem supervisionada, ainda que também permitiam a aprendizagem auto-organizativa. Entre estes modelos cabe destacar ARTMAP (CARPENTER; GROSSBERG; REYNOLDS, 1991), que atua de fundamento para todas as outras redes ART supervisionadas, ARTMAP Fuzzy (CARPENTER, et. al., 1992) que associa os padrões de E/S com os valores difusos e a partir dele, várias extensões foram desenvolvidas. Na Tabela 1 os principais modelos de redes ART e algumas de suas vertentes são apresentadas:

Tabela 1 - Marcos históricos da rede ART na literatura

Ano	Arquitetura	Referência
1976	GN	Grossberg
1987	ART1	Carpenter
	ART2	Carpenter
1990	ART3	Carpenter
1991	ART 2-A	Carpenter
	ARTMAP	Carpenter
	Fuzzy-ART	Carpenter
1992	Fuzzy-ARTMAP	Carpenter

	AFCL	Newton
	Fuzzy Min-Max Neural Network(FMMNN)	Simpson
1993	LAPART	Healy
	Simplified Fuzzy-ARTMAP	Kasuba
	Fusion ARTMAP	Asfour
1994	IAFC	Kim
1995	ART-EMAP	Carpenter
	Adaptive Hamming Net (AHN)	Hung
	ARAM	Tan
	PROBART	Marriot
1996	Gaussian-ARTMAP	Williamson
	FasArt	Cano-Izquierdo
	hART-J	Barfai
1997	Supervised AHN (SAHN)	Hung
	hART-S	Barfai
	Mart	Kim
	ART-LD	Zhou-J.
	FasBack	Cano-Izquierdo
	dART, dARTMAP	Carpenter
	Cascade-ARTMAP	Tan
	PNN-ART	Lim
1998	ARTMAP-IC	Carpenter
	LAPART2	Healy
	Multi-channel ART (MART)	Fernández-Delgado
	Boosted ARTMAP	Verzi
1999	Fuzzy (Supervised) AHN	Hung
	ARTEX	Grossberg
	FA Variant	Georgiopoulos
	Ofam	Dagher
	DfasArt	Parrado-Hernández
2000	MicroARTMAP (µARTMAP)	Gómez-Sánchez
	HART & HARTMAP	Anagnostopoulos
	FANNC	Zhou-Z.
2001	Ellipsoid ART (EA) & Ellipsoid ARTMAP (EAM)	Anagnostopoulos
	Safe µARTMAP	Gómez-Sánchez
2002	semi-supervised Ellipsoid ARTMAP (ssEAM)	Anagnostopoulos
	ART-C	He
2003	AFC	Sapozhnikova
	Default ARTMAP	Carpenter
2004	Simplex ARTMAP (SAM)	Gomes
	Analog-ART1	Rajasekaran
	ARTSTREAM	Grossberg
2005	GreyART	Yeh
	PolyTope ARTMAP (PTAM)	Gomes
	Overlapping PolyTope ARTMAP (OPTAM)	Gomes

Fonte: (AMORIM, 2006).

4.7 **Redes ART Não Supervisionada**

4.7.1 **ART1**

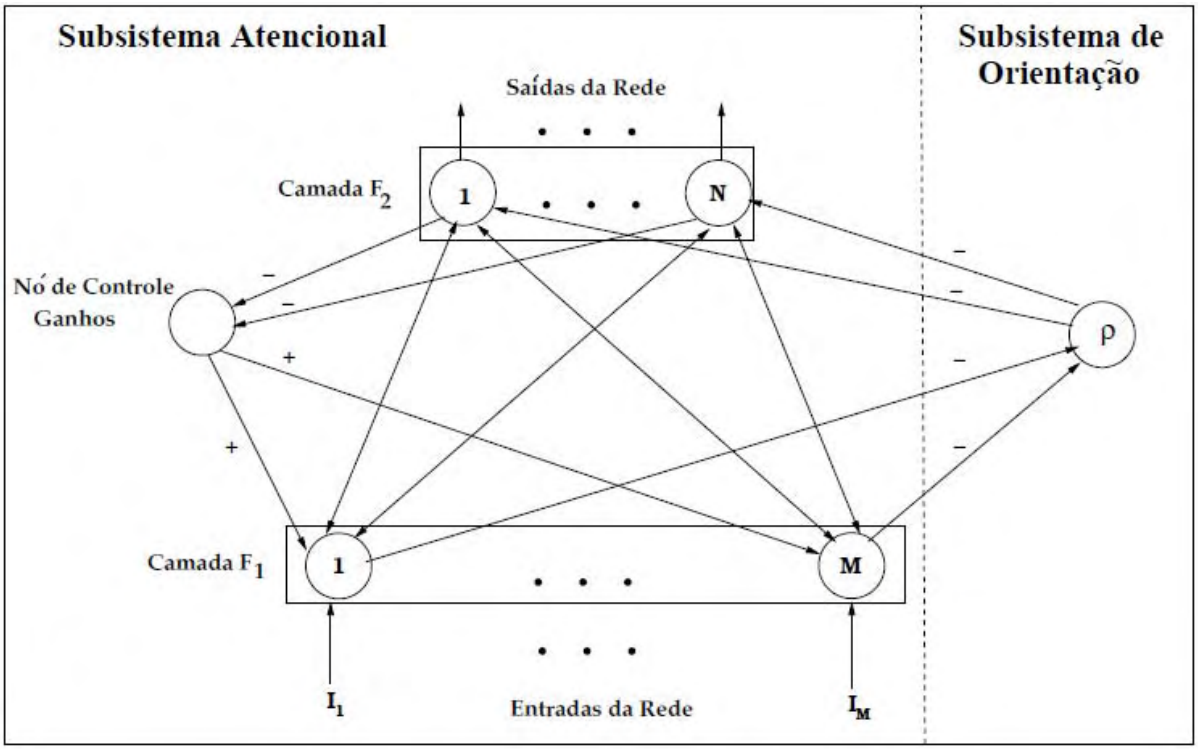
Utilizando como base a Grossberg Network, foi desenvolvido ART1 para aprendizagem não supervisionada de padrões de entrada binários, se tornando o mais famoso modelo da família ART, em termos de pesquisa relacionada e aplicações. Sua topologia básica encontra-se na Figura 7.

Em termos rigorosos a rede ART1 é caracterizada por um conjunto de equações diferenciais não-lineares, que implementam as propriedades de estabilidade-plasticidade, objetivando a aprendizagem incremental. Entretanto, para efeito da implementação computacional, esse modelo dinâmico de equações diferenciais pode ser resumido a um algoritmo sequencial comum, utilizando-se dos valores do regime estacionário das ativações da rede (FREEMAN; SKAPURA, 1991).

Arquitetura

O subsistema de atenção é formado por duas camadas de unidades denominadas F_1 e F_2, e pelo nó de controle de ganhos, como se encontra na Figura 7. O padrão binário de entrada I (cujas componentes $I_j \in \{0,1\}$) apresentam na camada F_1, são de natureza binária e de dimensão M. Os padrões de ativação desenvolvidos nestas unidades da rede são denominados traços de memória de curta duração (STM - Short-Term Memory), pois somente existem em decorrência da aplicação de uma entrada às unidades. Enquanto que na camada F_2, com N unidades de processamento (neurônios) associadas às distintas categorias internas (agrupamentos ou clusters) aprendidas pela rede, e seus pesos de conexão com F_1, integram os representantes ou valores esperados destas categorias. O fluxo de informações entre F_1 e F_2 é bidirecional, envolvendo pesos sinápticos bottom-up (de F_1 para F_2) e top-down (de F_2 para F_1), sendo, portanto uma rede com topologia feedback. Os pesos dessas conexões sinápticas entre as camadas F_1 e F_2 são denominadas de traços de memória de longa duração (LTM - Long-Term Memory), pois são responsáveis pela codificação das informações pela rede e representam a sua memória, que permanece quando se retira o padrão de entrada em F_1. Por último, o nó de controle de ganhos, que direciona o fluxo de informações entre as camadas, além de possuir um efeito excitatório sobre as unidades de F_1, recebe também um sinal inibitório proveniente de F_2 (AMORIM, 2006).

Figura 7 - Estrutura da rede ART1



Fonte: (AMORIM, 2006).

O subsistema de orientação (circuito de reset) é caracterizado pelo parâmetro ρ (de vigilância), e controla o processo de "busca por coincidência" de padrões e de aprendizagem da rede, atuando como inibidor das unidades em F_2. Somente no caso em que a atividade em F_1 seja suficientemente elevada, o subsistema de orientação se encontra desativado, e as unidades de F_2 podem se ativar (AMORIM, 2006).

Funcionamento

Quando a rede é inicializada para o treinamento, os pesos top-down das unidades da camada F_2 ainda não participaram de qualquer aprendizagem, e por isso as unidades são ditas não-rotuladas ou desativadas (uncommitted), e o nó de controle de ganhos está ativo. À medida que o treinamento prossegue, as unidades que já participaram de alguma aprendizagem, codificando os padrões em seus traços LTM top-down, passam a ser ditas rotuladas (committed) ou ativadas. As saídas de F_1 até o subsistema de orientação estão regidas pela regra de 2/3, ou seja, se ativando somente se 2 de suas 3 entradas se encontram ativas simultaneamente. O fato do nó de controle de ganhos se encontrar ativo e existir um padrão na entrada, faz com que a atividade inibidora sobre o subsistema de orientação seja suficiente para sua desativação. A propagação do padrão de entrada de F_1 a F_2 (propagação ascendente) determina as funções de escolha de categoria (Category Choice Function, FEC) $T_j(I)$, que representa a si-

milaridade entre o padrão de entrada e o vetor de pesos W_j da unidade j em F_2 (AMORIM, 2006):

$$T_j(I) = \frac{|I \cap W_j|}{|W_j|} \tag{3.7.1.1}$$

em que ∩ designa o operador AND binário e $I = (I_1, ..., I_M)$ é o padrão de entrada (M é a dimensão do espaço de entrada). O W_j é o vetor representante (valor esperado ou protótipo) da categoria interna j associada a unidade j de F_2, e |x| designa a soma dos componentes do vetor $x = (x_1, \dots, x_M)$:

$$|x| = \sum_{i=1}^{M} |x_i| \tag{3.7.1.2}$$

A camada F_2 possui um mecanismo de ativação competitivo (competição "Winner-Takes-All", WTA), onde somente a unidade J com entrada máxima T_j é a ganhadora (em caso de empate, ganha a categoria J com $T_J = \max_j\{T_j\}$ e índice J mínimo). A ativação exclusiva desta unidade, associada à categoria que apresenta uma função de escolha maior com o padrão de entrada, ocasiona a propagação descendente de seu representante W_j até a camada F_1 (AMORIM, 2006).

Por outro lado, a ativação da unidade ganhadora J em F_2 inibe o nó de controle de aquisição, de forma que a saída inibidora desde F_1 ao sistema de orientação é uma medida de coincidência ("matching") $m_j(I)$ entre os dois vetores de entrada a F_2: I ascendente, e w_j (representante (LTM) da classe ganhadora J) descendente. Esta coincidência está determinada pela expressão (AMORIM, 2006):

$$m_j(I) = \frac{|I \cap w_j|}{|I|} \tag{3.7.1.3}$$

Se $m_j(I)$ é superior ao parâmetro de vigilância ρ ($0 \leq \rho \leq 1$), se alcança ressonância entre o padrão de entrada e o valor esperado para a unidade ganhadora, situação caracterizada por uma coincidência $m_j(I)$ entre ambos os vetores (I e w_j) superior ao mínimo ρ (vigilância) exigido pela rede. O estado de ressonância se prolonga até a apresentação do padrão de entrada seguinte. Durante este intervalo de tempo, o valor esperado w_j da unidade resso-

nante J atualiza-se mediante um AND binário, componente a componente, com o padrão de entrada I, conforme equação 3.7.1.4.

$$w_j(n+1) = (1-\beta)w_j(n) + \beta(I \cap w_j(n)) \qquad (3.7.1.4)$$

em que $0 \leq \beta \leq 1$ é a velocidade de aprendizagem; no caso especial de $\beta = 1$, chama-se de aprendizagem rápida (equação 3.7.1.5).

$$w_j(n+1) = I \cap w_j(n) \qquad (3.7.1.5)$$

Se, ao contrário, o matching $m_j(I)$ é inferior a vigilância ρ, o subsistema de orientação se ativa, inibindo a unidade ganhadora em F_2 (rejeição ou reset). O mecanismo competitivo determina uma nova unidade ganhadora e, repete-se o processo até que se alcance a ressonância com alguma das unidades associadas às categorias, já aprendidas pela rede. Se isso não ocorre, porque todas as unidades estão rejeitadas, a rede cria uma categoria nova e associa-se a uma unidade livre (quer dizer, não associada a nenhuma categoria) J' em F_2, igualando-se seu representante (ou valor esperado) $w_{j'}$ ao padrão de entrada atual, $w_{j'} = I$ (AMORIM, 2006).

