

NEURAL NETWORKS IN E-MAILS SPAM IDENTIFICATION

Alisson Marques da Silva (CEFET-MG, MG, Brasil) – alisson@lsi.cefetmg.br

Gray Farias Moita (CEFET-MG, MG, Brasil) – gray@dppg.cefetmg.br

Paulo E. M. Almeida (CEFET-MG, MG, Brasil) – pema@dri.cefetmg.br

Abstract. *E-mail has become one of the most important ways of communication. The main problem in its use is the amount of undesired messages received daily. Besides being very annoying for the users, spam can cause a great deal of damage, such as the loss of bandwidth, productivity and time, delays and non transmission of valid messages. The continuous challenge is to develop tools to eliminate spam and to decrease the trouble caused by them. Therefore, this work proposes the implementation of an anti-spam system to filter the messages, sought to classify e-mails in order to differentiate between spam and legitimate e-mails, using Artificial Neural Networks (ANN). The messages will pass through a standardisation process, capable of transforming complex information into simpler parts, allowing for a better performance with regard to their classification. The feature selection method, known as Document Frequency (DF), is used to reduce the dimension of each message. Hence, the basic objective of this study is to present the Multilayer Perceptron Networks as a classifier agent and the variations of the DF feature selection method. Good results are obtained and are presented and evaluated in the present paper.*

Keywords: *E-mail, Spam, Artificial Neural Network, Multilayer Perceptron.*

REDES NEURAIIS NA IDENTIFICAÇÃO DE E-MAILS SPAM

Resumo. *O correio eletrônico tornou-se um dos mais importantes meios de comunicação. O principal problema na sua utilização é a quantidade de mensagens indesejadas recebidas diariamente. Além dos transtornos aos usuários, os spams podem causar grandes prejuízos como: desperdício de largura de banda da rede; perda de tempo e produtividade; atrasos ou não recebimento de mensagens legítimas. Muitos esforços de pesquisa têm sido feitos para desenvolver ferramentas de combate aos spams e reduzir os problemas causados por eles, sendo este, um desafio contínuo. Dentre os métodos utilizados para a detecção de e-mails spam estão as redes neurais artificiais. Este trabalho apresenta um filtro anti-spam no qual as mensagens passam por um processo de uniformização, responsável por transformar as informações complexas, presentes em cada mensagem em informações mais simples, permitindo um melhor desempenho na classificação. O método de seleção de características Distribuição por Frequência (DF) é empregado para reduzir a dimensionalidade das informações presentes em cada mensagem e as redes neurais Multilayer Perceptron (MLP) como agente classificador. O objetivo fundamental desta pesquisa é apresentar as redes neurais e as variações do método de extração de características DF. Bons resultados foram obtidos e são apresentados e avaliados no presente trabalho.*

Palavras-Chaves: *E-mail, Spam, Redes Neurais Artificiais, Multilayer Perceptron.*

1. INTRODUÇÃO

Cada vez mais presente na vida das pessoas, a internet revoluciona a maneira de se obter informações, de se fazer negócios e, até mesmo, a de se relacionar. Vários serviços são disponibilizados através da internet, entre eles o *e-mail* ou correio eletrônico, que pode ser definido como uma forma de criar, enviar e receber mensagens por intermédio de sistemas eletrônicos.

O correio eletrônico foi criado antes da internet, mas seu uso generalizado se deu com o surgimento e popularização desta, tornando-se um dos mais importantes meios de comunicação. Segundo pesquisa realizada pela *Meta Group* (Web, 2003), 80% dos usuários comerciais preferem o *e-mail* ao telefone. Entre outras vantagens citadas está a rapidez na comunicação, a possibilidade de se comunicar com várias pessoas simultaneamente além de gerar um registro dessa comunicação.

