Rede Neural ARTMAP fuzzy para detecção e classificação de e-mails indesejados

C. R. Santos Junior, A. D. P. Lotufo, M. C. G. Silveira
FEIS – Faculdade de Engenharia de Ilha Solteira
Universidade Estadual Paulista – UNESP
Ilha Solteira-SP, Brasil
carlos9_rsj@hotmail.com, annadiva@ieee.org, carmo@adm.feis.unesp.br

***Abstract*—O problema de mensagens não solicitadas pelos usuários em meios de comunicação eletrônica, embora tenha 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 principal característica ser dinâmico, ou seja, 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 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 sua eficiência.**

Keywords—E-mail, Spam, Filtro de spams, Redes Neurais Artificiais, ARTMAP Fuzzy.

I. 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 inúmeros destinatários com facilidade e sem aumento de custos. Estas vantagens também foram observadas pelos Spammers que através do envio de e-mails não solicitados levam aos destinatários, quase sempre de forma incômoda e inconveniente, conteúdos publicitários ou códigos maliciosos. 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 spams em relação aos e-mails trafegados no mundo [1]. Esses dados representam grande insatisfação dos usuários, já que os spams causam inúmeros problemas, como inundação nas caixas de e-mails consumindo tempo, dinheiro, largura de banda, além de fraudes, roubos, etc.

Os filtros anti-spam são utilizados para identificar e bloquear o maior número possível de spams de chegarem aos usuários. Na literatura encontram-se várias técnicas para filtragem, entre elas destacam-se o uso das Redes Neurais Artificiais (RNA), Sistema Imunológico Artificial, SVM (Support Vector Machines), filtros Bayesianos, Lógica Fuzzy, entre outros [2]. No trabalho de [3], o princípio da descrição mais simples auxiliado por fator de confiança (MDL-CF) com treinamento por erro (Train On Error), formaram o modelo de filtro anti-spam. As bases de dados TREC05, TREC06 e CEAS08 foram utilizadas nos testes. Em [4] a base de regras disponibilizada em [5] é utilizada para criar uma tabela de palavras válidas para pré-processar os e-mails e alimentar a RNA Perceptron Multicamadas com o algoritmo de treinamento Backpropagation. A base de dados utilizada é a SpamAssasin. No trabalho de [6] é utilizado no pré-processamento o software WEKA [7] e redes Bayesianas combinadas com RNA com treinamento baseado em algoritmo genético para classificação. Uma base de dados formada pelo próprio autor foi utilizada para os testes. Já nos trabalhos de [8] e [9] é utilizado o mesmo modelo de pré-processamento proposto por [8]. As RNAs SOM-WTA e SOM-LVQ foram as escolhidas por [9] e a RNA Perceptron Multicamadas em ambos os trabalhos. Além da base de dados SpamAssasin utilizadas nos trabalhos, [9] também inclui amostras de e-mails originadas no Brasil.

Uma RNA 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 [10].

Neste trabalho é proposto um modelo de filtro de spams utilizando a extração de características proposto em [8] e a RNA ARTMAP Fuzzy [11] para detecção e classificação de Spams.

A extração e seleção de características são realizadas na fase de pré-processamento dos e-mails, e tem como objetivo simplificar a tarefa de classificação da RNA, já que os tornam mais simples, uniformes e sem informações desnecessárias.

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

II. PRÉ-PROCESSAMENTO

O desempenho na tarefa de classificação das RNAs esta relacionado diretamente à qualidade dos dados utilizados na fase de treinamento. O pré-processamento tem a função de garantir a extração e seleção minuciosa das características da base de dados, de forma que os dados sem informações irrelevantes, mais simples e uniformes possibilita a RNA aperfeiçoar o processo de classificação. O pré-processamento pode ser divido em quatro etapas: Processamento HTML, Tokenização, Detecção de Padrões e Seleção de Características [8]. A metodologia completa é apresentada na fig. 2.

