

# PRÉ-PROCESSAMENTO DE DADOS DE BAIXA ESTATÍSTICA E SUA APLICAÇÃO NA PREVISÃO POR REDES NEURAIS DA ATENUAÇÃO POR CHUVAS EM ENLACES DE COMUNICAÇÃO VIA SATÉLITE

<sup>1</sup>Gilson A. Alencar e <sup>2</sup>L. P. Calôba

<sup>1</sup> UNIVERSIDADE GAMA FILHO

Dept. Eng. Elétrica - Rua Manuel Vitorino, 625 - Piedade - Rio de Janeiro  
gilsonalencar@aol.com

<sup>2</sup> COPPE/EE/UFRJ, C.P. 68564, Rio de Janeiro 21945-970  
caloba@lps.ufrj.br

## SUMÁRIO

A rede neural artificial implementada para realizar a previsão de atenuação por chuvas em enlaces de comunicação via satélite tem mostrado resultados bastante promissores. Após uma análise comparativa com alguns modelos fenomenológicos, verificou-se que a rede neural tem capacidade de estimar a atenuação com razoável precisão. No entanto, o pequeno número de experimentos disponíveis dificulta a obtenção de uma distribuição adequada dos dados entre os conjuntos de treinamento e teste, passo fundamental para garantir um processo de aprendizagem da rede neural preciso e confiável. Para contornar esta dificuldade, este trabalho apresenta um processo de agrupamento de base estatística que permite obter tais conjuntos de forma adequada para um treinamento eficaz. Após a implementação deste processo, os problemas inerentes à distribuição de dados deixaram de ser observados, o treinamento da rede tornou-se mais estável e os resultados muito mais precisos e confiáveis.

## 1. INTRODUÇÃO

A rede neural utilizada é constituída por seis unidades na camada de entrada, quinze neurônios (unidades de processamento) não-lineares na camada intermediária e um neurônio linear na camada de saída. A rede é do tipo “feedforward” com processo de aprendizagem utilizando o algoritmo “back-propagation” modificado super SAB, “speeding up adaptative back-propagation”. Foram realizados diversos experimentos com a rede neural e, em função dos resultados, todos os seus parâmetros foram devidamente ajustados. Para estimar a atenuação por chuvas foram treinadas, individualmente, redes neurais para diferentes valores de percentagem de tempo.

Para o processo de aprendizagem, os dados disponíveis (pares entrada-saída) foram divididos em dois conjuntos: um para treinamento e outro para teste. O conjunto de treinamento (dados conhecidos da rede) foi utilizado para ajustar as sinapses da rede, de forma que a mesma aprendesse o mapeamento não-linear da função que representa o problema da atenuação por chuvas. A função objetivo à ser minimizada durante o processo de aprendizagem corresponde ao erro relativo médio quadrático do conjunto de treinamento. Desta forma, o conjunto de treinamento necessita conter um número relativamente grande de pares entrada-saída de modo a garantir a boa representatividade estatística do problema em questão. O conjunto de teste (dados desconhecidos pela rede durante o treinamento) foi utilizado para avaliar o desempenho da rede neural e estabelecer o instante de parada do processo de aprendizagem. O critério de parada do treinamento é o instante em que a rede neural atinge a melhor generalização do problema que está sendo mapeado. É usualmente aceito na literatura que a melhor generalização é atingida quando o erro relativo médio quadrático medido sobre o conjunto de teste atinge seu mínimo. Geralmente, ao longo do treinamento, o erro relativo médio quadrático do conjunto de treinamento decai continuamente, tendendo assintoticamente para um mínimo, enquanto o do conjunto de teste decai, passa por um mínimo, e volta a crescer. O instante do treinamento em que este mínimo ocorre é em princípio o ponto de melhor generalização da rede. Por isto, o conjunto de teste também necessita ser estatisticamente representativo do universo que esta sendo modelado.

Selecionando-se randomicamente os pares entrada-saída disponíveis entre os conjuntos de treinamento e teste, os melhores resultados haviam sido obtidos com uma proporção de 75% do total dos pares para o conjunto de