A grande popularização do correio eletrônico fez com que este serviço se tornasse bastante utilizado para envio de *spam* - termo é utilizado para denominar mensagens eletrônicas que, na maioria das vezes são de intuito publicitário, visando à promoção de serviços, produtos ou eventos, e são enviadas sem o consentimento prévio dos destinatários. São mensagens não solicitadas e endereçadas a um grande número de pessoas (Jessen e Chaves, 2003). Um dos maiores motivadores para o envio de *spam* é o baixo custo para atingir grande número de destinatários.

Desde 1994, quando surgiu o primeiro *spam*, a prática de enviar *e-mails* não solicitados tem sido aplicada com objetivos diversos, utilizando para isso, diferentes aplicativos e meios de propagação na rede. O *spam* tem evoluído, acompanhando o desenvolvimento da internet, das novas aplicações e tecnologias. Atualmente, está associado não somente ao envio de mensagens publicitárias, mas também a ataques à segurança da internet e do usuário, propagando vírus e golpes.

O *spam* pode causar prejuízos a usuários e empresas, como: não recebimento de *e-mails* legítimos; perda de produtividade e tempo gasto para identificar e remover os *spams*; acesso a conteúdo impróprio ou ofensivo; prejuízos financeiros causados por fraude; impacto no *link* de conexão com a internet através do aumento no tráfego de rede; má utilização dos servidores devido à necessidade de armazenar e processar os *spams*; inclusão em listas de bloqueio; investimento em pessoal e equipamentos (Antispam.BR, 2008).

O desenvolvimento de mecanismos de prevenção, detecção e proteção dos correios eletrônicos tornaram-se imprescindíveis para o combate ao *spam*. Vários métodos para identificar e classificar os *spams* foram propostos, entre eles: listas de bloqueios, *greylisting* e filtros de conteúdos com diferentes técnicas, porém segundo Ozgur et. al. (2004) nenhum deles é completamente satisfatório.

Este trabalho avalia e compara três variações do método de seleção de características Distribuição por Frequência e as redes *Multilayer Perceptron* (MLP) empregados na classificação de *e-mails* em legítimos e *spam*.

Para apresentação do trabalho, este texto encontra-se organizado como se segue: em um primeiro momento, na Seção 2, será apresentada uma introdução à categorização de textos, destacando as principais etapas do processo e os métodos de seleção de características. Posteriormente, será apresentada uma introdução às redes neurais com ênfase nas MLP, na Seção 3. Depois disso, na Seção 4 será apresentado um relato dos experimentos executados e as medidas de desempenho utilizadas. Os resultados alcançados são apresentados na Seção 5. Por fim, na Seção 6 são apresentadas as principais conclusões obtidas com os experimentos.

2. CLASSIFICAÇÃO DE TEXTOS

A classificação de textos é um processo cujo objetivo é classificar documentos com base no seu conteúdo em relação a um conjunto de categorias pré-definidas, isto é, uma técnica para atribuir automaticamente um documento textual a um ou mais conjuntos (Rizzi et. al., 2000).

A primeira etapa do processo consiste em definir o conjunto de dados e as categorias. Em seguida, os documentos passam pelo processo de preparação no qual o texto é uniformizado. O próximo passo é a seleção de características que identificam as palavras que melhor representam as classes. Essas palavras são escolhidas para compor um vetor de entrada do agente classificador. Os passos descritos acima são ilustrados na Figura 1. Este mesmo procedimento pode ser empregado na classificação de mensagens eletrônicas em legítimas e *spams*, conforme apresentado nas próximas seções.

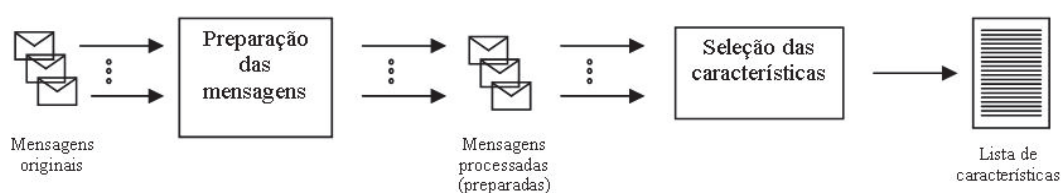


Figura 1 – Fases do processo de categorização.