4.7.2 **ART Fuzzy**

A rede ART Fuzzy constitui uma evolução da rede ART1, orientada a categorizar de forma estável padrões de entrada analógicos com componentes compreendidos entre 0 e 1. Por isso ART Fuzzy substitui os operadores de intersecção (∩) e de união (∪) de ART1 pelos operadores difusos MIN(∧) e MAX(∨), respectivamente, da teoria de lógica difusa. Esta mudança, com a ajuda da codificação em complemento (complement coding), preserva a amplitude da informação ao mesmo tempo que normaliza os vetores de entrada, permite implementar o algoritmo de classificação não supervisionado com grande rapidez de aprendizagem.

Em comparação com a rede ART1, a rede ART Fuzzy possui duas diferenças: (1) No que se refere aos traços LTM, o fluxo bidirecional de informações entre as camadas F_1 e F_2 é realizado por um único conjunto de pesos sinápticos, w_j; (2) a rede ART Fuzzy necessita de um tipo específico de pré-processamento das entradas, de modo a evitar um problema inerente de proliferação de categorias. A solução proposta foi a utilização de padrões com norma constante, através da codificação em complemento. Significa que se duplica o número de entradas,

transformando o padrão de entrada $a = (a_1, \ldots, a_M)$ no $I = (a, a^c) = (a_1, \ldots, a_M, 1 - a_1, \ldots, a_m)$, garantindo assim que $|I| = M$, o qual evita o decrescimento nos valores dos representantes e a proliferação de categorias derivado do contínuo decrescimento na norma $|w_j|$ dos pesos de conexão por usar o operador MIN difuso ^ na sua atualização.

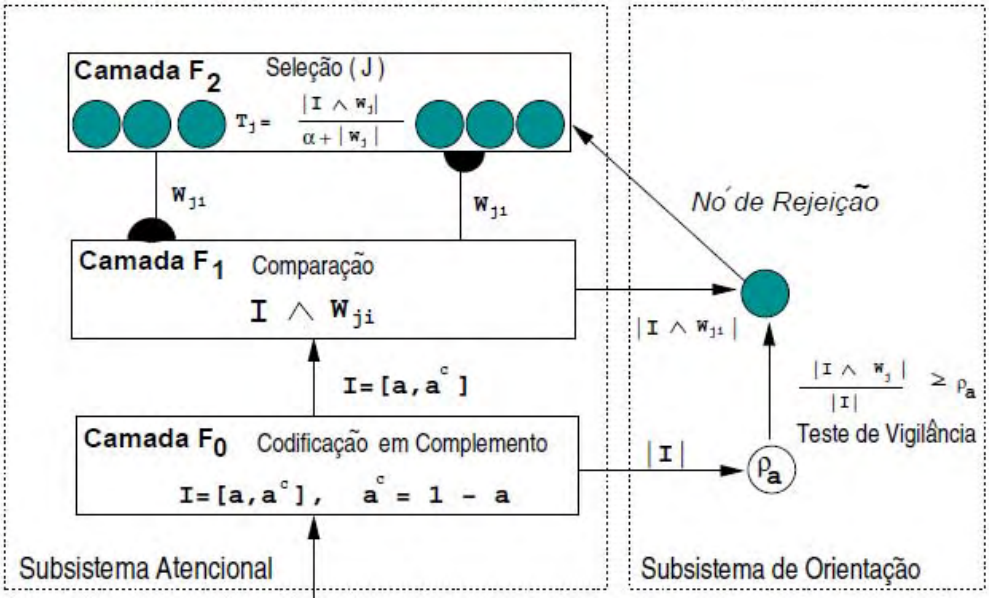
ART Fuzzy pode ser considerado um "método baseado a exemplos" para o agrupamento (clustering) de padrões de entrada. Estes padrões agregam-se em categorias ou agrupamentos com forma de hiper-retângulos (em geral, sobrepondo-se), que cobrem distintas regiões do espaço de entrada. A formação de grupos é uma maneira de compressão, na qual se formam regras abstratas sobre a distribuição dos dados que permite certa generalização. Por outro lado, ART Fuzzy usa a aprendizagem baseada em coincidência (matching) para incorporar padrões a uma categoria (AMORIM, 2006).

Arquitetura

O estudo precedente sobre a rede ART1 permite um entendimento mais direto da rede ART Fuzzy. Na Figura 8 é apresentada a arquitetura básica desta rede. ART Fuzzy possui como padrão de entrada a um vetor M-dimensional $(a_a, \ldots, a_M)$. A cada categoria interna lhe corresponde um vetor $w_j = (w_{j1}, \ldots, w_{j(2M)})$ de pesos adaptativos, onde $j = 1, \ldots, N$ e N é o número de possíveis categorias internas (ou unidades em F_2). O vetor de pesos de ART Fuzzy é o equivalente, em um único vetor, aos vetores de pesos top-down e bottom-up de ART1.

Assim, ART Fuzzy possui dois subsistemas, o de atenção e o de orientação, semelhante a ART1. O de atenção consiste em três camadas de unidades. Se a dimensionalidade do padrão de entrada é M, a camada do módulo F_0 tem M unidades e é uma etapa de pré-processamento a qual codifica o complemento do padrão de entrada que é entrada da camada F_1. A camada F_1 então possui 2M unidades. A camada F_2 é do tipo competitiva, e as suas unidades estão completamente interconectadas via conexões laterais e cada unidade caracteriza um retorno inibitório (negativo), esta também é chamada "camada de representação de categoria", já que cada unidade está associada a uma categoria interna; enquanto que o subsistema de orientação consiste de apenas uma unidade, a de rejeição (reset), que recebe as entradas de F_0. Sua função é inibir as unidades de F_2 durante a busca de categorias.

Figura 8 - Estrutura da rede ART Fuzzy



Fonte: (AMORIM, D. G., 2006).

Funcionamento

A rede ART Fuzzy possui além da estrutura, também o mecanismo de funcionamento muito semelhante a ART1. A Figura 8 apresenta um resumo das operações de ART1 e a sua correspondência em Fuzzy ART.

Durante a etapa de aprendizagem, assume-se que um novo padrão a é apresentado a rede. Depois de ser codificado seu complemento em F_0 de modo que $I = (a, a^c)$, este é propagado através da camada F_1 às unidades de F_2. Na aprendizagem, F_2 consiste à de unidades rotuladas (committed) e não-rotuladas, e sendo uma camada competitiva, todas as unidades competirão nos termos da função de escolha de categoria (FEC), definida pela equação 3.7.3.1.

$$T_j(I) = \frac{|I \wedge w_j|}{\alpha + |w_j|} \qquad (3.7.3.1)$$

em que ^ designa o operador *fuzzy* AND que é definido como $(x \wedge y)_i \equiv min(x_i, y_i)$.O α é o parâmetro de escolha ($\alpha \geq 0$). Esta função de escolha de categoria (FEC) é a denominada Lei de *Weber*. Após o cálculo de todas as FEC das unidades de F_2, seleciona-se a categoria (unidade de F_2) J com maior $T_j(I)$:

$$J = argmax_j\{T_j(I)\} \qquad (3.7.3.2)$$

Caso exista mais de um T_J máximo, a categoria j com menor índice é escolhida. A ressonância ocorre se a FEC da categoria ganhadora J satisfaz o teste de vigilância:

$$m_J(I) = \frac{|I \wedge w_j|}{|I|} \geq \rho \Rightarrow |I \wedge w_j| \geq \rho M \tag{3.7.3.3}$$

em que se considera $|I| = M$ pela codificação em complemento. Em caso de superar o teste de vigilância, o processo de aprendizagem é efetuado de forma análoga a ART1, mas com o operador ^, gerando o decrescimento na norma $|w_J|$ do vetor de pesos:

$$w_J(n+1) = \beta(I \wedge w_J(n)) + (1-\beta)w_J(n) \tag{3.7.3.4}$$

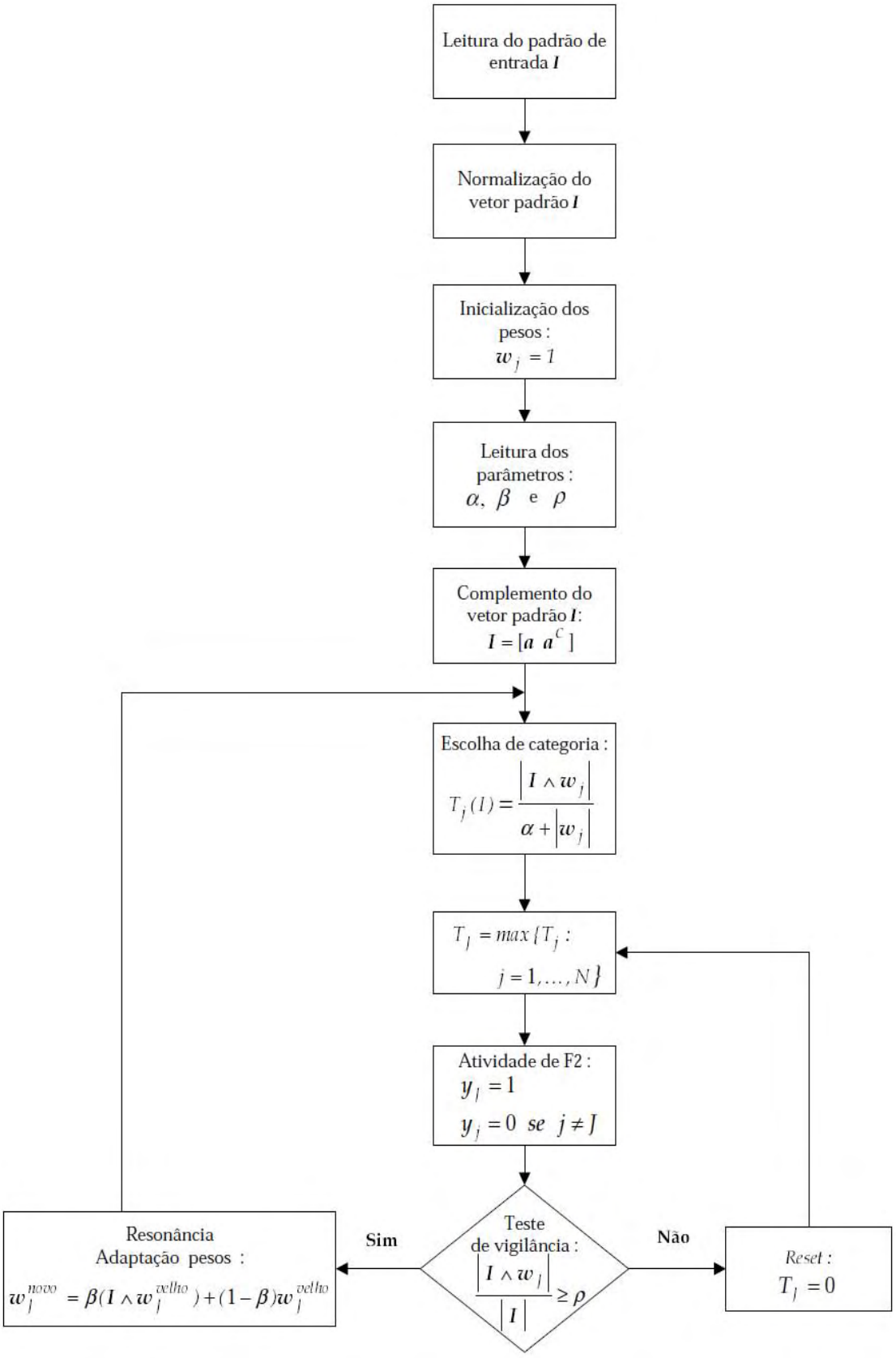
em que $\beta \in (0,1)$ é a velocidade da aprendizagem e é especificada na etapa de aprendizagem de ART *Fuzzy*. A aprendizagem rápida ocorre quando $\beta = 1$ (conforme equação 3.7.3.5).

$$w_J(n+1) = I \wedge w_J(n) \tag{3.7.3.5}$$

Caso a categoria vencedora não satisfaça o teste de vigilância, ela recebe um sinal de reset (rejeição) procedente do subsistema de orientação, permanecendo inativa, enquanto o padrão de entrada atual estiver presente. Neste caso, o processo de competição seleciona uma nova categoria ganhadora, com a que se repete o teste de vigilância na (equação 3.7.3.1). De modo similar à ART1, este processo continua até que a categoria ganhadora selecionada satisfaça o teste de vigilância ou, se nenhuma categoria o satisfaz, ART Fuzzy cria uma categoria nova J′, associada a uma unidade não rotulada (uncommitted) em F_2, cujo representante se iguala ao padrão de entrada: $w_{J'} = I$ (AMORIM, 2006).