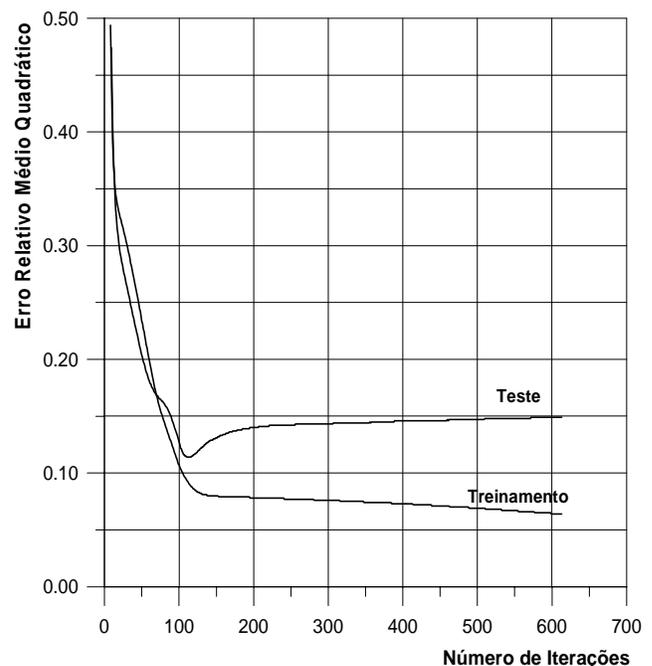
treinamento e 25% para o conjunto de teste. No entanto, como não há muitos dados, o conjunto de teste acaba recebendo um número ainda muito pequeno de pares entrada-saída. Após a realização de inúmeros experimentos, verificou-se que este pequeno número de dados no conjunto de teste influenciava significativamente no processo de aprendizagem da rede. Com um número muito pequeno de pares entrada-saída, a baixa representatividade estatística do conjunto de teste ocasiona uma certa instabilidade e imprevisibilidade no processo de aprendizagem. E mudar a proporção de pares entre os dois conjuntos apenas piorava a situação, devido a redução da representatividade do crítico conjunto de treinamento. Daí, surgiu a necessidade de se desenvolver um processo não completamente randômico de distribuição de dados de forma que o conjunto de teste e, também o conjunto de treinamento, pudessem ter melhor representatividade estatística do problema. Este trabalho descreve o método proposto e os resultados obtidos.

## 2. PROCESSO DE DISTRIBUIÇÃO

O processo de distribuição de dados que vinha sendo utilizado antes do que está sendo proposto neste trabalho era inteiramente aleatório. O conjunto total de pares entrada-saída para uma determinada percentagem de tempo era misturado aleatoriamente e dividido na proporção mencionada no item anterior. No entanto, foi observado que diferentes conjuntos de treinamento e teste, obtidos aleatoriamente, produziam resultados muito distintos após o processo de aprendizagem. Verificou-se que, em determinados casos, o erro relativo médio quadrático do conjunto de teste crescia rapidamente após alcançar um valor mínimo em um curto período de treinamento [1]. Este comportamento foi definido como super-especialização (“overtraining”) prematura da rede, com o correspondente encerramento prematuro do treinamento. Após a ocorrência de um treinamento com super-especialização prematura, o erro médio e o desvio padrão determinados sobre o conjunto total de dados (treinamento mais teste) apresentam valores bem elevados: em determinados casos, o desvio padrão alcançou valores superiores a 35%. A Figura 1 ilustra a evolução de um processo de aprendizagem com super-especialização prematura. Nesta figura, é possível perceber que o erro relativo médio quadrático para o conjunto de teste cresce rapidamente após ter atingido um valor mínimo em um curto período de treinamento (neste caso, após apenas 100 iterações). Nota-se também que, a curva relativa ao conjunto de treinamento esta ainda longe de atingir a estabilização quando o mínimo para o conjunto de teste é alcançado. Isto significa que o erro relativo médio quadrático para o conjunto de treinamento ainda teria possibilidade de decrescer. Este comportamento ocorreu

em virtude da baixa representatividade estatística no conjunto de teste. Para identificar dados isolados dos agrupamentos com maior representatividade foi implementada uma rede neural para efetuar um processo de redução de dimensionalidade de dados conhecida como análise por componentes principais, PCA [2]. Com este processo é possível representar graficamente através de um ponto em um espaço bidimensional - um plano - os parâmetros operacionais (frequência, polarização, altitude da estação, latitude e elevação) e mensuráveis (taxa de precipitação e atenuação) de cada enlace, da forma mais precisa possível. Com os pares entrada-saída de cada enlace representados por um ponto em um plano, fica fácil a identificação de dados em regiões de baixa representatividade estatística, ou seja, dados isolados dos agrupamentos formados.

Uma vez identificado o problema relativo aos conjuntos de treinamento e teste, foi desenvolvido um processo de base estatística para a distribuição adequada dos dados. Este processo avalia a estrutura estatística do conjunto de dados disponíveis para uma determinada percentagem de tempo e identifica aqueles que, efetivamente, fazem parte do grupo de maior representatividade.



**Figura 1:** Evolução do erro relativo médio quadrático para os conjuntos de treinamento e teste de uma rede neural treinada para 0,01% do tempo.