2.1 Conjunto de Dados

A definição de um bom conjunto de dados, com mensagens representativas das categorias definidas, é de grande relevância para o sucesso na classificação. Devido à existência de diversos tipos de mensagens legítimas e *spams*, é importante definir um conjunto de dados que contemple satisfatoriamente todos os tipos.

Na internet estão disponíveis diversos *corpus* de mensagens que são utilizados pela comunidade científica em seus experimentos, dentre eles: UCI-R

Hopkins et. al. (2006); Sakkis et. al. (2003) realizou seus experimentos com a base do Ling-Spam; as bases PU1 e U5Spam foram empregadas no trabalho de Clark et. al. (2003); Carpinteiro et. al. (2006) e Chuan et. al. (2005) utilizaram o conjunto do *SpamAssassin*.

Neste trabalho utiliza-se um *corpus* de mensagens criado a partir de um dos subdomínios do CEFET-MG. O conjunto foi compilado de mensagens pessoais de todos os usuários do subdomínio, cada qual com um diferente grau de exposição na rede e perfil de utilização do *e-mail*, juntamente com os *spams* recebidos nos mesmos endereços eletrônicos. As mensagens estão em seu formato original, sem remoção de *tags* HTML, anexos ou conteúdo. Isso é essencial para a proposta deste trabalho em utilizar todo o conteúdo da mensagem. Este conjunto representa fielmente as mensagens recebidas por um usuário comum. O conjunto é composto por 12.687 mensagens, sendo 8.867 legítimas e 4.020 spams.

2.2 Preparação das Mensagens

O objetivo dessa etapa é realizar uma uniformização do conteúdo das mensagens, transformando as informações complexas presentes em cada mensagem, em informações mais simples, permitindo um melhor desempenho na classificação. Neste processo, todos os *e-mails* do conjunto devem ser analisados e uniformizados.

Várias técnicas de preparação de textos podem ser empregadas, entre elas: remoção de *stopwords*; remoção de caracteres inválidos; identificação de termos, identificação de termos compostos e a técnica de *stemming*.

Nesta pesquisa emprega-se o método proposto por Carpinteiro et. al. (2006), no qual: todos os caracteres são convertidos em minúsculos; imagens, anexos, *links*, endereços eletrônicos, moeda, porcentagem e palavras longas são substituídos por *strings* específicos; os acentos são removidos; *tags* html são tratadas, sendo que algumas são utilizadas integralmente e outras parcialmente ou descartadas; palavras pequenas são descartadas; entre outras.

2.3 Seleção de Características

Em um conjunto de dados uma característica é uma palavra, e o total de características é o número de palavras que ocorrem em todo o conjunto. Este geralmente é um número elevado (alta dimensão do espaço característico), o que dificulta ou até impossibilita a utilização das técnicas de classificação devido ao seu elevado custo computacional e, conseqüentemente, a uma classificação não eficiente.

A seleção de características consiste na extração das informações mais relevantes de um conjunto de dados, identificando as palavras que melhor representam uma categoria, ou seja, é um método utilizado para reduzir a

dimensionalidade do espaço característico. É recomendável que essa redução seja realizada de forma automática e, principalmente sem sacrificar a precisão na classificação dos documentos.

Diferentes algoritmos podem ser empregados para realizar essa tarefa, dentre eles: Ganho de Informação (IG) usado por Sakkis et. al. (2003); Informação Mútua (Mi) empregado por Androutsopoulos et. al. (2000), Ozgur et. al. (2004); χ^2 statistic (Qui-quadrado) utilizado no trabalho de Yan e Petersen (1997) e; Distribuição por Freqüência (DF).