A. 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. Seguem o seguinte padrão: < nome-tag parâmetro > Texto da tag < nome-tag >.

O processo HTML, de acordo com o grau de importância, divide as tags em três categorias, e as submete a tratamentos distintos:

1) Na primeira categoria tudo é ignorado, isto é, o nome da tag, seus parâmetros e conteúdo. 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.

B. Tokenização

Emails no formato de texto original, ou e-mails no formato HTML depois de processados, são enviados ao processo de tokenização. O processo simplesmente separa o texto em tokens, ou seja, simples palavras. São utilizados como delimitadores os seguintes caracteres: espaço, nova linha, tabulação, exclamação, interrogação, vírgula e ponto e vírgula. Também neste processo todos os caracteres são passados a forma minúscula e é removida toda acentuação.

C. Detecção de Padrões

Identifica padrões de textos conhecidos e utilizados por spammers como técnicas para evadir filtros de spams, e unifica padrões de texto para se obter uma única saída. Segue os padrões que deverão ser detectados.

- Parâmetros de tags HTML. Exemplo: "< table color = blue >", a saída será "!_table_color";
- E-mail. Exemplo: "compreaqui@loja.com.br", a saída será "!_mail";
- URLs e Hiperlinks. Exemplo: "http://compreaqui.com", a saída será "!_link";
- Palavras que contenham caracteres inválidos. Exemplo: ".m.o-n-e_y", a saída será "!_HIDEWORDS";
- Palavras acima de 20 caracteres será substituída por "!_BIGTEXT";
- Quantidades monetárias: Exemplo: "R$ 20,00", a saída será "!MONEY";
- Porcentagem. Exemplo: "27,39 %", a saída será "!PORCENTAGEM".

D. 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[8]:

$$FD(t) = \frac{n[t \in C]}{T} \tag{1}$$

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:

$$FD_{s-h} = |FD_s - FD_h| \tag{2}$$

sendo

$$FD_s = \frac{n_s[t \in spam]}{T_s} \tag{3}$$

e

$$|FD_h = \frac{n_h[t \in ham]}{T_h} \tag{4}$$

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 uma 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 [8].

	t_0	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
E-mail 1	1	1	0	1	1	1	0	1	0
E-mail 2	1	1	1	1	1	0	1	0	1
E-mail 3	1	1	1	1	0	1	0	1	0
E-mail 4	1	1	0	0	1	0	0	0	0

Fig. 1. Composição de vetores característicos

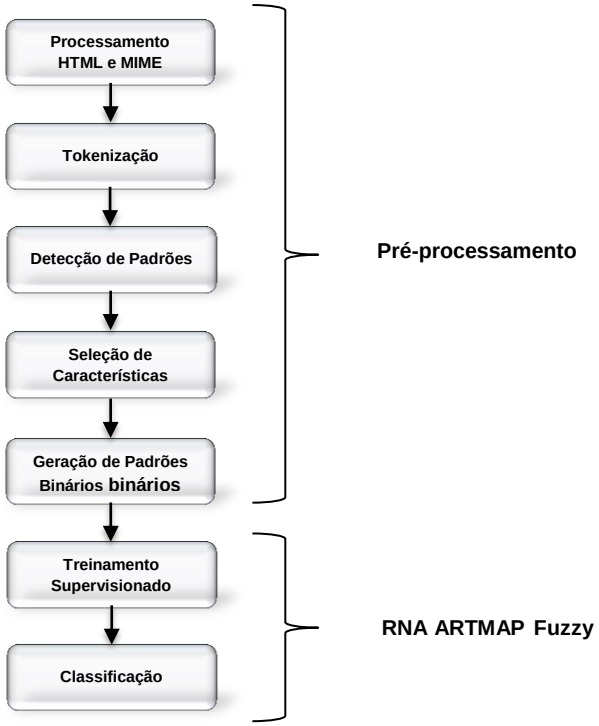

Fig. 2. Diagrama de blocos da metodologia proposta

III. RNA ARTMAP FUZZY

A RNA ARTMAP Fuzzy é caracterizada por um conjunto de equações que implementam as propriedades da Teoria da Ressonância Adaptativa (ART), objetivando a aprendizagem incremental e supervisionada [12].