O processo de distribuição de dados ocorre em duas fases: Na primeira fase é executado um algoritmo de agrupamento ou “clustering” que tem como função

separar o conjunto total de dados em agrupamentos ou classes. Este processo de agrupamento é parte de um conjunto mais geral de técnicas comumente referidas como aprendizagem não-supervisionada. As técnicas de aprendizagem não-supervisionada estão relacionadas com o problema de formação de classes a partir de dados de treinamento sem o benefício de uma supervisão específica. O algoritmo implementado para executar a primeira fase do processo de seleção de dados efetua a formação de classes identificadas através de seus respectivos centróides. Este processo é iniciado com a determinação do centróide relativo a toda a população de dados. A partir deste centróide são obtidos novos centróides através de uma operação de divisões sucessivas ou “splitting”. Após a determinação do primeiro centróide, obtido a partir da média das componentes de todos os vetores de dados, são obtidos novos centróides por meio de uma operação de partição e deslocamento do centróide inicial. Os vetores representativos dos novos centróides são obtidos um pelo acréscimo e o outro pelo decréscimo de um pequeno valor (0,01) em cada componente do vetor do centróide inicial. Uma vez obtido os novos centróides, é feita a re-alocação dos vetores de treinamento às novas classes, baseado na medida da distância entre cada vetor de treinamento e os centróides. Cada vetor é alocado à classe cujo centróide apresenta a menor distância ao vetor. A partir daí, os centróides das classes criadas na operação anterior são recalculados e um novo processo de divisão é iniciado. Até este ponto o processo é uma alternância entre os processos “splitting” e “K-means” [3].

No algoritmo implementado para realizar o processo de agrupamento foi adotado um critério para a divisão das classes. Segundo este critério, uma classe só pode ser dividida se o número de vetores a ela associado ultrapassar um certo valor. Para o processo de agrupamento foi estipulado que cada classe deve conter, no máximo, seis vetores. Deste modo, a quantidade de classes que podem ser formadas pelo referido processo depende do número de vetores de treinamento e do número máximo de vetores estipulado para cada classe.

Na segunda fase, é executado o algoritmo de otimização das classes geradas pelo processo de agrupamento. Este algoritmo efetua a minimização da dispersão total nas classes. A dispersão de cada classe é obtida pelo somatório das distâncias entre os vetores da classe e o respectivo centróide. O processo de minimização da dispersão total consiste em transferir um dado vetor pertencente a uma classe  $i$  para uma classe  $j$ , se o decréscimo na dispersão na classe  $i$  for maior que o acréscimo na dispersão na classe  $j$ . Após a conclusão da segunda fase, as classes estarão totalmente otimizadas com relação às distâncias entre os vetores de treinamento

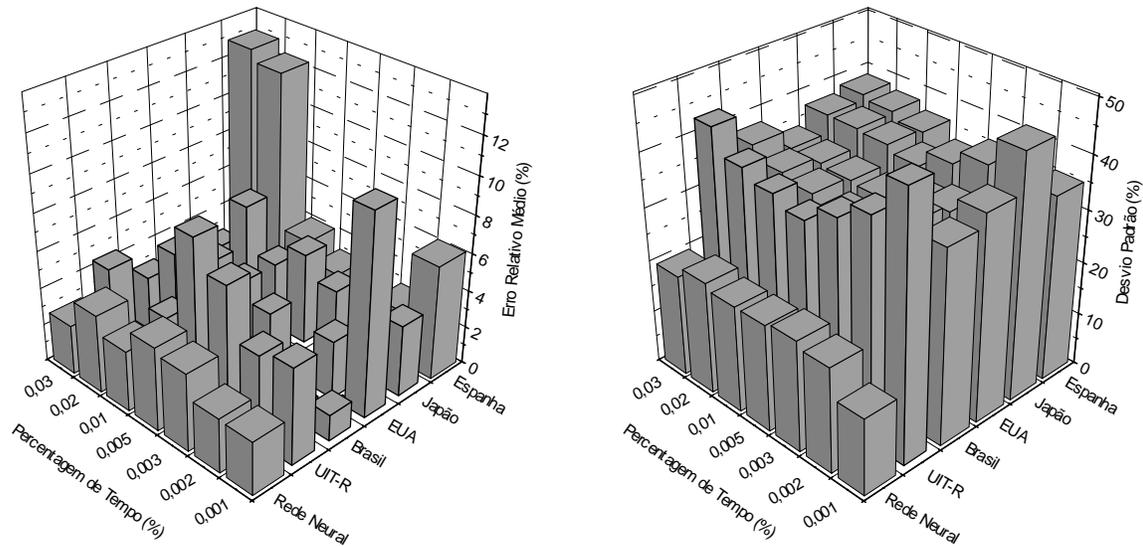
e os respectivos centróides. O processo final de distribuição de dados para os conjunto de treinamento e teste é realizado com base em um histograma das distâncias entre cada vetor e o centróide de sua respectiva classe. Com este histograma, é possível identificar e eliminar dados que fazem parte das classes de menor representatividade estatística.