O método de Distribuição por Freqüência objeto de estudo deste trabalho é descrito na próxima seção. A escolha desse método se deu pelo seu desempenho nos experimentos de Yang e Pedersen (2003) e Carpinteiro (2006) e, também por ser o mais simples dos métodos, necessitando menores recursos computacionais. Outro motivador é a possibilidade de utilizar variações nos cálculos que não foram implementadas nos trabalhos supracitados.

2.4 Distribuição por Freqüência (DF)

É uma das técnicas mais simples para redução da dimensionalidade. Possui uma complexidade computacional aproximadamente linear, o que possibilita seu uso em grandes conjuntos de dados a um custo computacional relativamente pequeno.

A distribuição por freqüência é definida pelo número de ocorrência de um termo em um conjunto de elementos (Yan e Pedersen, 1997). O cálculo do DF de uma palavra se dá por meio da Equação 1,

$$DF = \frac{N[x \in \{legitima, spam\}]}{T} \quad (1)$$

onde, N é o número de ocorrência da palavra x na classe (legítima ou *spam*) e T , o número total de palavras na classe.

Para representar as classes são escolhidas as palavras com valores de DF mais altos, considerando-se que as palavras com baixa freqüência de ocorrência são menos significativas para identificação das classes. Nessa técnica a palavra possui um valor de DF para cada uma das classes. Nesse caso específico, um DF para o conjunto de mensagens legítimas e outro para o conjunto de *spams*. A proposta deste trabalho é variar a forma de cálculo do DF unificado da palavra (DF Legítimo e DF *Spam*). O objetivo dessas diferentes formas de cálculo é conseguir encontrar as características que melhor definam cada uma das classes e, conseqüentemente melhorem o desempenho na classificação. Abaixo são apresentadas as três variações implementadas nesta pesquisa.

DF Soma (DF+): O cálculo comumente utilizado pela comunidade científica para encontrar o DF de uma palavra é somar o DF da palavra em cada uma das classes, como pode ser visto na Equação 2. As palavras com valores de DF mais

altos são selecionadas para compor o vetor de características. Esse método foi empregado no trabalho de Carpinteiro et. al. (2006).

$$DF+ = DF_Legítimo + DF_Spam \quad (2)$$

DF Exclusão de Termos Comuns (DF ETC): Algumas palavras possuem um elevado número de ocorrência nas duas classes e conseqüentemente um alto valor de DF em ambas. Baseado no princípio de que se uma palavra possui alta representatividade em duas classes distintas, não será uma boa representante para uma classe específica.

Diante dessa premissa, este método busca excluir as palavras com elevado valor de DF que estão presentes nas duas classes. Selecionando somente, as palavras possuem elevado DF em apenas uma delas. Por exemplo, num conjunto de 100 palavras de cada uma das classes, são selecionadas para compor o vetor de características as palavras que estão presentes em somente uma das classes.

DF Subtração (DF-): Nesta técnica são selecionadas as palavras que possuem a maior diferença entre o seu valor de DF nas duas classes. O DF é obtido através do módulo (valor absoluto) da subtração entres os valores de DF da palavra nas duas classes. Após o cálculo descrito na Equação 3, as palavras com maior DF são selecionadas para compor o vetor de características.

$$DF- = |DF_Legítimo - DF_Spam| \quad (3)$$

3 REDES NEURAIIS ARTIFICIAIS

As redes neurais artificiais são modelos matemáticos que se assemelham às estruturas neurais biológicas e que têm capacidade computacional adquirida por meio de aprendizagem e generalização (Haykin, 1999). Esses modelos almejam semelhança com o sistema nervoso dos seres vivos e com a sua capacidade de processar informações. Trata-se de uma metáfora da maneira como o cérebro humano processa as informações utilizadas em computação.