Todos os procedimentos da rede ART Fuzzy estão descritas na Figura 9:

Figura 9 - Fluxograma da rede ART Fuzzy



Fonte: (LOPES, 2005).

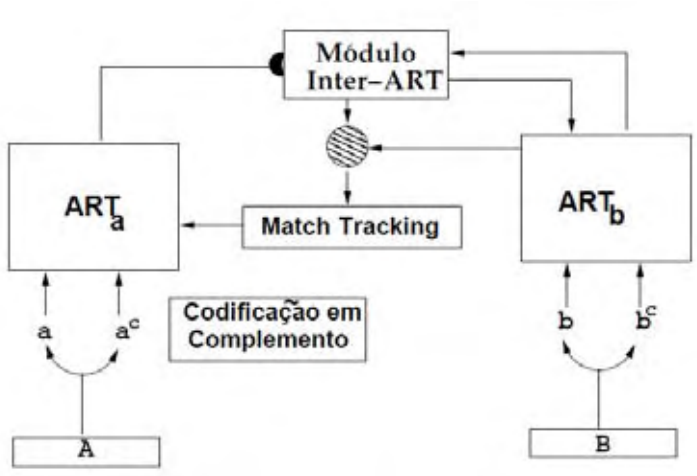
4.8 **Redes ART Supervisionadas**

Os modelos ART mais recentes permitem a aprendizagem supervisionada baseado nos princípios e conceitos (categorias internas, representantes de categorias, funções de escolha, aprendizagem competitiva, teste de vigilância, aprendizagem on-line) que aparecem nas redes ART não supervisionadas. Em geral, os modelos ART supervisionados, que veremos nas seguintes seções, permitem também a aprendizagem não supervisionada, mediante a desativação de certos módulos da rede, recuperando assim a funcionalidade dos modelos já vistos.

4.8.1 **ARTMAP**

Os modelos ART foram originalmente projetados para o agrupamento ou categorização não supervisionada de padrões, quer dizer, para o estabelecimento de uma correspondência entre um padrão de entrada e uma saída ativa (categoria ressonante). Porém, a rede ARTMAP (CARPENTER ; GROSSBERG; REYNOLDS, 1991) se orienta à construção de correspondências entre as múltiplas entradas e saídas, convertendo-se num associador de padrões. No caso particular em que um dos padrões seja a predição de saída desejada para o padrão de entrada, a rede permite o tratamento de problemas de classificação supervisionada de padrões. Por isso, baseia-se em duas redes ART1 (ART_a e ART_b) e um módulo F^{ab} (no módulo Inter-ART) entre as camadas F_2 de ambas as redes, como observa-se na Figura 10. Desativando os módulos ART_b e F^{ab} tem-se o funcionamento não supervisionado da rede ART1 original.

Figura 10 - Estrutura da rede ARTMAP



Fonte: (AMORIM, 2006).

A ARTMAP utiliza uma aprendizagem supervisionada onde se apresentam padrões nas camadas F_1^a e F_1^b. Estes padrões ressonam com as unidades em F_2^a e F_2^b, respectivamente,

e F^{ab} estabelece correspondências entre elas. Durante o treinamento, a ARTMAP permite a extração automática, a partir de um conjunto de amostras, de regras de decisão com a seguinte estrutura (AMORIM, 2006):

$$IF\ (A_1\ AND \ldots AND\ A_N)\ THEN\ (B_1\ AND \ldots AND\ B_M)$$

Os antecedentes $A_1, \ldots, A_N$ estão associados às unidades de F_1^a (entradas de ARTMAP), cada uma das quais representa uma condição, e os consequentes $B_1, \ldots, B_M$ se correspondem com as unidades de F_1^b (saídas de ARTMAP) e representam consequências inferidas a partir das condições. Por sua vez, durante o processamento ou teste, a ARTMAP opera no modo de sistema especialista, usando as correspondências aprendidas, durante o treinamento, para codificar os padrões de entrada pelos padrões de saída. Os antecedentes em F_1^a provocam a ressonância com uma unidade em F_2^a. Esta, por sua vez, determina a unidade ganhadora em F_2^b, cujo representante se propaga à saída F_1^b e, determina os consequentes $B_1, \ldots, B_M$. Estas regras proporcionam certa flexibilidade tanto que, distintamente dos sistemas expertos ou especialistas tradicionais, não requerem o cumprimento exato de todos os antecedentes, senão um cumprimento global aproximado com um nível de aproximação determinado pelo parâmetro de vigilância ρ. Por último, esta arquitetura permite a alternância entre as aprendizagens supervisionada e não supervisionada, proporcionando a qualidade na aprendizagem associada aos sistemas supervisionados, juntamente com a autonomia dos sistemas auto-organizáveis (AMORIM, 2006).

Uma rede de aprendizagem associativa e um controlador interno une estes módulos para fazer o sistema ARTMAP operar em tempo real. O controlador cria o número mínimo de categorias de reconhecimento em ART_a, ou "unidades ocultas", necessárias para atingir o critério de precisão. Uma regra de aprendizagem mínimo-máximo (minimax) permite a ARTMAP aprender rápida e eficientemente, minimizando o erro de treinamento e maximizando a compressão do código (AMORIM, 2006).

O módulo Inter-ART em ARTMAP forma associações entre as categorias através da saída da aprendizagem e disparos de buscas através de uma regra de "Match-Tracking". Quando a categoria ganhadora J em F_2^a está associada a um padrão de saída b' distinto do padrão desejado b, ARTMAP incrementa o parâmetro de vigilância ρ_a a mínima quantidade necessária para que a categoria J seja rejeitada em ART_a e uma nova categoria saia ganhadora. A base do parâmetro de vigilância $\bar{\rho}_a$ calibra um nível mínimo de confiança em que ART_a aceitará uma categoria selecionada. Os valores baixos de $\bar{\rho}_a$ permitem formar menos categorias, maximizando a compressão do código. Inicialmente $\rho_a = \bar{\rho}_a$. Durante a aprendizagem,

uma falha na predição em ART_b provoca o Match-Tracking, que incrementa ρ_a somente o suficiente para disparar uma busca em ART_a assim, corrigindo o erro de predição. O módulo ART_a seleciona uma nova categoria, que enfoca a atenção a um agrupamento de características do padrão de entrada que seja mais capaz de predizer b (AMORIM, 2006).

4.8.2 **ARTMAP Fuzzy**

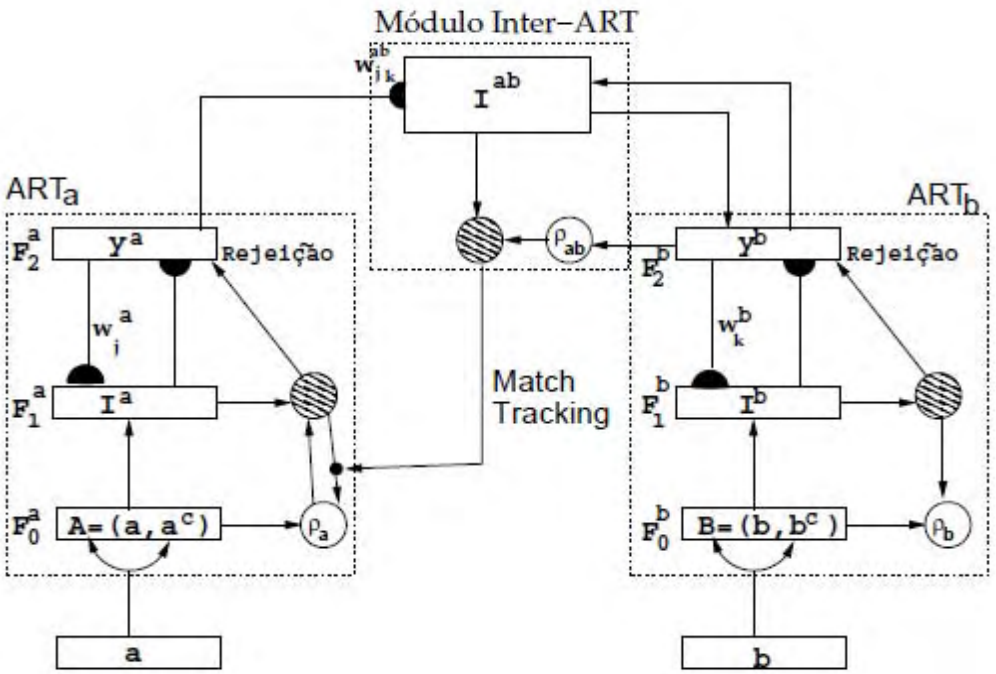
O sistema Fuzzy ARTMAP (CARPENTER et. al., 1992) aprende a associar padrões de entrada e saída com valores difusos, quer dizer, no intervalo [0, 1] e que portanto se pode interpretar como graus de pertinência aos conjuntos difusos. Esta generalização é alcançada pela substituição dos módulos ART1 do ARTMAP binário por módulos ART Fuzzy. A Figura 11 detalha a arquitetura de ARTMAP Fuzzy. A sua operação está baseada na similaridade entre o padrão de entrada I e os representantes w_j das categorias internas aprendidas pela rede. A função de escolha, Lei de Weber (equação 3.8.2.1), é a mais usada como FEC (Função de Escolha de Categorias) $T_j(I)$ entre os padrões e a categoria j, associada à unidade j em F_2^a:

$$T_j^{WL} = \frac{|I \wedge w_j^a|}{\alpha + |w_j^a|} \tag{3.8.2.1}$$

em que I é o padrão de entrada, codificada em complemento, w_j^a é o protótipo para a categoria interna j de ARTa, $\alpha \gtrsim 0$ e M é a dimensão do espaço de entrada. Ambas as funções definem implicitamente uma região geométrica hiper-retangular, cujas esquinas são definidas pelos componentes do vetor w_j (os valores máximo e mínimo dos componentes dos padrões de entrada codificados pela categoria j). Como em todos os modelos ART examinados, existe um processo de competição *WTA* em ART_a, no qual a categoria ganhadora J é aquela com maior $T_j(I)$: $J = argmax_j\{T_j(I)\}$. O mesmo ocorre em ART_b. O teste de vigilância determina se esta categoria ganhadora J alcança a ressonância com o padrão de entrada (equação 3.8.2.2) (AMORIM, 2006).

$$|I^a \wedge w_j^a| \geq \rho^a M \tag{3.8.2.2}$$

Figura 11 - Estrutura da rede ARTMAP Fuzzy



Fonte: (AMORIM, 2006).

assim como no caso de ART *Fuzzy*, se leva em consideração que $|I^a| = M$, pela codificação em complemento. O teste de vigilância é igual em ART_b operando sobre I^b, w_j^b e ρ^b. A aprendizagem em ARTMAP *Fuzzy* é similar ao caso de ART *Fuzzy*. Em caso de ressonância, o representante w_j^a da categoria ressonante J se aproxima ao mínimo $I \wedge w_J^a$ de modo similar à ART *Fuzzy* (equação 3.8.2.3).

$$w_J^a(n+1) = (1-\beta) w_J^a(n) + \beta(I^a \, w_J^a(n)) \qquad (3.8.2.3)$$

em que $0 \leq \beta \leq 1$ é a velocidade de aprendizagem (o caso $\beta = 1$ é a aprendizagem rápida). A mesma expressão é válida para ART_b. Como no caso de ART Fuzzy, a codificação em complemento faz com que w_J^a (de dimensão $2M$, sendo M a dimensão do padrão de entrada antes da codificação) codifique, nos seus M primeiros componentes, os valores mínimos dos componentes dos padrões de entrada codificados pela categoria J. Analogamente, os últimos M componentes de w_J^a são os valores máximos dos componentes dos padrões de entrada codificados pela categoria J. Portanto, o representante de uma categoria interna da ARTMAP Fuzzy, armazena os intervalos dos distintos componentes dos padrões codificados pela categoria.