É orientada a categorizar de forma estável 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. Nas operações são utilizados operadores difusos $MIN(\wedge)$ e $MAX(\vee)$, respectivamente, da teoria de lógica difusa [13].

A ARTMAP Fuzzy constrói 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 RNAs ART Fuzzy (ART_a e ART_b) e em um módulo F_{ab} (módulo Inter-ART) entre as camadas F_2 de ambas as redes [13].

A. Teoria da Ressonância Adaptativa

A ART é um paradigma de RNA desenvolvido no Centro de Sistemas Adaptativos da Universidade de Boston (EUA) por Caperter 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 seres vivos sobrevivem porque aprendem a se adaptar continuamente ao ambiente mutante [14].

As RNAs com arquitetura ART incorporam um modelo de aprendizagem competitiva dentro de uma estrutura de controle auto-organizável, cujo reconhecimento e aprendizado autônomo continuam em resposta a uma sequencia arbitrária de padrões de entrada [14].

B. RNA ART Fuzzy

A RNA ART Fuzzy é formada por três camadas de unidades denominadas por F_0, F_1 e F_2. A camada F_0 possui como padrão de entrada corrente um vetor de atividade M-dimensional $(I_1, \dots, I_M)$, e é uma etapa de pré-processamento que codifica o complemento da entrada para servir a camada F_1 (bottom-up), logo esta com dimensão $2M$ denotada por $x = x_1, \dots, x_{2M}$. F_1 ainda pode receber entrada da camada F_2 (top-down), que representa a categoria ativa, denotada por $y = y_1, \dots, y_{2N}$. Cada categoria interna de F_2 lhe corresponde um vetor de pesos adaptativos $w_j = (w_{j1}, \dots, w_{j(2M)})$, sendo que $j = 1, \dots, N$ e N é o número de possíveis categorias internas de F_2. Inicialmente, cada categoria de F_2 é dita como uncommitted, e após ser selecionada para uma codificação se torna committed, ou ativa [11].

A ART Fuzzy é determinada por um parâmetro de escolha $\alpha > 0$; um parâmetro de taxa de aprendizagem $\beta \in [0,1]$; e um parâmetro de vigilância $\rho \in [0,1]$.

O funcionamento da RNA ART Fuzzy se resume as seguintes etapas [11][15]:

1) *Inicialização dos pesos e parâmetros da ART Fuzzy* – Inicialmente, o vetor de pesos adaptativos e os parâmetros da RNA são arbitrados como segue: $w_{ij} = 1$, $\rho \in [0,1]$, $\alpha > 0$ e $\beta \in [0,1]$.

2) *Processo de normalização* – A proliferação de categorias é evitada na ART Fuzzy se as entradas são normalizadas:

$$I = \frac{a}{|a|} \tag{5}$$

sendo:

a = vetor de entrada;

$I = [I_1\, I_2 \dots I_M]$ (Normalizado);

$|\,.\,|$ = Função norma.

3) *Codificação em complemento* – Um novo padrão de entrada a, sendo cada elemento a_i um número real pertencente ao intervalo [0,1], tem uma codificação complementar. Isto produz um vetor de entrada I com $2M$ elementos, tais como:

$$I = (a, a^c) = (a_1, \dots, a_M, a_1^c, \dots, a_M^c) \tag{6}$$

sendo:

$$a_1^c = 1 - a_1 \tag{7}$$

A codificação em complemento e a regra de normalização preservam a amplitude do padrão de entrada:

$$|I| = |a, a^c| = \sum_{i=1}^{M} a_i + \left(M - \sum_{i=1}^{M} a_i \right) = M \quad (8)$$

4) *Escolha de categoria* – Para cada entrada I, a função de escolha T_j é calculada para todos nós j da camada F_2, a função é definida por:

$$T_j(I) = \frac{|I \wedge w_j|}{\alpha + |w_j|} \quad (9)$$

sendo o símbolo ∧ o operador fuzzy AND definido por:

$$(p \wedge q)_i = \min(p_i, q_i) \quad (10)$$

e a norma | . | definida por:

$$|p| = \sum_{i=1}^{M} |p_i| \quad (11)$$

para qualquer vetores M-dimensionais p e q.