A capacidade de aprender com exemplos, robustez, velocidade de processamento, generalização e adaptabilidade, possibilita a utilização das redes neurais na solução de uma grande variedade de problemas, entre eles, problemas de classificação, otimização, categorização, aproximação, análise de sinais ou imagens e predição (Braga e Carvalho, 2000).

Os problemas de categorização e classificação envolvem a descoberta de características estatisticamente relevantes de um determinado conjunto de dados, e como estes, podem ser divididos em classes. Nesses casos, somente os dados de entradas são conhecidos, cabendo à rede encontrar as classes presentes nesse conjunto de dados. Por todas as suas características e pelo sucesso de sua utilização na categorização de informação, as redes neurais artificiais podem ser utilizadas para detecção de spam em mensagens de correio eletrônico.

3.1 Rede Neural Multicamada

As redes neurais com uma ou mais camadas escondidas são denominadas de redes multicamadas, também conhecidas como redes MLP (*Multilayer Perceptron*). Sua arquitetura consiste de um conjunto de unidades sensoriais que formam a camada de entrada, uma ou mais camadas ocultas (intermediárias) de unidades computacionais e uma camada de saída (Ramos, 2001). Os sinais de entrada são propagados em sentido unidirecional, camada por camada, da entrada para a saída.

Como pode ser visto na arquitetura de rede MLP apresentada na Figura 2, os neurônios de uma determinada camada estão completamente conectados, isto é, cada neurônio desta camada é conectado a cada um dos neurônios da camada subsequente. Não existem conexões entre os neurônios de camadas não adjacentes, e nem conexões entre neurônios em uma mesma camada (Almeida, 2006).

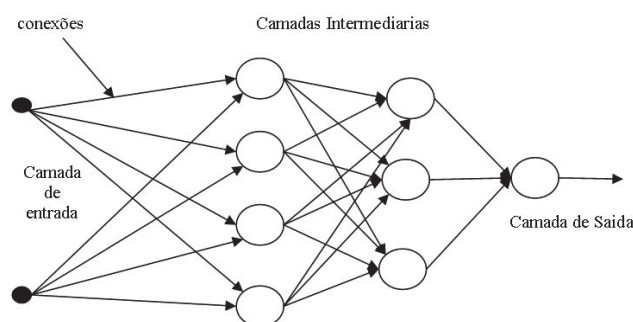


Figura 2 – Arquitetura de uma rede neural MLP

Geralmente a implementação de uma rede neural MLP passa pelas seguintes etapas: definição do número de entradas e saídas da rede; escolha da arquitetura da rede (número de camadas, número de neurônios por camada e respectivos tipos e função de ativação); ajuste dos pesos das conexões ou aprendizagem (treinamento) da rede. Uma MLP pode ter ou não retroalimentação.

Em uma rede neural a função ativação de um neurônio pode ser linear, tangente *sigmoidal*, logarítmica *sigmoidal* (logsig), entre outras. A função logarítmica *sigmoidal*, descrita na Equação 4 apresenta uma saída compreendida entre [0 1].

$$y = \int(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

No algoritmo de treinamento com retro-propagação (*Backpropagation*) os sinais são propagados pela rede, camada por camada, até a apresentação do resultado na camada de saída. O resultado é comparado com o desejado para esse padrão particular. Se a saída obtida comparada com a saída esperada não

estiver correta, o erro é calculado. Na iteração de treinamento t , o sinal do erro na saída do neurônio n é definido pela Equação 5,

$$erro_n(t) = y_{Dn}(t) - y_n(t), \quad (5)$$

onde $y_{Dn}(t)$ é a saída desejada para o neurônio n na iteração t . O erro é propagado a partir da camada de saída até a camada de entrada. Os pesos das conexões dos neurônios das camadas ocultas vão sendo modificados conforme o erro é retro-propagado (Almeida, 2006).