Assim como em ARTMAP, o módulo Inter-ART forma associações entre as categorias de ART_a com as unidades em F^{ab}, através da saída da aprendizagem e disparos de busca através de uma regra de Match-Tracking. Estas unidades em F^{ab} representam as categorias em F_2^b, uma vez que as conexões entre F_2^b e F^{ab} possuem pesos iguais a 1. As categorias (unidades em F_2^a) representam grupos (clusters) de padrões de treinamento no espaço de entrada. Os vetores máximos v e mínimo u dos padrões de treinamento ressonantes com essa categoria são os valores esperados de cada categoria como em ART Fuzzy (AMORIM, 2006).

A aprendizagem dos pesos w^{ab} segue a mesma regra que em ART Fuzzy, conforme equação 3.8.2.4, mas empregando a ativação y_b em F_2^b ao invés de I.

$$w_J^{ab}(t+1) = (1-\beta)w_J^{ab}(t) + \beta(y_b \wedge w_J^{ab}(t)) \qquad (3.8.2.4)$$

em que, inicialmente $w_J^{ab}(t=0) = 1$ e y_b, em problemas de classificação, é a predição desejada para o padrão de entrada.

Se a aprendizagem é rápida $\beta = 1$, então w_J^{ab} somente varia no momento que se cria a categoria J. De fato, se supomos que a categoria J é a ganhadora e, se a predição desejada é K, então $y_b = w_J^{ab}$, ($w_{JK}^{ab} = 1$ e $w_{jk}^{ab} = 1, j \neq J$ ou bem $k \neq K$) e w_J^{ab} não varia. Se, ao contrário, a predição desejada não é K, então a categoria J é inibida pelo Match-Tracking e já não alcança a ressonância (AMORIM, 2006).

Quando a predição de ART_a é distinta de ART_b, a inibição do módulo Inter-ART F^{ab} ativa o processo de Match-Tracking, uma vez que satisfaz a condição de $|y^b \wedge w_J^{ab}| < \rho^{ab}|y^b|$, e assim, incrementa-se o valor da vigilância (ρ_a) em ART_a através da equação 3.8.2.5 [01]:

$$\rho_a = \frac{|I \wedge w_J^a|}{M} + \epsilon \qquad (3.8.2.5)$$

sendo $\epsilon \gtrsim 0$. Este processo rejeita a categoria J ganhadora em F_1^a e força as futuras categorias ganhadoras em ART_a ter um tamanho menor que $M(1-\rho_a)$ para superar o teste de vigilância. Se não existir uma categoria com estas características, cria-se uma nova categoria que associa-se a predição correta. Em tal caso, a categoria nova teria o hiper-retângulo (inicialmente um único ponto, coincidente com o padrão de entrada) dentro do hiper-retângulo associado à primeira categoria, e as duas categorias teriam predições distintas. Portanto, em AR-

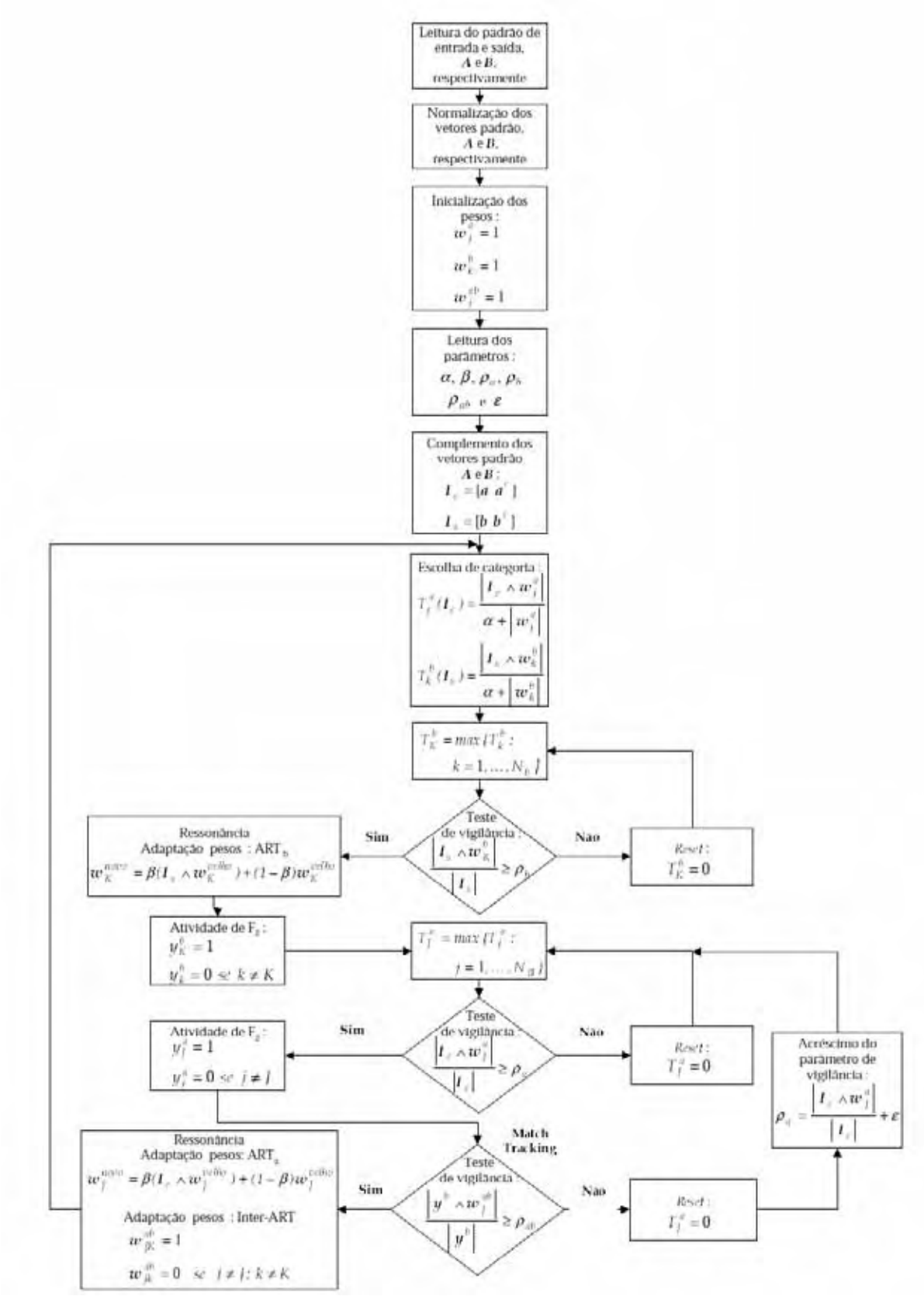
TMAP *Fuzzy* pode existir hiper-retângulos "aninhados" (sobrepostos completamente), e as categorias podem se sobrepor parcialmente, quando tenham predições distintas. De fato, se um padrão cai dentro dos hiper-retângulos de várias categorias, codifica-se por $\boldsymbol{a}$ de menor tamanho, e portanto, a mais específica destas duas. Considere as categorias C_1 e C_2 em ART_a. Se $R_1 > R_2$, como $R_j = M - |w_j|$, então $|w_1| < |w_2|$. Por outro lado, como $|I \wedge w_j| = |w_j| \; com \; j = 1, 2$ por cair $\boldsymbol{a}$ dentro dos dois hiper-retângulos, tem-se que:

$$T = \frac{|I \wedge w|}{\alpha + |w|} = \frac{|w|}{\alpha + |w|} \tag{3.8.2.6}$$

Esta função é crescente com $|w|$, e como $|w_1| < |w_2|$, chega-se a que $T_2 > T_1$ e a categoria com menor hiper-retângulo ganha a competição. Assim vai alcançar a ressonância, porque $|I \wedge w| = |w| > \rho M$ uma vez que, como sempre se verifica $R = M - |w| < M(1 - \rho)$, então $|w| > \rho M$ para todas as categorias (AMORIM, 2006).

Todos os procedimentos da rede ARTMAP Fuzzy estão descritas na Figura 12:

Figura 12 - Fluxograma da rede ARTMAP Fuzzy



Fonte: (LOPES, 2005)

5 Metodologia

5.1 Introdução

O objetivo desta pesquisa é desenvolver um modelo de identificação e classificação de e-mails indesejados utilizando os conceitos de extração e seleção de características proposto em Carpinteiro, Lima e Assis (2006) e o processo de classificação através da RNA ARTMAP Fuzzy.

A metodologia proposta, então, consiste nos seguintes procedimentos:

- ✓ Construção de um modelo de pré-processamento de dados para extração de características dos e-mails;
- ✓ Utilizar o método estatístico Frequency Distribution (FD) para seleção das características mais relevantes;
- ✓ Definição do número de entradas do módulo ART_a da RNA ARTMAP Fuzzy;
- ✓ Treinamento da RNA ARTMAP Fuzzy;
- ✓ Validação do modelo usado.

A linguagem de programação JAVA é utilizada em toda execução do trabalho, já que oferece recursos tanto para a realização de testes acadêmicos quanto para uma futura implementação em um sistema real. Algumas características que se pode citar oferecidas pela linguagem JAVA é sua simplicidade, compatibilidade com redes de computadores, portabilidade, Múltiplos Threads, entre outros recursos.

5.2 Pré-Filtro

O Pré-Filtro é de grande importância para o sistema de classificação. Além de eliminar informações irrelevantes e facilitar o treinamento da RNA tornando-a mais rápida, ele otimiza a qualidade dos dados possibilitando a rede neural obter resultados mais precisos, já que a rede identifica características que não seriam possível sem o pré-filtro. O processo pode ser dividido em quatro etapas (ASSIS, 2006):

- ✓ Processamento HTML e MIME;
- ✓ Tokenização;
- ✓ Detecção de Padrões;
- ✓ Seleção de Características;

Antes de enviado ao processamento de texto plano, ou seja, texto em seu formato original, é verificado no campo Content-Type do e-mail qual é seu tipo, e só assim enviado ao processamento adequado. Após processado adequadamente conforme seu tipo os dados são processados como texto plano.

Processamento HTML: O formato HTML permite adicionar ao corpo do e-mail, formatação de texto, tabelas, hiperlinks, imagens, etc. Essas personalizações são possíveis por uso das chamadas tags, palavras chaves que indicam uma instrução e possibilitam customizar como as informações do texto vão ser interpretadas pelos aplicativos. Seguem o seguinte padrão:

< nome-tag parâmetros > Texto da tag < nome-tag >.

A proposta de Carpinteiro, Lima e Assis (2006) foi a divisão das tags em três categorias de acordo com o grau de importância para o processo, com tipos de processamentos distintos. Seguem as três categorias:

1. Na primeira categoria tudo é ignorado, isto é, o nome da tag, seus parâmetros e conteúdos. Supõe-se que toda informação presente seja irrelevante.
2. Na segunda categoria as tags têm seus atributos removidos. A tag em si é substituída por outra específica, composta dos caracteres “!_in_” mais a tag.
3. Na terceira categoria a tag é processada integralmente. Neste caso o nome da tag, os parâmetros e o conteúdo são utilizados e adicionados à saída.

A Tabela 2 representa algumas tags e suas respectivas categorias.

Tabela 2 - Categorias das Tags HTML

Tag	Category	Tag	Category
a	3	html	2
abbr	2	i	2
acronym	2	img	3
b	2	input	3
base	3	ins	2
blockquote	3	kbd	2
body	2	label	2
br	2	li	2
button	3	map	3
caption	2	marquee	1
col	2	ol	2
comment	1	option	2
del	2	p	2
em	2	select	2
font	3	style	1
form	3	table	2

frame	2	textarea	2
h1–h6	2	title	1
head	2	tr	2
hr	2	var	2

Fonte: (CARPINTEIRO ; LIMA,; ASSIS, 2006).

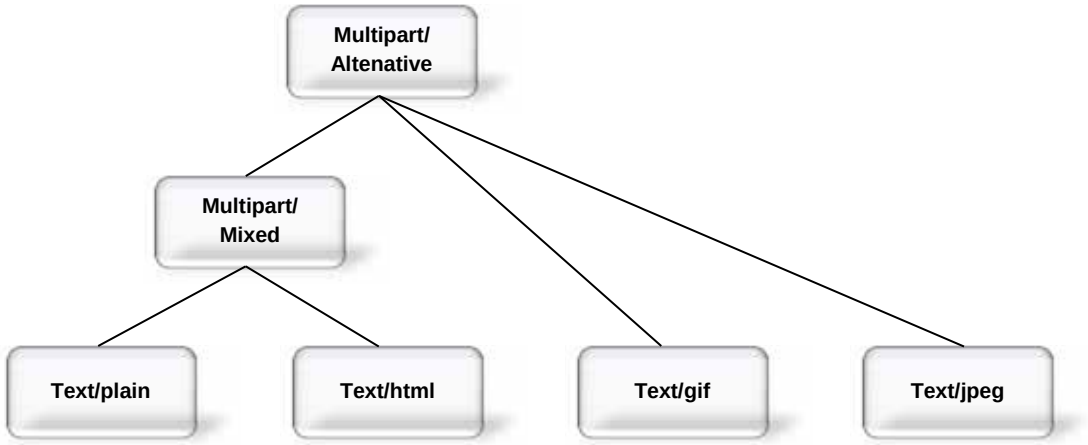
Processamento MIME: Um cabeçalho MIME é incluído antes de cada conteúdo MIME, apresentando o formato:

tipo:sub-tipo

O conteúdo MIME ainda pode se estender ao chamado MIME Multipart, que possibilita um MIME possuir outros tipos MIME. Assim é necessário um processo recursivo para processar o e-mail com estruturado de árvore quando no formato MIME Multipart.