### 3. RESULTADOS

Uma vez implementado o processo de distribuição de dados descrito no item anterior, a ocorrência de super-especialização prematura durante a fase de treinamento da rede não foi mais observada. Para avaliar o desempenho da rede foi realizada uma análise comparativa com cinco modelos fenomenológicos [4]: UIT-R, Americano, Brasileiro, Japonês e Espanhol. Esta análise se desenvolveu com base no erro relativo médio e no desvio padrão obtidos para cada modelo e pela rede neural treinada com dados selecionados pelo método de distribuição proposto neste trabalho. O erro relativo médio e o desvio padrão foram determinados com relação a todos os pares entrada-saída (conjunto de treinamento mais o conjunto de teste). A representação gráfica dos resultados obtidos está ilustrada na Figura 2 e os resultados numéricos correspondentes podem ser visualizados na Tabela 1. Como pode ser observado na Figura 2, o desvio padrão obtido para a rede neural ficou em torno de 20%, o que representa um excelente resultado se comparado com os valores obtidos pelos modelos fenomenológicos.

### 4. CONCLUSÕES

Os resultados apresentados pela rede neural após a implementação do processo de organização de dados proposto neste trabalho foram bastante satisfatórios. Destaca-se, neste caso, o desvio padrão que, para a rede neural, tem se mantido próximo dos 20 %, enquanto para os modelos fenomenológicos, varia entre os 30 e 40 %. Em análises anteriores à implementação do processo de distribuição de dados, foi observado que o valor absoluto do erro relativo médio apresentava grandes variações de um modelo para outro. Quando a rede neural passou a ser treinada com conjuntos de treinamento e teste obtido a partir do processo de distribuição de dados, o valor absoluto do erro relativo médio se mostrou mais uniforme. A não uniformidade do erro relativo médio observada em análises anteriores pode ser atribuída à baixa representatividade estatística dos conjuntos de treinamento e teste obtidos naquela ocasião. O processo de distribuição de dados tem se mostrado eficiente e, a ocorrência de “overtraining” prematuro, não tem sido mais observada após a sua implementação.



**Figura 2:** Valor absoluto do erro relativo médio e o desvio padrão obtidos com relação a todos os dados disponíveis para cada porcentagem de tempo.

**Tabela 1:** Erro relativo médio (erm) e desvio padrão (dp).

(%)	UIT-R		Japão		EUA		Brasil		Espanha		Rede	
	erm	dp	erm	dp	erm	dp	erm	dp	erm	dp	erm	dp
<b>0,001</b>	5,1	49,26	3,8	46,09	10,7	38,89	-1,4	36,53	6,1	35,14	-2,83	14,54
<b>0,002</b>	-4,7	41,68	-2,4	41,06	3,1	36,12	0,6	37,43	2,8	33,78	-2,89	19,88
<b>0,003</b>	-7,3	38,26	-3,8	38,51	0,2	34,86	-0,9	32,74	0	32,21	-4,13	21,07
<b>0,005</b>	-8,8	34,79	-4,9	35,41	-2,7	32,77	-1,7	32,39	-3,0	32,51	-4,46	20,34
<b>0,01</b>	-3,7	36,60	-3,5	36,71	-3,5	33,14	-1,3	33,94	-4,0	36,02	-3,23	20,33
<b>0,02</b>	-0,5	38,49	-5,9	37,07	-3,8	33,63	-5,2	34,99	-12,3	37,47	-4,16	21,37
<b>0,03</b>	4,5	39,18	-2,8	37,13	-2,1	33,79	-3,0	37,40	-12,9	38,27	-2,64	19,32

## 5. REFERÊNCIAS

[1] G. A. Alencar, L. P. Calôba and M. S. Assis, “Rain attenuation Prediction in Earth-Space Paths Performed by Artificial Neural Networks”, CLIMPARA, Canadá, Abril, 1998.

[2] Jolliffe, I.T., “Principal Component Analysis”, New York: Springer-Verlag, 1986.

[3] John R. Deller, John G. Proakis, John H. L. Hansen, “Discret-Time Processing of Speech Signals”, Macmillan Publishing Company, New York, 1993.

[4] G. A. Alencar, L. P. Calôba and M. S. Assis, “Performance Investigation of an Artificial Neural Network Designed to Predict the Rain Attenuation in

Earth-Space Paths”, Wave Propagation and Remote Sensing, Commission F Trienal Open Symposium, Aveiro, Portugal, Setembro, 1998.