Existem diversos métodos para fazer a correção dos pesos, entre eles o de *Levenberg-Marquart* (LM). O LM pode ser visto como uma combinação das vantagens do método de *Gauss-Newton* e do gradiente descendente. Usa a velocidade de convergência do *Gauss-Newton* quando a solução está longe do mínimo local e o gradiente descendente quando a região próxima da função objetivo é alcançada. O cálculo da iteração para esse método pode ser descrita pela Equação 6,

$$X_{k+1} = X_k - [J^T J + \mu I]^{-1} J^T e \quad (6)$$

onde J é a Matriz Jacobiana, que contém as derivadas primeiras dos erros da rede com relação aos pesos e bias, $H = J^T J$ é a aproximação do Hessiano, o qual contém as derivadas segundas dos erros da rede com relação aos pesos e bias. O valor escalar μ define qual método de otimização prevalece. Quando μ assume um valor elevado, o método utilizado é o gradiente descendente e, ao diminuir μ , a técnica empregada aproxima-se do método *Gauss-Newton*. Assim, a manipulação de μ visa usufruir as vantagens dos dois métodos, minimizando o efeito das suas desvantagens (Eyng, 2006).

4 EXPERIMENTOS

Nos 18 experimentos realizados, foram empregadas as três variações supracitadas do método de seleção de características Distribuição por Freqüência. A técnica de validação cruzada, que objetiva encerrar o treinamento quando o erro de validação começa a aumentar, indicando que a rede está aprendendo com o ruído presente nos dados, foi implementada neste trabalho.

Segundo Haykin (1999), na validação cruzada o conjunto de dados é dividido aleatoriamente em um conjunto de treinamento, validação e teste. O conjunto de treinamento é utilizado no processo de aprendizagem da rede. O de validação serve para interromper a aprendizagem, quando o erro médio quadrado para os dados de validação começa a aumentar de forma contínua. Os dados de testes são utilizados para verificar a capacidade de generalização da rede. O conjunto de dados foi distribuído da seguinte forma: 60% para treinamento, 20% para teste e 20% para validação (Demuth, 2008).

A arquitetura da rede MLP é composta por uma camada de entrada, uma oculta e uma de saída. A função de ativação *logsig* foi empregada na camada

oculta e na de saída. Para treinamento da rede foi utilizado o algoritmo de *Levenberg-Marquart*. Os vetores de entrada foram gerados usando indexação por peso binário.

Os casos de testes foram gerados para os vetores de entrada com 25 e 50 características, isto é, número de neurônios na camada de entrada da rede. Para a camada oculta foram utilizados 5, 25 e 50 neurônios e na camada de saída um único neurônio. Cada uma dessas combinações foi implementada com as três variações do método de Distribuição por Frequência. Cada experimento foi executado 30 vezes, considerando nas avaliações as suas médias e o melhor resultado.

4.1 Medidas de Desempenho

O desempenho na classificação e categorização pode ser avaliado através de diferentes métodos. Esses podem ir desde o cálculo da abrangência e precisão simples, até o cálculo da acurácia (*accuracy*), *ranking*, micro e macro-média, entre outros (Yang e Petersen, 1997).

Em tarefas de classificação o desempenho é freqüentemente mensurado pelo processo de acurácia (Acc) ou pela taxa de erro (Androutsopoulos et. al., 2000). Neste trabalho optou-se por medir a acurácia na classificação dos *e-mails* legítimos e *spams*. As letras L e S foram usadas para *e-mails* legítimos e *spam* respectivamente, n_L e n_S como o número total de mensagens legítimas e *spams*, $n_{L \rightarrow L}$ e $n_{S \rightarrow S}$ como o total de mensagens legítimas e *spams* classificados corretamente.

A acurácia, calculada através da Equação 7, indica a proporção de classificações corretas.

$$Acc = \frac{n_{S \rightarrow S} + n_{L \rightarrow L}}{n_S + n_L} \quad (7)$$

5 RESULTADOS

A Tabela 1 apresenta os resultados alcançados nos experimentos. Ela mostra o percentual de acerto, ou seja, a acurácia (exatidão) na classificação dos *e-mails* legítimos e dos *spams*, através da média e do melhor resultado de cada experimento.