A **Erro! Fonte de referência não encontrada.** representa a estrutura de um e-mail no or

Figura 13 - Exemplo de um e-mail com estrutura MIME Multipart



Fonte: Próprio autor

Os anexos não textuais são descartados, e então adicionados no corpo do e-mail a string "!_ANEXO_TIPO?nome_anexo".

Tokenização: E-mails no formato de texto original, ou e-mails no formato MIME e HTML depois de processados, são enviados a tokenização. O processo simplesmente separa o texto em tokens, ou seja, simples palavras. Para separação das palavras são utilizadas com caracteres delimitadores:

- ✓ Espaço;
- ✓ Nova linha;
- ✓ Tabulação;
- ✓ Exclamação;
- ✓ Interrogação;

- ✓ Vírgula;
- ✓ Ponto e vírgula.

Também na uniformização todos os caracteres são transformados em minúsculos e removida toda acentuação. Após o processo todos tokens são submetidos à detecção de padrões.

Detecção de Padrões: Tem com sua principal função identificar padrões de textos conhecidos e utilizados por spammers como técnicas para enganar filtros de spams, e unificar padrões de texto que desejamos ter uma saída única. Seguem padrões que deverão ser detectados como propostos em (ASSIS, 2006).

- ✓ Valores de saída de um parâmetro de uma tag HTML. Exemplo: "< table color=blue >", a saída seria "!_table_color";
- ✓ Todo endereço de e-mail encontrado no texto. Exemplo: "Mande e-mail para compreaqui@loja.com.br", a saída seria "Mande e-mail para !_email";
- ✓ URLs ou Hiperlinks, como no caso de e-mails terá como saída "!_link";
- ✓ Qualquer palavra que contenha caracteres inválidos no meio é substituída por "!_HIDEWORDS". Exemplos: ".M.O-N-E_Y" terá a saída "!_HIDEWORDS";
- ✓ Palavras muitos grandes sem sentido, acima de 20 caracteres será substituído por "!_BIGTEXT";
- ✓ Número ou strings inválidas colocadas após o sujeito será substituída pela string padrão "!_NUMERO_SUBJECT". Exemplo: "Get the best Life can offter you @e90jaakdfd". A saída nesse caso seria "Get the best Life can offter you !_NUMERO_SUBJECT";
- ✓ Qualquer ocorrência de quantidades monetárias e porcentagem são substituídas pelas strings "!_MONEY" e "!_PORCENTAGEM", respectivamente.

Seleção de Características: A maior dificuldade na classificação de textos é a alta dimensionalidade do espaço característico, já que cada palavra é considerada um espaço característico. Para contornar essa dificuldade é necessário o uso de métodos estatísticos a fim de selecionar as palavras mais relevantes para representar as classes spam ou ham.

Neste trabalho a seleção dessas palavras foi auxiliada pelo método Frequency Distribution (FD), que tem como objetivo medir o grau de ocorrência de um termo t em um conjunto C. O FD do termo t é calculado conforme equação 4.2.1:

$$FD(t) = \frac{n[t \in C]}{T} \tag{4.2.1}$$

sendo $n[t \in \{C\}]$ o número de ocorrências do termo t no conjunto C, e T o número total de termos no conjunto. Com o propósito de se reduzir o impacto das palavras com baixa incidência e/ou baixa significância, o cálculo para seleção de palavras considera apenas poucas palavras que carregam fortes indícios, e são descartadas as palavras que aparecem de maneira igual nas classes spam e ham. Dessa forma a seleção é dada pela equação 4.2.2:

$$FD_{s-h} = |FD_s - DF_h| \tag{4.2.2}$$

sendo

$$FD_s = \frac{n_s[t \in spam]}{T_s} \tag{4.2.3}$$

e

$$FD_h = \frac{n_h[t \in ham]}{T_h} \tag{4.2.4}$$

As palavras com maior valor de FD_{s-h} são selecionadas a fim de compor o chamado vetor característico, sendo que cada elemento do vetor representa uma entrada da RNA. O método Binary Weighting é utilizado para compor o vetor característico, assim se termo t_i ocorre pelo menos uma vez no e-mail, t_i recebe o valor 1, caso contrário, t_i recebe o valor 0. A Tabela 3 apresenta a composição de alguns vetores característicos:

Tabela 3 - Composição de vetores característicos

	t_0	t_1	t_2	t_3	t_4	t_5	t_6	...	t_M
E-mail 1	1	1	0	1	1	1	0	...	0
E-mail 2	1	1	1	1	1	0	1	...	1
E-mail 3	1	1	1	1	0	1	0	...	0
E-mail 4	1	1	0	0	1	0	0	...	0
E-mail 5	1	0	0	0	1	1	1	...	1

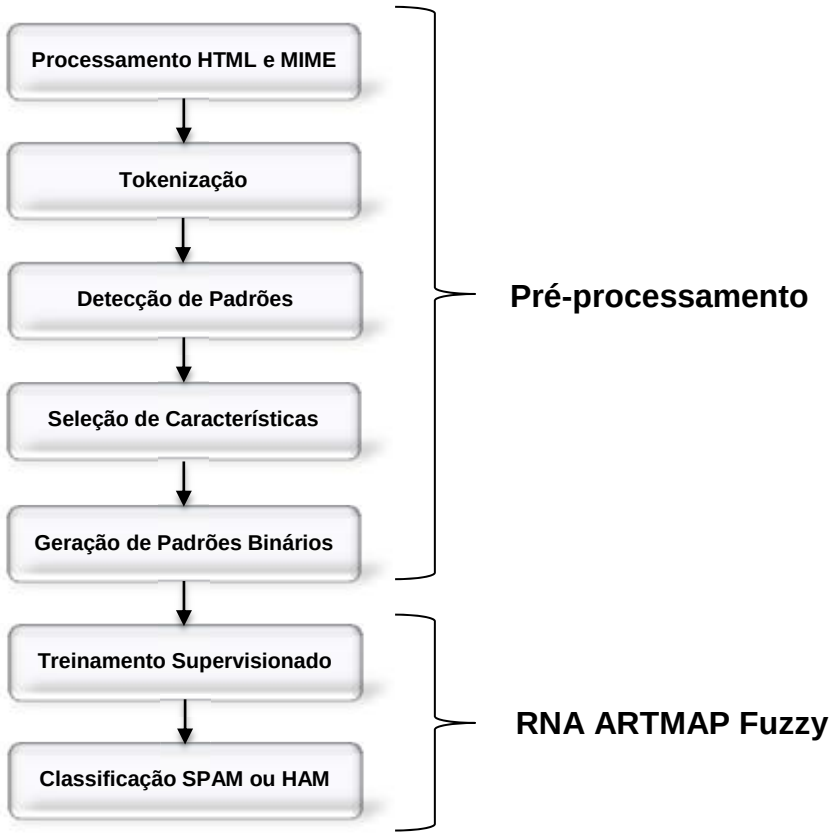
5.3 **RNA ARTMAP Fuzzy**

A Rede Neural Artificial ARTMAP Fuzzy, como já detalhada no item 3.8.2, possui treinamento supervisionado e uma grande capacidade de generalização, o que a torna uma opção bastante promissora dentro do problema de identificação e classificação de e-mails indesejados.

Os testes realizados neste trabalho foram com dados amplamente utilizados na literatura, portanto dados antigos, o que nos leva a um processo de treinamento e classificação off-line. No entanto, filtros de Spams em sistemas reais trabalham no modo on-line, exigindo assim que as técnicas de treinamento e classificação sejam também on-line. A arquitetura da RNA ARTMAP Fuzzy possibilita o emprego do chamado Treinamento Continuado (MORENO, 2010), o qual a possibilita estar em um treinamento permanente mesmo na etapa de classificação. Esse recurso reforça a ideia de que a RNA ARTMAP Fuzzy é uma opção promissora nas pesquisas de filtros de Spams, e uma das motivações para o uso dessa arquitetura de RNA neste trabalho.

A Figura 14 apresenta o diagrama de blocos da metodologia proposta.

Figura 14 - Diagrama de blocos da metodologia proposta



Fonte: Próprio autor

6 Resultados

6.1 Medida de Desempenho

Os métodos para avaliação de desempenhos para classificadores de e-mails apresentados na literatura ainda não deixam claro qual o melhor sistema de comparação, nem quais medidas oferecem uma avaliação justa, o que dificulta a comparação entre trabalhos. (CORMACK, 2008).

De acordo com Cormack (2008), os filtros devem ser julgados em quatro aspectos: autonomia, imediatismo, identificação de Spams e identificação de Hams. Entretanto, não é obvio um procedimento capaz de mensurar cada uma dessas dimensões separadamente, nem como combinar essas medidas em uma única expressão cujo propósito seja comparar o desempenho dos filtros de Spams.

Sejam S e H os conjuntos de mensagens de Spams e Hams, respectivamente, e todas as análises possíveis de predição apresentados na **Erro! Fonte de referência não encontrada.**.

Tabela 4- Possíveis resultados de predição de um classificador de e-mails

Notação	Denominação	Descrição
VP	Verdadeiros positivos	Spams classificados corretamente
VN	Verdadeiros negativos	Hams classificados corretamente
FN	Falso negativos	Spams classificados como hams
FP	Falso positivos	Hams classificados como spams

Fonte: Próprio autor.

Os e-mails classificados erroneamente como spam (falso positivo) pode ser altamente prejudicial ao usuário, já que um e-mail legítimo, muitas vezes de grande importância ao usuário, pode não chegar ao destinatário. Já os e-mails classificados erroneamente como ham (falso negativo), apesar de muitas vezes não causar tantos transtornos ao usuário quanto os falsos positivos, devem ser evitados ao máximo.

As medidas: taxa de erro, taxa de falso positivo e taxa de falso negativo, utilizadas para avaliar a eficiência do modelo proposto, são apresentadas na Tabela 5.

Tabela 5 - Medidas de desempenho para avaliar métodos de classificação

Medida	**Equação (%)**
Taxa de FP	$Tfp = \frac{\|FP\|}{H}$
Taxa de FN	$Tfn = \frac{\|FN\|}{S}$
Taxa de erro	$Ter = \frac{\|FP\| + \|FN\|}{\|S\| + \|H\|}$

Fonte: Próprio autor.

As medidas escolhidas possibilitam a análise da eficiência do modelo através da taxa de erro que indica o total de erros na classificação em relação a todo conjunto, e a taxa de falso positivo e falso negativo que indicam o total de erros na classificação de spams e hams separadamente.

6.2 **Base de Dados**

A base de dados utilizada para realização dos testes foi o SpamAssassin public mail corpus, criado especialmente para testes em sistemas de filtros anti-spams. Segue suas principais características (SPAMASSASSIN, 2011):

- ✓ Todos os cabeçalhos são reproduzidos na íntegra, com exceção de alguns que tiveram o hostname alterado para spamassassin.taint.org, apenas por questões de privacidade. Porém, na maioria dos casos, o cabeçalho permanece inalterado;
- ✓ Todas as mensagens do conjunto foram enviadas com o conhecimento que poderiam ser disponibilizadas como domínio público. Mensagens do tipo newsletter de servidores públicos também foram adicionadas ao conjunto;
- ✓ O SpamAssassin public mail corpus possui um total de 6.047 mensagens, formados por três subconjuntos: Spam - Contem 1.897 mensagens de spams, Easy Ham – 3.900 mensagens legítimos, facilmente diferenciáveis de spams por não conterem assinaturas de spams como por exemplo tags HTML, Hard Ham - 250 mensagens legítimas, porém mais difíceis de diferenciar dos spams por conterem tags HTML entre outras assinaturas. Portanto, o total de mensagens possui aproximadamente 31% de spams.