É dito ao sistema para fazer a escolha de categoria apenas quando mais de um nó da camada F_2 estiver ativo. A categoria escolhida é indexada por J, sendo:

$$T_J = \max\{T_j : j = 1 \dots N\} \quad (12)$$

Caso exista mais de um T_J máximo, a categoria j com menor índice é escolhida.

5) *Teste de vigilância (ressonância ou reset)* – A ressonância ocorre se a função de combinação:

$$\frac{|I \wedge w_J|}{|I|} \quad (13)$$

da categoria ativa satisfaz o critério de vigilância:

$$\frac{|I \wedge w_J|}{|I|} \geq \rho \quad (14)$$

Caso o critério de vigilância não seja satisfeito, o sinal de reset é enviado a ART Fuzzy. O valor da função de escolha T_J é fixado em 0 enquanto o padrão de entrada atual estiver presente afim de evitar a seleção persistente da mesma categoria durante a busca. Uma nova categoria de índice J é escolhida pela equação (9), e o processo de escolha de categoria continua até que o índice J satisfaça (11) e a RNA ART Fuzzy entre em ressonância e consequentemente em processo de aprendizagem.

6) *Adaptação dos pesos (aprendizagem)* – Finalizado o processo de escolha de categoria e teste de vigilância, o vetor de pesos w_j é atualizado de acordo com a equação:

$$w_j^{new} = \beta\left(I \wedge w_j^{old}\right) + (1 - \beta)\, w_j^{old} \quad (15)$$

O processo de aprendizagem rápido corresponde em atribuir o valor 1 ao parâmetro de aprendizagem $\beta = 1$.

C. Módulo Inter-ART

A ARTMAP Fuzzy incorpora dois módulos ART Fuzzy, ART_a e ART_b, que são ligadas através do módulo Inter-ART, F^{ab}, chamado map field. O map field é usado para formar associações preditivas entre as categorias e aplicar a regra match tracking, através do qual o parâmetro de vigilância aumenta em resposta a uma incompatível predição de ART_b. Match tracking reorganiza a estrutura de categorias de modo que erros de predição não se repitam na apresentação das entradas posteriores. As interações mediadas por map field F^{ab} pode ser operacionalmente caracterizado como segue:

a) ART_a *e* ART_b *:* Entradas de ART_a e ART_b na forma de codificação em complemento: para ART_a, $I = A = (a, a^c)$; e para ART_b, $I = B = (b, b^c)$ Variáveis de ART_a e ART_b são designados por subscritos e sobrescritos a e b. Para ART_a, $x^a = x_1^a \dots x_{2M_a}^a$ denota o vetor de saída da camada F_1^a; $y^a = y_1^a \dots y_{2N_a}^a$ denota o vetor de saída da camada F_2^a; e $w_j^a = w_{j1}^a, w_{j2}^a \dots w_{j,2M_a}$denota o $j-ésimo$ vetor de pesos de ART_a. Para ART_b, $x^b = x_1^b \dots x_{2M_b}^b$ denota o vetor de saída da camada F_1^b; $y^b = y_1^b \dots y_{2N_b}^b$ denota o vetor de saída da camada F_2^b; e $w_k^b = w_{k1}^a, w_{k2}^a \dots w_{k,2M_a}$denota o $k-ésimo$ vetor de pesos de ART_b. Para o map field, $x^{ab} = x_1^{ab} \dots x_{2N_b}^{ab}$ denota o vetor de saída de F^{ab}, e $w_j^{ab} = w_{j1}^{ab}, \dots, w_{jN_b}^{ab}$ denota o $j-ésimo$ vetor de pesos do F_2^a para F^{ab}. É atribuído o valor 0 para os vetores x^a, y^a, x^b, y^b e x^{ab} entre cada novo padrão de entrada apresentado a RNA.