Os resultados obtidos são bastante promissores, principalmente em função da utilização de um método de seleção de características simples e com baixo custo computacional.

Experimento	N.º de Entradas	Camadas Ocultas	Método de Seleção de Característica	Classificação Correta (%)	
				Acc -	
				Média	Melhor
1	25	5	DF+	98.16	98.74
2	25	25	DF+	98.22	98.78
3	25	50	DF+	98.17	98.74
4	50	5	DF+	97.77	98.25
5	50	25	DF+	97.77	98.33
6	50	50	DF+	97.85	98.49
7	25	5	DF -	98.32	98.77
8	25	25	DF -	98.50	98.81
9	25	50	DF -	98.55	98.89
10	50	5	DF -	98.23	98.61
11	50	25	DF -	98.16	98.85
12	50	50	DF -	98.27	98.85
13	25	5	DF ETC	98.26	98.97
14	25	25	DF ETC	98.57	98.97
15	25	50	DF ETC	98.66	99.10
16	50	5	DF ETC	98.34	98.72
17	50	25	DF ETC	98.24	98.88
18	50	50	DF ETC	98.34	98.88

Tabela 1 – Resultado dos experimentos

Nos experimentos com 25 elementos de entrada da rede o melhor resultado foi obtido no experimento 15 pelo DF ETC. O qual classificou corretamente 99,14% dos *e-mails*. Este experimento também obteve a melhor média com 25 neurônios de entrada com 98,66% de acerto.

Com 50 características na entrada da rede, o DF Subtração com 50 camadas ocultas obteve o melhor resultado com 98,97% de classificações corretas. O melhor desempenho com base na média dos experimentos também foi obtido pelo DF Subtração com 50 camadas ocultas, classificando corretamente 98,55% das mensagens.

O melhor resultado alcançado nos experimentos 15 no qual o erro foi inferior a 1% na classificação dos *e-mails*. Para entender melhor o bom desempenho dos resultados, estes foram comparados com os obtidos em recentes pesquisas. Em Andrade (2006), as redes neurais MLP obtiveram 94% de acerto e o sistema Neuro-Fuzzy 88%. No trabalho de Chuan et. al. (2005), as redes MLP obtiveram 91.26% de acerto, as redes LVQ 93,58% enquanto o método Naive Bayesian 86,48%. Estes resultados são inferiores aos apresentados na Tabela 1.

6 CONCLUSÃO

Este trabalho propôs um filtro anti-spam utilizando redes MLP. A rede foi treinada e testada no *corpus* de mensagens do CEFET-MG. As mensagens passaram por um pré-processamento com o objetivo de uniformizar as informações, ou seja, torná-las mais simples e eliminar informações desnecessárias. Foram empregadas três variações do método de seleção de características Distribuição por Frequência para extrair as informações mais relevantes dos conjuntos de mensagens e compor o vetor de entrada da rede.

Os resultados dos experimentos se mostraram satisfatórios com alto percentual de classificações corretas, mostrando a capacidade da rede MLP de generalizar em diferentes padrões de *e-mails* legítimos e *spams*. A implementação das variações no cálculo do DF se mostrou bastante satisfatória. Os dois métodos propostos, DF Subtração e DF Exclusão Termos Iguais, apresentaram resultados superiores aos do DF Soma, que é o método comumente empregado pela comunidade científica.

Em continuação a esta pesquisa, novos métodos de seleção de características, arquiteturas de redes neurais e algoritmos de treinamento estão sendo testados.