6.3 **Resultados Obtidos**

O método proposto foi submetido a testes a fim de se obter os melhores parâmetros e consequentemente os melhores resultados. A avaliação é feita através das medidas de desempenho Tfp (Taxa de falso positivo), Tfn (Taxa de falso negativo) e Ter (Taxa de erro).

Alguns parâmetros da RNA ARTMAP Fuzzy foram os mesmos em todos os testes, como apresenta a Tabela 6:

Tabela 6 - Parâmetros RNA ARTMAP Fuzzy

Parâmetros	**Valor**
$\boldsymbol{\rho_{ab}}$	1
$\boldsymbol{\rho_{b}}$	1
β	1
α	0.01

Fonte: Próprio autor.

Os valores dos parâmetros de vigilância ρ_{ab} e ρ_b foram mantidos com o valor 1 em todos os testes, já que o problema possui apenas duas opções de classes, Spam ou Ham, logo a vigilância deve exigir alto grau de semelhança. O parâmetro de escolha α apresentou melhores resultados para o problema com valores baixos, como o escolhido 0.01, assim como a taxa de treinamento β com o valor 1.

O parâmetro de vigilância ρ_a, a porcentagem de e-mails destinados ao treinamento e a dimensão do vetor de entrada (M) da RNA ARTMAP Fuzzy são cruciais para um bom desempenho do modelo proposto. O parâmetro ρ_a determina o grau de semelhança entre as classes existentes e os novos padrões apresentados à RNA ARTMAP Fuzzy; a porcentagem de e-mails utilizados no treinamento determina a quantidade de informações adquiridos pela RNA durante o processo, essencial na fase de classificação; e a dimensão do vetor M representa a quantidade de características extraídas no pré-processamento que é usada para distinguir Spams de e-mails legítimos.

A escolha de valores menores para M influenciaram na maior ocorrência dos chamados ruídos nos padrões destinados ao treinamento da RNA. Ruídos podem ser definidos como padrões iguais, mas de classes diferentes, ou seja, um padrão gerado a partir de um ham é exatamente igual ao padrão gerado a partir de um Spam. Na fase de aprendizagem os padrões identificados como ruídos foram excluídos, pois é incoerente apresentar a RNA ARTMA Fuzzy dois padrões iguais indicando saídas distintas, principalmente por se tratar de uma RNA com treinamento supervisionado. A escolha de valores maiores para M indica um maior nú-

mero de características para formar um padrão, o que possibilita maior diversificação dos padrões e diminuição na ocorrência de ruídos. Porém essas características podem não possuir um bom valor de DF_{s-h} (método utilizado na seleção de características) e contribuir para o aumento da taxa de erros (Ter).

A seguir são apresentados os resultados para M = 20, 50, 100 e 200, e treinamento com 50% a 90% da base de dados. A coluna "Ruídos%" representa a porcentagem de padrões excluídos em relação a base de dados:

Tabela 7 - Resultados para M = 20

Treino %	Ruídos%	ρ_a	Tfn%	Tfp%	Ter%
50	7.640	0.75	8.496	1.591	3.367
60	**7.425**	0.60	**10.363**	1.566	**3.841**
70	8.020	0.75	7.765	**2.492**	3.835
80	**13.775**	0.99	0.935	0.483	0.575
90	**13.775**	0.99	**0.000**	**0.000**	**0.000**

Fonte: Próprio autor

A tabela 7 apresenta os resultados para $M = 20$ e treinamento de 50% a 90%. A taxa de falso negativo, taxa de falso positivo e taxa de erro chegaram a 0.0% quando a RNA é submetido a treinamento de 90%, sendo a taxa de ruído de 13.775%. A menor taxa de ruído encontrada foi de 7.425% quando a RNA é submetida a treinamento de 60%, porém a taxa de falso negativo é de 10.363%, taxa de falso positivo de 1.566% e taxa de erro de 3.841%.

Tabela 8 - Resultados para M = 50

Treino %	Ruídos%	ρ_a	Tfn%	Tfp%	Ter%
50	**2.215**	0.94	12.585	5.063	7.307
60	**2.215**	0.94	**13.050**	3.795	6.554
70	**1.951**	0.74	**10.487**	**7.390**	**8.319**
80	**1.951**	0.82	10.674	5.422	6.998
90	1.967	0.72	12.360	**1.446**	**4.722**

Fonte: Próprio autor

A tabela 8 apresenta os resultados para $M = 50$ e treinamento de 50% a 90%. A menor taxa de erros encontrada é de 4.722% quando a RNA é submetido a treinamento de 90%, sendo a taxa de falsos negativos de 12.360%, taxa de falsos positivos de 1.446% e taxa de ruído de 1.967%. A menor taxa de ruído encontrada foi de 1.951% quando a submetido a trei-

namento de 70% e 80%, porém a taxa de falso negativo chega a 10.674%, taxa de falso positivo a 7.390% e taxa de erro a 8.319%.

Tabela 9 - Resultados para M = 100

Treino %	**Ruídos%**	ρ_a	**Tfn%**	**Tfp%**	**Ter%**
50	0.479	0.82	11.765	4.822	6.979
60	0.463	0.97	**15.909**	5.361	**8.638**
70	**0.446**	0.67	**9.447**	7.149	7.863
80	0.496	0.67	9.920	**7.590**	8.313
90	**0.512**	0.98	10.638	**2.657**	**5.150**

Fonte: Próprio autor

A tabela 9 apresenta os resultados para $M = 100$ e treinamento de 50% a 90%. A menor taxa de erros encontrada é de 5.150% quando a RNA é submetido a treinamento de 90%, sendo a taxa de falsos negativos de 10.638%, taxa de falsos positivos de 2.657% e taxa de ruído de 0.512%. A menor taxa de ruído encontrada foi de 0.446% quando a submetido a treinamento de 70%, porém a taxa de falso negativo chega é de 9.447%, taxa de falso positivo de 7.149% e taxa de erro a 7.863%.

Para M = 200 não houve caso de ruídos, seguem os resultados na Tabela 10:

Tabela 10 - Resultados para M = 200

Treino %	ρ_a	**Tfn%**	**Tfp%**	**Ter%**
50	0.90	**11.052**	5.111	6.965
60	0.92	15.160	4.099	7.549
70	0.99	**16.014**	**5.145**	**8.527**
80	0.98	14.209	3.865	7.077
90	0.98	14.130	**1.208**	**5.184**

Fonte: Próprio autor

A tabela 10 apresenta os resultados para $M = 200$ e treinamento de 50% a 90%. A menor taxa de erros encontrada é de 5.184% quando a RNA é submetido a treinamento de 90%, sendo a taxa de falsos negativos de 14.130% e taxa de falsos positivos de 1.208%.

7 Conclusão e Sugestões para Trabalhos Futuros

7.1 Conclusões

O uso de RNAs com arquitetura baseada na teoria da ressonância adaptativa para classificação de padrões é destaque na literatura com resultados bastante satisfatórios. Este trabalho teve como objetivo o desenvolvimento de uma metodologia para detecção de e-mails indesejados, sendo utilizado o modelo de pré-processamento proposto em Carpinteiro, Lima e Assis (2006) e a RNA ARTMAP Fuzzy para classificação.

No modelo proposto, o modelo de pré-processamento foi responsável por extrair as características mais relevantes dos e-mails pertencentes às classes Spam e Ham, a fim de se gerar o vetor característico. O padrão binário referente a cada e-mail foi formado a partir do vetor característico e a RNA ARTMAP Fuzzy responsável por classificar os padrões de entrada entre Spam ou Ham.

Foram realizados testes com diferentes dimensões do vetor de entrada M do módulo ART_a da RNA, 20, 50, 100 e 200; e diferentes porcentagens de treinamento em relação a base de dados, de 50% a 90%. O modelo apresentou ótimos resultados principalmente em relação a taxa de falsos positivos, que sempre se manteve abaixo da taxa de falsos negativos, além de pouca variação dos resultados quando a RNA é submetida a menos amostras na fase de aprendizagem. Ainda é possível afirmar que quanto maior o índice M, se têm características de menor qualidade, já que o valor de FD_{s-h} é menor, mas, no entanto, contribui para a não ocorrência de ruídos. Logo se pode concluir que o vetor M de tamanho menor e a não ocorrência de ruídos contribui para a melhora dos resultados.

7.2 Sugestões para Trabalhos Futuros

Um dos principais objetivos deste trabalho é contribuir para o surgimento de novos modelos que possuam a capacidade de adquirir informações de treinamento e classificar e-mails de um modo dinâmico e adaptativo. Segue sugestões para trabalhos futuros:

- ✓ Utilizar métodos para a extração de características com abordagem de contexto, e não só ocorrência de palavras;
- ✓ Acrescentar a extração de características de imagens e anexos através de métodos de processamento de imagens;

- ✓ Modificar o treinamento da RNA ARTMAP Fuzzy para o chamado treinamento continuado e dessa forma obter um filtro capaz de se adaptar a novos modelos de Spams;
- ✓ Criar classes intermediárias para identificar e-mails com características muito próximas de Spams e Hams;

Referências

A BRIEF history of email. Learning Center. Disponível em: <http://www.vicomsoft.com/learning-center/history-of-email/>. Acesso em: 15 mar 2012.

ACKLEY, D. H.; HINTON, G. E.; SEJNOWSKY, T. J. A Learning algorithm for Boltzmann machines. **Cognitive Sciences,** v. 9, p. 147-169, 1985. Disponível em: < http://onlinelibrary.wiley.com/doi/10.1207/s15516709cog0901_7/pdf>. Acesso em: 12 abr. 2012.

ALMEIDA, T. A. **SPAM:** do surgimento à extinção. 2010. 114 f. Tese (Doutorado em Engenharia Elétrica) – Faculdade de Engenharia Elétrica e Computação, Universidade Estadual de Campinas, Campinas, 2010.

AMORIM, D. G. **Redes art com categorias internas de geometria irregular.** 2006. 246 f. Tese (Doutorado em Física) – Universidade de Santiago de Compostela, Santiago de Compostela, 2006.

ASSIS, J. M. **Detecção de e-mails spam utilizando redes neurais artificiais.** 2006. 112 f. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Itajubá, Itajubá, 2006.

BARRACUDA Networks. **Spam Data.** Disponível em: <http://www.barracudacentral.org/data/spam> Acesso em: 21 jan. 2013.

BELLIS, M. **History of e-mail & ray tomlinson**. About.com Inventors. Disponível em: <http://inventors.about.com/od/estartinventions/a/email.htm>. Acesso em: 12 mar.2012.

BRAIN, M. **Como funciona o e-mail.** Disponível em: <http://informatica.hsw.uol.com.br/e-mail.htm>. Acesso em: 12 mar. 2012.

CAPUANO, E. A. O poder cognitivo das redes neurais artificiais modelo Art1 na recuperação de informação. **Ciência da Informação**, Brasiília, DF, v. 38, p. 9-30, 2009.

CARPENTER, G. A.; GROSSBERG, S. ART2: Self-organization of stable category recognition codes for analog input patterns. **Applied Optics,** Washington, v. 26, p. 4919-4930, 1987.

CARPENTER, G. A.; GROSSBERG, S.; MARKUZON, N.; REYNOLDS, J. H.; D. ROSEN. Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. **IEEE Transactions on Neural Networks**, New York, v. 3, n. 5, p. 698–713, 1992.

CARPENTER, G. A.; GROSSBERG, S.; REYNOLDS, J. H. ARTMAP: supervised real-time learning and classification of nonstationary data by a self-organizing neural network. **Neural Networks**, New York, v. 4, p. 565-588, 1991.

CARPENTER, G. A.; GROSSBERG, S.; ROSEN, D. B. Fuzzy ART: fast stable learning and categorization of analog patterns by an adaptive resonance system. **Neural Networks**, New York, v. 4, p. 759-771, 1991.

CARPINTEIRO, O. A.; LIMA, I.; ASSIS, J. M. A Neural model in anti-spam systems. In: Proceedings of the 16th international conference on Artificial Neural Networks, 16, 2006, Heidelberg. **Proceedings of the…** Heidelberg: Springer-Verlag Berlin, 2006. p. 847-855. Disponível em: < http://link.springer.com/chapter/10.1007%2F11840930_88#page-1>. Acesso em: 15 dez. 2012.

CORMACK, G. Email spam filtering: a systematic review. **Foundations and Trends in Information Retrieval**, Hanover, v. 1, n. 4, p. 335-455, 2008.

ELEVEN Research. Eleven E-Mail security report. December 2012. Disponível em: <http://www.eleven.de/eleven-security-reports-reader.612/items/eleven-e-mail-security-report-december-2012.html>. Acesso em: 20 dez. 2012.