b) Map Field: O map field F^{ab} é ativado sempre que uma categoria de ART_a ou ART_b estiver ativa. Se o nó J de F_2^a é escolhido, então seus pesos w_j^{ab} ativam F^{ab}. Se o nó K em F_2^b estiver ativo, então o nó K de F^{ab} é ativado entre as vias de F_2^b e F^{ab}. Se ambas os módulos, ART_a e ART_b estiverem ativos, então F^{ab} torna-se ativo somente se ART_a prever a mesma categoria de ART_b através dos pesos w_j^{ab}. O vetor de saída x^{ab} de F^{ab} obedece:

$$x^{ab} = \begin{cases} y^b \wedge w_J^{ab} & \text{nó } J \text{ de } F_2^a \text{ ativo e } F_2^b \text{ ativo} \\ w_J^{ab} & \text{nó } J \text{ de } F_2^a \text{ ativo e } F_2^b \text{ inativo} \\ y^b & F_2^a \text{ inativo e } F_2^b \text{ ativo} \\ 0 & F_2^a \text{ inativo e } F_2^b \text{ inativo} \end{cases} \quad (16)$$

Por (13), $x^{ab} = 0$ se a predição de w_J^{ab} não é confirmada por y^b. Tal incompatibilidade dispara a busca por novas categorias em ART_a, como segue:

c) Match Traking: No início de cada padrão de entrada apresentado a ART_a, o parâmtro de vigilânica é igual ao valor de vigilânica base, $\bar{\rho}_a$. O parâmetro de vigilância do map field é ρ_{ab}. Se

$$|x^{ab}| < \rho_{ab}|y^b| \quad (17)$$

então ρ_a é ligeiramente incrementado até que seja maior que:

$$|x^a| = |A \wedge w_j^a| < \rho_a|A| \quad (18)$$

sendo J o nó ativo de F_2^a. Quando isto acorre, ART_a busca novamente outro nó J de F_2^a para ativação com:

$$|x^a| = |A \wedge w_J^a| \geq \rho_a |A| \tag{19}$$

e

$$|x^{ab}| = |y^b \wedge w_J^{ab}| \geq \rho_a |y^b| \tag{20}$$

d) Aprendizagem do map field: As regras de aprendizagem determinam como os vetor de pesos do map field w_{jk}^{ab} muda ao longo do tempo, como se segue. O vetor de pesos w_{jk}^{ab} via $F_2^a \rightarrow F^{ab}$ inicialmente satisfaz

$$w_{jk}^{ab} = 1 \tag{21}$$

Com a aprendizagem rápida, uma vez que J aprende a predizer a categoria K de ART_b, $w_{JK}^{ab} = 1$, esta associação é permanente.