REFERÊNCIAS BIBLIOGRÁFICAS

- Almeida, P. E. M. (2006). *Notas de aula da disciplina de Inteligência Computacional*, Curso de Mestrado em Modelagem Matemática e Computacional, CEFET-MG, Belo Horizonte-MG.
- Andrade, L. M. (2006), *Análise Comparativa de Técnicas de Inteligência Computacional para a Detecção de Spam*, Dissertação de Mestrado, UFMG - Programa de Pós-Graduação em Engenharia Elétrica, Belo Horizonte-MG.
- Androutsopoulos, I.; Koutsias, J. (2000), *An experimental comparison of naïve bayesian and keyword-based anti-spam filtering*, ACM In: Proceedings of the 23rd Annual International, Athens, Greece, p. 160-167.
- AntiSpam.BR (2008), *AntiSpam*, Página na Internet. Disponível em: <http://www.antispam.br>.
- Braga, A. P.; Carvalho, A. P. L. F. (2000), *Redes Neurais Artificiais: Teoria e aplicações*. Rio de Janeiro, LTC.
- Carpinteiro, O. A. S.; Lima, I.; Assis, J. M. C.; Souza, A. C. Z.; Moreira, E. M.; Pinheiro, C. A. M. (2006), *A neural model in anti-spam systems*, Lecture Notes in Computer Science, v. 4132, p. 842-855.
- Chuan, Z.; Xianling, L.; Menshu, H.; Xu, Z. (2005), *A LVQ-based neural network anti-spam e-mail approach.*, ACM SIGOPS Operating System Review, v. 39, p.34-39.
- Clark, J.; Koprinska, I.; Poon, J. (2003), *Linger - A Smart Personal Assistant for E-mail Classif*, ICANN/ICONIP.
- Demuth, H.; Beale, M.; Hagan, M. (2008), *Neural Network Toolbox 6: User's Guide*, The MathWorks.

- Eyng, E. (2006), *Controle FeedForward Baseado em Redes Neurais Aplicada a Coluna de Absorção do Processo de Produção de Etanol*, Dissertação de Mestrado, Unicamp, Campinas-SP.
- Haykin, S. (1999), *Neural Networks: A Comprehensive Foundation*. 2ª. ed. Prentice-Hall.
- Hopkins, M.; Reeber, E.; Forman, G.; Suermondt J. (2008). *UCI – Machine Learning Repository: Spambase Data Set*, Página da Internet, Disponível em < <http://archive.ics.uci.edu/ml/datasets/Spambase> >.
- Jessen, K. S.; Chaves, M. H. P. C.; Hoepers C. (2003), *Projeto de desenvolvimento de um sistema de controle e acompanhamento e notificações de spam*, V Simpósio de Segurança em Informática – São José dos Campos.
- Ozgur, L.; Gungor, T.; Gurgun F. (2004), *Adaptive anti-spam filtering for agglutinative languages: a special case for turkish*. Pattern Recognition Letters 25, p. 1819-1831.
- Ramos, J. P. S. (2001), *Aplicação de Redes Neurais Artificiais Multicamadas Estáticas no Processo de Seleção de Frutos*, Tese de Doutorado, Unicamp, Campinas-SP.
- Rizzi, C.; Wives, L.; Engel, P. M.; Oliveira, J. P. M. (2000), *Fazendo Uso da Categorização de Textos em Atividades Empresariais*. In: International Symposium on Knowledge Management/Document Management (ISKM/DM 2000), III, PUC-PR, Curitiba/PR, p.251-268.
- Sakkis, G.; Androutsopoulos, I.; Paliouras, G.; Karkaletsis, V.; Spyropoulos, C. D.; Stamatopoulos, P. (2003), *A memory-based approach to anti-spam filtering for mailing list*. Information Retrieval, v. 6, p. 49-73.
- Web, Agência (2003), *Pesquisa detecta preferência do e-mail na comunicação corporativa*, Página da Internet, Disponível em: <<http://www.agenciaweb.com.br/noticias.asp?pldMateria=622>>.
- Yang, Y.; Pedersen, J. O. (1997), *A comparative study of feature selection in text categorization*, Proceedings of the International Conference on Machine Learning.