FREEMAN, J. A.; SKAPURA, D. M. **Neural Networks:** algorithms, applicationsand programming techniques. New York: Addison-Wesley, 1991. 401 p.

GROSSBERG, S. Adaptive pattern classification and universal recoding II. Feedback, expectation, olfaction, illusions. **Biological Cybernetics,** Heidelberg, v. 23, p. 187-202, 1976.

GROSSBERG, S. Competitive learning: from interactive activation to adaptive resonance. **Cognitive Science,** Wheat Ridge, v. 11, p. 23-63, 1987.

GUDKOVA, D. **Kaspersky security bulletin:** Spam Evolution 2012. Disponível em: <http://www.securelist.com/en/analysis/204792276/Kaspersky_Security_Bulletin_Spam_Evolution_2012> Acesso em: 18 jan. 2013.

HAYKIN, S. **Neural networks:** a comprehensive foundation. 2. ed. New Jersey: Prentice-Hall, 1999. 842 p.

HUNT, C.; LOUKIDES, M. **TCP/IP network administration.** 2. ed. Cambridge: O`Reilly, 1997. 630 p.

KOHONEN, T. Self-organized formation of topologically correct of feature maps. **Biological Cybernetics,** Heidelberg, v. 43, p. 159-168, 1982. Disponível em: < http://cns-classes.bu.edu/cn510/Papers/kohonen-82.pdf>. Acesso em: 12 dez. 2012.

KROSE, B.; SMAGT, P. V. **An introduction to neural networks.** Amsterdam: University of Amsterdam, 1996. 135 p.

LEVINE J. **Why is spam bad?.** Disponível em: <http://spam.abuse.net/overview/spambad.shtml>. Acesso em: 12 jan. 2013.

LOPES, M. L. **Desenvolvimento de um sistema previsor de cargas elétricas via redes neurais.** 2000. 100 f. Dissertação (Mestrado em Engenharia Elétrica) – Faculdade de Engenharia, Universidade Estadual Paulista, Ilha Solteira, 2000.

LOPES, M. L. **Desenvolvimento de redes neurais para previsão de cargas elétricas de sistemas de energia elétrica.** 2005. 169 f. Tese (Doutorado em Engenharia Elétrica) – Faculdade de Engenharia, Universidade Estadual Paulista, Ilha Solteira, 2005.

MA, Q.; QIN, Z.; ZHANG, F.; LIU, Q. Text spam neural network classification algorithm. communications, circuits and systems. In: IEEE INTERNATIONAL CONFERENCE IN COMMUNICATIONS CIRCUITS AND SYSTEMS (ICCCAS), 2010, Chengdu. **Proceedings of the...** Chengdu: IEEE, 2010. p. 466-469. Disponível em: < http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5581954>. Acesso em: 12 mar. 2013.

MALANGE, F. C. **Rede neuro-fuzzy-wavelet para detecção e classificação de anomalias de tensão em sistemas elétricos de potência.** 2010. 125 f. Tese (Doutorado em Engenharia Elétrica) – Faculdade de Engenharia, Universidade Estadual Paulista, Ilha Solteira, 2010.

MANJUSHA, K.; KUMAR, R. Spam mail classification using combined approach of bayesian and neural. In: IEEE INTERNATIONAL CONFERENCE COMPUTATIONAL INTELLIGENCE AND COMMUNICATION NETWORKS (CICN), 2010, Bhopal. International Conference. **Proceedings of the...** Bhopal: IEEE Computer Society, 2010, p. 145-149. Disponível em: < http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5701953>. Acesso em: 12 abr. 2013.

MARCHIORI, S. C. **Desenvolvimento de um sistema para análise da estabilidade transitória de sistemas de energia elétrica via redes neurais.** 2006. 130 f. Dissertação (Mestrado em Engenharia Elétrica) – Faculdade de Engenharia, Universidade Estadual Paulista, Ilha Solteira, 2006.

MARDWARE. **A origem do SPAM.** Disponível em: <http://www.mardware.com/2010/07/origem-do-spam.html>. Acesso em: Dez. de 2013.

MCCULLOCH, W. S.; PITTS, W. H. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, New York, v. 5, p. 115-133, 1943.

MENDES, E. F.; CARVALHO, A. C. **Tutorial introdutório sobre redes neurais artificiais.** São Carlos: USP/EESC/Departamento de Ciências de Computação e Estatística, 1997.

MINUSSI, C. R.; LOTUFO, A. D. **Redes neurais:** introdução e principais conceitos. Ilha Solteira: UNESP/FE/DEE, 2008. (Apostila).

MORENO, A. L. **Análise da estabilidade transitória via rede neural ART-ARTMAP fuzzy euclidiana modificada com treinamento continuado.** 2010. 104 f. Tese (Doutorado em Engenharia Elétrica) – Faculdade de Engenharia, Universidade Estadual Paulista, Ilha Solteira, 2010.

ROSEMBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological Review**, Washington, v. 65, n. 5, p. 386-408, 1958. Disponível em: < http://psycnet.apa.org/journals/rev/65/6/386.pdf>. Acesso em: 15 abr. 2013.

RUAN, G.; TAN, Y. A Three-layer back-propagation neural network for spam detection using artificial immune concentration. **Soft Computing**, Heidelberg, v. 14, p. 139-150, 2009. Disponível em: < http://link.springer.com/article/10.1007%2Fs00500-009-0440-2#page-1>. Acesso em: 12 abr. 2013.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by Back-propagation errors. **Nature,** London, v. 329, p. 533-536, 1986. Disponível em: < http://www.iro.umontreal.ca/~vincentp/ift3395/lectures/backprop_old.pdf>. Acesso em: 12 jan. 2013.

SILVA, A. M. **Utilização de redes neurais artificiais para classificação de spams.** 2009. 126 f. Dissertação (Mestrado em Modelagem Matemática e Computacional) – Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, 2009.

SILVA, T. A. A. **Previsão de cargas elétricas através de um modelo híbrido de regressão com redes neurais**. 2012. 63 f. Dissertação (Mestado em Engeharia Elétrica) – Universidade Estadual Paulista, Ilha Solteira, 2012.

SILVEIRA, M. G. **Análise de estabilidade transitória de sistemas elétricos por redes neurais ARTMAP nebulosas modulares.** 2003. 107 f. Tese (Doutorado em Engeharia Alétrica) – Faculdade de Engenharia, Universidade Estadual Paulista, Ilha Solteira, 2003.

SPAMASSASSIN. **The apache spamassassin project.** Disponível em: <http://spamassassin.apache.org/publiccorpus/> Acesso em: 12 set. 2011.

TEAM, R. **Email statistics report, 2012-2016.** Disponível em: <http://www.radicati.com/?p=8262>. Acesso em : 17 jan. 2013.

TEMPLETON'S B. **Origin of the term "spam" to mean net abuse.** Disponível em: < http://www.templetons.com/brad/spamterm.html>. Acesso em: 17 jan. 2013.

UPASANA, S. C.; CHAKRAVARTY, S. A Survey of Text Classification Techniques for E-mail Filtering. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND COMPUTING, 2, 2010 Washington. **Proceedings of the…**. Washington: IEEE, 2010. p. 32-36, 2010. Disponível em: < http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=05460695>. Acesso em 123 jan. 2013.

VAN, T. V. **The history of electronic mail:** multicians. Disponível em: <http://www.multicians.org/thvv/mail-history.html>. Acesso em: 12 mar. 2012.

WEKA. **Data mining software in Java.** Disponível em: <http//www.cs.waikato.ac.nz/ml/weak> Acesso em: 12 jul. 2012.

WERBOS, P. J. **Beyond regression:** new tools for prediction and analysis in the behavioral sciences. 1974. 102 f. Dissertation (Master) – Harvard University, Massachusetts, 1974.

WIDROW, B.; LEHR, M. A. 30 years of adaptive neural networks: perceptron, madaline, and backpropagation. **Proceedings of the IEEE**, New York, v. 78, n. 9, p. 1415-1442, 1990. Disponível em: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=58323>. Acesso em: 12 mar. 2013.

Apêndice A - Redes Neurais Artificiais

A.1 Reflexão Histórica

As redes neurais artificiais (RNAs) são modelos computacionais inspirados no sistema nervoso dos seres vivos e tem como principal característica simular o seu funcionamento (HAYKIN, 1999).

As características mais importantes envolvidas com as aplicações das RNAs são a adaptação por experiência, capacidade de aprendizado, habilidade de generalização, organização de dados, tolerância a falhas e o armazenamento distribuído.

Embora os primeiros trabalhos em redes neurais artificiais tenham sido publicados há mais de 50 anos, tal tema começou a ser fortemente pesquisado a partir de 1990, sendo até hoje um potencial de pesquisa atrativo, pois as RNAs são capazes de resolver problemas complexos como previsão de séries temporais, aproximações de funções e identificação de padrões, sem a necessidade da modelagem matemática do problema a ser resolvido.

A era moderna das redes neurais começou em 1943, com o trabalho pioneiro de (MCCULLOCH ; PITTS, 1943), que descreveu o primeiro neurônio artificial. Em 1958 ROSEMBLATT, F. (1958), apresentou seu trabalho sobre o perceptron, uma das primeiras RNAs, com apenas um neurônio e aprendizagem supervisionada. Em 1960 (WIDROW; LEHR, 1990) introduziram o algoritmo do mínimo quadrado (LMS - Least Mean-Square) e o usaram para formular o Adaline (Adaptive linear element). A diferença entre o perceptron e o Adaline está no procedimento de aprendizagem. Após dois anos Widrow propôs uma das primeiras RNAs de estrutura múltipla (Madaline, Multiadaline). Nos anos 70 surgiram vários trabalhos a respeito dos mapas auto-organizáveis utilizando aprendizagem competitiva. Em 1976, surgem os primeiros trabalhos de (GROSSBERG, 1976) a respeito da teoria da ressonância adaptativa (ART - Adaptative Ressonance Theory). Nos anos 80 as redes neurais com realimentação atraíram muita atenção, e tornaram-se conhecidas como redes de Hopfield; e os mapas auto-organizáveis de (KOHONEN, 1982), utilizando uma estrutura de rede unidimensional ou bidimensional. Em 1985 surgiu a máquina de Boltzmann, sendo esta desenvolvida por (ACKLEY et al. ,1985), que foi a primeira realização bem sucedida de uma rede neural de múltiplas camadas. Em 1986 (RUMELHART et. al. 1986) apresentam o algoritmo retropropagação (backpropagation), primeiramente proposto por (WERBOS, 1974). O algoritmo de treinamento por retropropagação (backpropagation) é um dos mais utilizados até hoje. Gros-

Sberg, (1987) apresentou sua primeira rede baseada na teoria da ressonância adaptativa em 1987, tal rede é conhecida como ART1 (SILVA, 2012).

Várias outras redes foram desenvolvidas nos últimos 20 anos, e o uso de técnicas híbridas vem sendo cada vez mais estudadas.

A.2 Modelo do Neurônio Biológico

Um neurônio artificial é baseado no funcionamento do neurônio biológico e uma rede neural se compõe por uma quantidade pré definida de neurônios artificiais que possibilita a solução de problemas de um número considerável de áreas distintas.

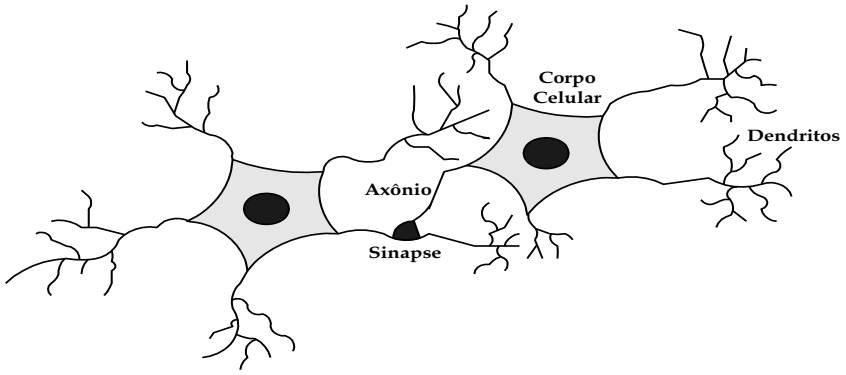
O cérebro humano pode ser considerado um processador complexo formado por bilhões de neurônios responsáveis pelo controle motor, da percepção, bem como do reconhecimento de padrões. A ligação desses neurônios se dá através de sinapses, as quais transmitem estímulos e que formam uma grande rede que estende o resultado pelo corpo humano (SILVA, 2012).