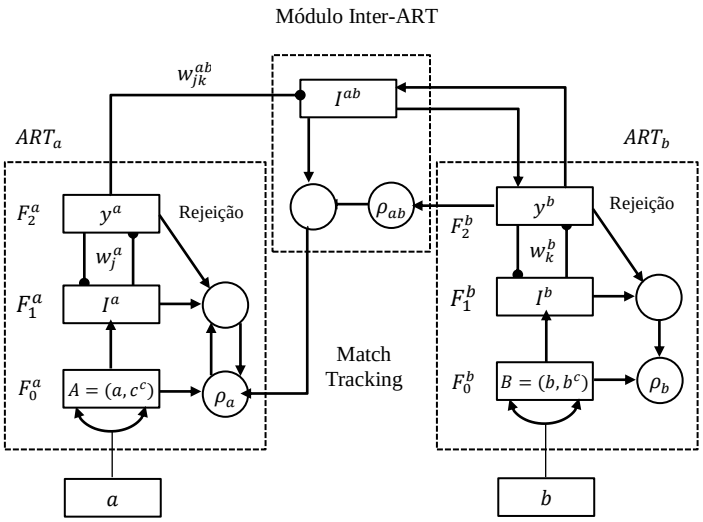

Fig. 3. Estrutura da RNA ARTMAP Fuzzy

IV. RESULTADOS

A. Forma de análise dos resultados

O método proposto foi submetido a testes a fim de se obter os melhores valores para os parâmetros e consequentemente os melhores resultados.

Sejam S e H os conjuntos de mensagens de Spams e Hams, respectivamente, e todos os resultados possíveis da predição apresentados na Tabela I.

TABELA I. POSSÍVEIS RESULTADOS NA PREDIÇÃO DE E-MAILS

Notação	Denominação	Descrição
VP	Verdadeiro Positivo	Spam classificado corretamente
VN	Verdadeiro Negativo	E-mail legítimo classificado corretamente
FN	Falso Negativo	Spam classificado como e-mail legítimo
FP	Falso Positivo	E-mail legítimo classificado como Spam

A avaliação é feita através das medidas de desempenho Tfp (Taxa de falso positivo), Tfn (Taxa de falso negativo) e Ter (Taxa de erro), apresentados na Tabela II.

TABELA II. MEDIDAS DE DESEMPENHO PARA AVALIAR MÉTODOS DE CLASSIFICAÇÃO

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

A base de dados utilizada para realização dos testes foi a SpamAssasin [5], formado por 6.047 e-mails divididos em três subconjuntos: Spam – Contém 1.897 Spams; EasyHam – 3.900 e-mails legítimos facilmente diferenciáveis dos spams por não conterem assinaturas de spams como por exemplo tags HTML; Hard Ham – 250 e-mails legítimos, porém mais difíceis de diferenciar dos spams por conterem assinaturas de spams. Portanto, o total de e-mails possui aproximadamente 31% de spams.

Alguns parâmetros da RNA ARTMAP Fuzzy se mantiveram os mesmos em todos os testes, como apresenta a Tabela III:

TABELA III. PARÂMETROS RNA ARTMAP FUZZY

Parâmetro	Valor
$\boldsymbol{\rho_{ab}}$	1
$\boldsymbol{\rho_b}$	1
$\boldsymbol{\beta}$	1
$\boldsymbol{\alpha}$	0.01

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 menores, como o escolhido 0.01, assim como a taxa de treinamento β com o valor 1. Conforme a seção III estes parâmetros devem estar entre [0,1]. Neste trabalho os valores foram escolhidos por tentativa e erro de acordo com aqueles que apresentaram os melhores resultados.

O parâmetro de vigilância ρ_a, a porcentagem de e-mails destinados ao treinamento, e o tamanho do vetor de entradas M do módulo ART_a 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 a RNA; a porcentagem de e-mails utilizados no treinamento determina a quantidade de informações adquiridas pela RNA durante o processo de aprendizagem, essencial na fase de classificação; e o tamanho do vetor de entrada M representa o número de características extraídas na fase de pré-processamento que compõem o vetor característico.

A escolha de tamanhos menores do vetor M influencia 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 dois padrões iguais indicando saídas distintas, principalmente por se tratar de uma RNA com aprendizagem supervisionada. A escolha de valores maiores para M indica um maior número 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} e contribuir para o aumento da taxa de erros (Ter).

B. Resultados

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

TABELA IV. 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**

TABELA V. 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**

TABELA VI. RESULTADOS PARA $M = 200$

Treino(%)	Ruídos(%)	ρ_a	Tfn	Tfp	Ter
50	0.00	0.90	**11.052**	5.111	6.965
60	0.00	0.92	15.160	4.099	7.549
70	0.00	0.99	16.014	5.145	8.527
80	0.00	0.98	14.209	3.865	7.077
90	0.00	0.98	14.130	**1.208**	**5.184**

CONCLUSÃO

O uso de RNAs com arquitetura baseada na teoria da ressonância adaptativa para problemas de 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 [8] e a RNA ARTMAP Fuzzy no processo de classificação.