O neurônio biológico é composto por:

- ✓ Dentritos: têm a função de receber estímulos de outros neurônios;
- ✓ Corpo celular (soma): coleta e combina informações recebidas;
- ✓ Axônio: composto por uma fibra tubular, que transmite estímulos a outras células;
- ✓ Sinapses: conecta o axônio aos dentritos de outros neurônios, propagando pulsos nervosos e excitando ou inibindo tais pulsos.

Na Figura 15 é mostrada a ilustração de um neurônio biológico.

Figura 15 - Neurônio biológico

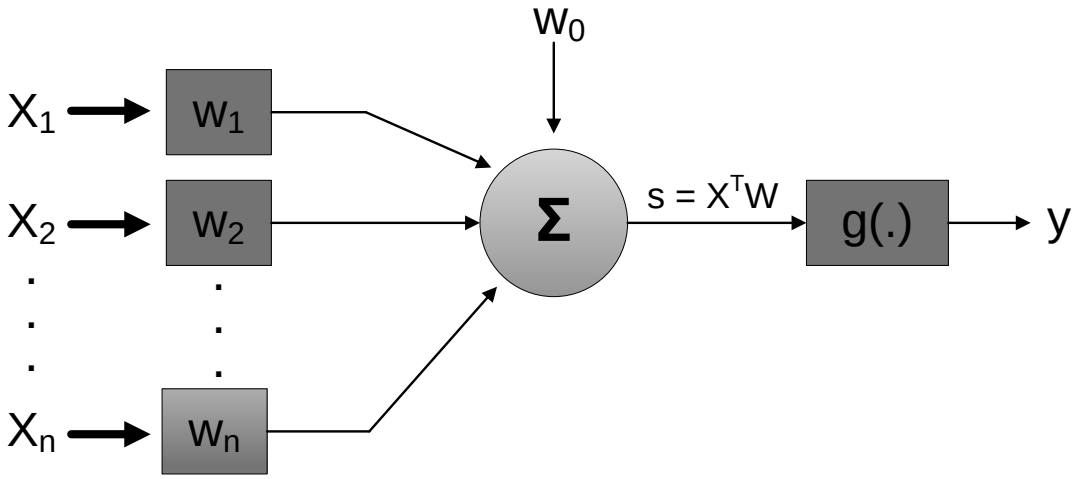


Fonte: (MENDES; CARVALHO, 1997).

A.3 Modelo do Neurônio Artificial

A RNA é um processador paralelo composto por neurônios que possui a capacidade de armazenar conhecimento. Semelhante ao cérebro humano, a rede neural possui um neurônio artificial, que busca imitar as características do neurônio biológico. A Figura 16 ilustra o modelo de um neurônio artificial proposto por (MCCULLOCH ; PITTS, 1943) com n entradas.

Figura 16 - Modelo MCP (McCulloch-Pitts) não-linear de um neurônio artificial



Fonte: (SILVA, T. A. A., 2012)

O processo do treinamento de um neurônio artificial se desenvolve da seguinte forma (LOPES, 2005):

- ✓ Um conjunto de entradas X, $x_1, x_2, \dots, x_n$, são aplicadas na camada de entrada do neurônio, onde elas representam os sinais dentro de um neurônio biológico;
- ✓ Cada sinal é ponderado por um peso associado W, $w_1, w_2, \dots, w_n$, que indica sua influência na saída da unidade;
- ✓ As entradas ponderadas são aplicadas ao bloco somatório, onde se adiciona também uma entrada independente, $x_0 = \pm 1$, denominada bias;
- ✓ Se a soma ponderada das entradas for maior que o valor de w_0 (limiar) o neurônio é ativado provocando um pulso na saída, como mostra a Equação 49. Caso contrário o neurônio não é ativado (WIDROW, B.; LEHR, M. A., 1990).

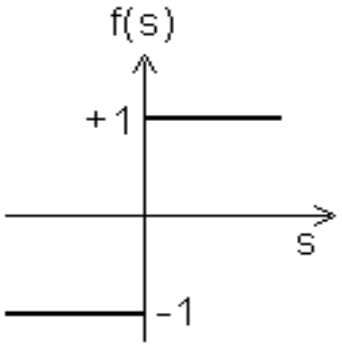
A saída intermediária do neurônio é, então, dada por:

$$s_i = \sum_j^n w_{ij} x_j \quad (49)$$

A ativação do neurônio artificial é feita através de uma função de ativação. As funções de ativação mais conhecidas e utilizadas são mostradas nas Figuras 17, 18, 19 e 20.

✓ Função degrau bipolar → $f_d(s) = \begin{cases} +1, & se\ s \geq 0 \\ -1, & se\ s \leq 0 \end{cases}$

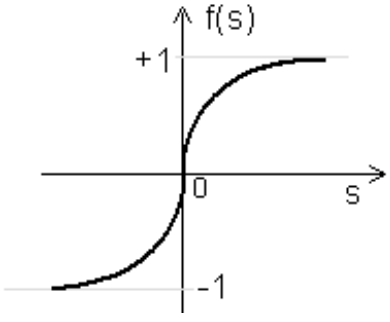
Figura 17 - Função degrau bipolar



Fonte: (MINUSSI, C. R.; LOTUFO, A. D., 2008).

✓ Função tangente hiperbólica → $f_t(s) = \frac{1-e^{-\lambda s}}{1+e^{-\lambda s}}$

Figura 18 - Função tangente hiperbólica

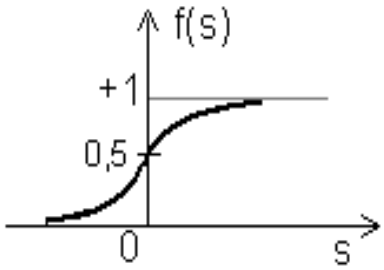


Fonte: Fonte: (MINUSSI, C. R.; LOTUFO, A. D., 2008).

sendo λ a inclinação da curva.

✓ Função Logística → $f_l(s) = \frac{1}{1+e^{-\lambda s}}$

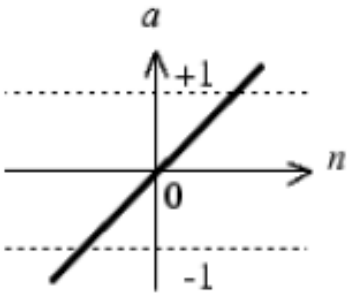
Figura 19 - Função logística



Fonte: Fonte: (MINUSSI, C. R.; LOTUFO, A. D., 2008).

- ✓ Função Linear → $f_{li}(n) = n$ para todo n

Figura 20 - Função linear



Fonte: (MINUSSI, C. R.; LOTUFO, A. D., 2008).

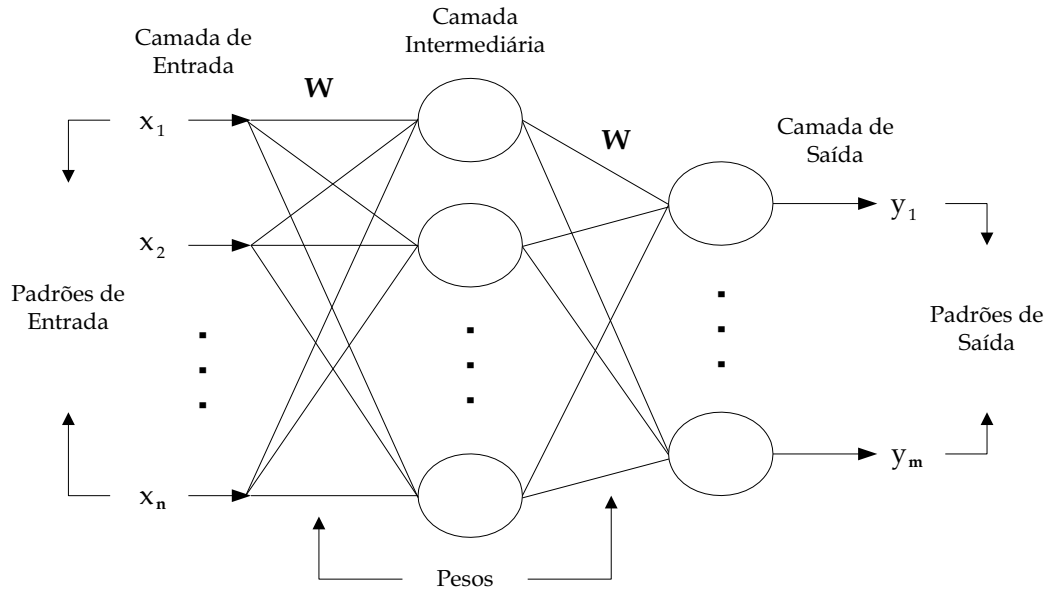
As funções tipo relé são apropriadas para sistemas binários, enquanto que as funções sigmoidais podem ser empregadas tanto para sistemas contínuos como binários (LOPES, 2000).

A.4 Estrutura das Redes Neurais Artificiais

As RNAs são definidas como conjuntos estruturados de unidades de processamento denominadas neurônios, interligadas entre si que forma uma disposição estrutural de camadas e conexões entre as camadas.

Basicamente as redes neurais são formadas por três camadas: camada de entrada, camada intermediaria (oculta, escondida ou invisível) e camada de saída (HAYKIN, 1999). como ilustra a Figura 21.

Figura 21 - Disposição das redes neurais artificiais



Fonte: (LOPES, 2005).

- ✓ Camada de entrada: é responsável por receber informações dos dados;
- ✓ Camada intermediaria: é responsável por extrair as características associadas ao processo ou sistema a ser inferido. Quase todo o processamento interno da rede é realizado nessa camada. O número de camadas escondidas e a quantidade de neurônios nessa camada dependem principalmente da complexidade do problema a ser mapeado pela rede;
- ✓ Camada de saída: é também constituída de neurônios, sendo esta responsável pela apresentação dos resultados obtidos pela rede.

As principais arquiteturas das RNAs em relação à propagação dos dados podem ser classificadas em: redes feedforward, redes recorrentes e redes reticuladas.

- ✓ Redes feedforward: também são conhecidas como redes não recorrentes, os dados fluem das unidades de entrada para as unidades de saída, ou seja, a saída é exclusivamente determinada em função das entradas e dos valores dos pesos. O processamento dos dados pode se estender sobre múltiplas unidades (camadas). As principais redes cuja arquitetura é do tipo feedforward são o Perceptron multicamadas (Multilayer Perceptron – MLP) e as redes de base radial (radial basis function - RBF) (KROSE ; SMAGT, 1996);
- ✓ Redes recorrentes: contém conexões feedback, e desenvolvem uma memória a longo prazo nos neurônios internos. Devido à realimentação, as redes com essa

arquitetura produzem saídas levando-se em consideração os valores das saídas anteriores recentes. A principal rede que possui realimentação é denominada rede de Hopfiled (KROSE; SMAGT, 1996);

- ✓ Redes Reticuladas: Tem como principal característica considerar a disposição espacial dos neurônios visando à extração de características, isto é, a localização espacial dos neurônios está relacionada com o processo de ajuste de seus pesos e limiares. A principal rede desse tipo de arquitetura é o mapa auto-organizável de Kohonen (MINUSSI ; LOTUFO, 2008).

A.5 Treinamento da Rede Neural

A propriedade mais importante das redes neurais é a habilidade de aprender e com isso melhorar seu desempenho. A rede é treinada de maneira que um conjunto de entrada produza o conjunto de saída desejada ou que pelo menos seja consistente (MENDES ; CARVALHO, 1997).

A aprendizagem das RNAs ocorre na fase de treinamento. Os tipos de treinamento mais utilizados são os supervisionados e não supervisionado (MINUSSI; LOTUFO, 2008).

- ✓ Treinamento supervisionado: consiste em um método de aprendizagem em que as combinações dos padrões de entrada e saída são arbitrados por um tutor ou professor na fase de aprendizado, como é o caso da rede Perceptron Multicamadas;
- ✓ Treinamento não supervisionado: não possui tutor, ou seja, a rede é auto-organizada, sendo capaz de descobrir estaticamente, padrões relevantes aos dados de entrada, o que ocorre na rede ART;
- ✓ Treinamento com reforço: o conjunto de treinamento é formado somente por entradas, no entanto, há um elemento externo que, em vez de retornar o erro de saída da rede, retorna um sinal de satisfatório ou não em relação à última ação da rede. Caso a ação não seja satisfatória, ela será reforçada, tendo menor chance de ocorrer futuramente. Caso contrário, se a ação resultou em uma melhora no desempenho, ou seja, foi satisfatória, ela será reforçada aumentando a probabilidade de ocorrência da mesma no futuro, servindo de exemplo a rede de Kohonen.