No modelo proposto, o pré-processamento foi responsável por extrair as características mais relevantes dos e-mails pertencentes às classes dos Spams e dos e-mails legítimos, a fim de se gerar o vetor característico. Os padrões binários referentes a cada e-mail foram formados a partir do vetor característico e a RNA ARTMAP Fuzzy responsável por classificar os padrões de entrada entre Spam e e-mail legítimo.

Foram realizados testes com diferentes dimensões do vetor de entrada M do módulo ART_a da RNA, 20, 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 de M, se têm características de menor qualidade, já que o valor de DF_{s-h} é menor, mas, no entanto, contribui para a não ocorrência de ruídos. Logo pode-se concluir que o vetor M de tamanho menor e a não ocorrência de ruídos contribui para a melhora dos resultados.

REFERÊNCIAS

[1] D.Gudkova, "Kaspersky Security Bulletin: Spam Evolution 2012," Aceso em: 18 de Janeiro de 2013, disponível em: http://www.securelist.com/en/analysis/204792276/Kaspersky_Security_Bulletin_Spam_Evolution_2012, 2013.

[2] Upasana e S. Chakravarty, "A Survey of Text Classification Techniques for E-mail Filtering," Second International Conference on Machine Learning and Computing, IEEE Computer Society, 2010.

[3] T.A. Almeida, "SPAM: do Surgimento à Extinção", Tese (Doutorado) Faculdade de Engenharia, Universidade Estadual de Campinas, 2010, pp. 114.

[4] Q. Ma, Z. Qin, F. Zhang, e Q. Liu, "Text Spam Neural Network Classification Algorithm," IEEE, pp. 466-469, 2010.

[5] SpamAssassin, Acesso em: Julho de 2012, disponível em: http://spamassassin.apache.org/.

[6] K. Manjusha e R. Kumar, "Spam Mail Classification Using Combined Approach of Bayesian and Neural", IEEE Computer Society, 2010, pp.145-149.

[7] Weka, disponível em: http//www.cs.waikato.ac.nz/ml/weak.

[8] O.A. Carpinteiro, I. Lima e J. M. Assis, "A Neural Model in Anti-spam Systems," Springer-Verlag Berlin Heidelberg, 2006, pp. 847-855.

[9] A. M. Silva, "Utilização de Redes Neurais Artificiais para Classificação de Spams". *Dissertação (Mestrado)* - Centro Federal de Educação Tecnológica de Minas Gerais, 2009, pp. 126.

[10] S. Haykin, "Redes Neurais Princípios e Práticas", 2ª ed., Bookman Companhia Editora, 2008, pp. 908.

[11] G. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds e D. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," IEEE Transactions on Neural Networks, vol. 3(5), 1992, pp. 698 – 713.

[12] J.A. Freeman e D. M. Skapura, "Neural Networks: Algorithms, applications, and programming techniques," Addison-Wesley, 1991.

[13] D.G. Amorim, "Redes Art com Categorias Internas de Geometria Irregular," Tese (Doutorado) – Universidade de Santiago de Compostela – Departamento de Eletrônica e Computação, Santiago de Compostela, 2006, pp. 246.

[14] E.A. Capuano, "O poder cognitivo das redes neurais artificiais modelo ART1 na recuperação de informação," Ci. Inf, Brasília, 2009, pp. 9-30.

[15] M. C. G. Silveira, A.D. P. Lotufo e C. R. Minussi, "Transient Stability Analysis of Electrical Power Systems Using a Neural Network Based on Fuzzy ARTMAP," IEEE Bologna Tech Conference, Bologna, Italy, 2003.