

PS-1084

TIME SERIES PREDICTION USING ARTIFICIAL NEURAL NETWORKS

Augusto Cesar Espíndola Baffa (Centro Universitário da Cidade do Rio de Janeiro, Rio de Janeiro, RJ) – abaffa@bol.com.br

Ronaldo R. Goldschmidt (Centro Universitário da Cidade do Rio de Janeiro, Rio de Janeiro, RJ) – rrg@univercidade.edu.br

Jorge de Abreu Soares (Centro Universitário da Cidade do Rio de Janeiro, Rio de Janeiro, RJ) – jorge@univercidade.edu.br

Claudia Ferlin (Centro Universitário da Cidade do Rio de Janeiro, Rio de Janeiro, RJ) – cferlin@univercidade.edu.br

Given speed and competitiveness of business processes and contemporary globalized world, there is an huge demand for reliable and agile computing resources that help man in decision-making activities. Planning of production, estimates of demand for power and price of shares in the stock exchange are examples of decision-making processes that use the so-called time series forecast. Some searches in estimates of time series are assessing the suitability of various algorithms based on different paradigms (Nogueira, 2005) (Wang and Mendel, 1992) (Omar et al, 2007). This paper seeks to complement such searches describing the use of algorithms based on artificial neural networks in the task of time series forecast. Thus, experiments were performed using Neural Networks MLP Back-Propagation, RBF and combinations of these networks in ensemble machines. The results were analyzed and compared with those obtained directly from (Omar et al, 2007), demonstrating a good performance of the connectionist approach.

Keywords: neural networks, time series, prediction, forecasting

PREVISÃO DE SÉRIES TEMPORAIS UTILIZANDO REDES NEURAS ARTIFICIAIS

Diante da rapidez e da competitividade dos processos e negócios do mundo globalizado contemporâneo, existe uma grande demanda por recursos computacionais confiáveis e ágeis que auxiliem o homem em atividades de tomada de decisão. Planejamento da produção, previsão de demanda de energia elétrica e estimativa da cotação de ações na bolsa de valores são exemplos de processos de tomada de decisão que utilizam a chamada tarefa de previsão de séries temporais. Algumas pesquisas em previsão de séries temporais vêm avaliando a adequação de diversos algoritmos baseados em diferentes paradigmas (Nogueira, 2005) (Wang e Mendel, 1992) (Omar et al, 2007). O presente trabalho procura complementar tais pesquisas descrevendo a utilização de algoritmos baseados redes neurais artificiais na tarefa de previsão de séries temporais. Para tanto, foram realizados experimentos utilizando Redes Neurais MLP Back-Propagation, RBF e combinações destas redes em comitês de aprendizado. Os resultados foram analisados e comparados diretamente com os obtidos em (Omar et al, 2007), mostrando um bom desempenho da abordagem conexionista.

Palavra-Chave: redes neurais, séries temporais, previsão

1 Introdução

Prever o futuro é uma tarefa complexa, porém necessária em diversos setores da economia. Ser capaz de estimar corretamente acontecimentos futuros pode determinar o sucesso de um projeto ou de uma organização inteira. A previsão de dados possui inúmeras aplicações, dentre as quais podemos citar estimativas de preços das ações na bolsa de valores, processos estratégicos de produção industrial, demandas de energia elétrica, previsão de temperaturas médias, altura das marés, além de prevenir comportamentos sazonais futuros.

Como a previsão de valores futuros é baseada nos eventos passados, é necessário que se disponha de uma memória histórica de dados ocorridos anteriormente em um determinado contexto. Esta base de dados distribuídos no tempo é denominada série temporal.

As séries temporais podem representar dados históricos da macroeconomia (crescimento econômico, taxas de inflação, etc), finanças (previsão de evoluções de mercados financeiros, investimentos, etc), gestão empresarial (procura de produtos, consumo, etc), gestão pública (previsões de tráfego em pontos ou estradas, etc) e as áreas científicas (meteorologia, etc) (Dumont e Ribeiro, 2006).

Como as séries temporais representam dados passados, é necessária a utilização de algoritmos, técnicas ou métodos de previsão, que podem envolver cálculos simples ou procedimentos complexos.

O problema de previsão de séries temporais consiste em definir, a partir de dados passados, o próximo ponto da série através de modelos não lineares. No entanto, a criação de um algoritmo que resolva esta questão pode apresentar problemas, como a dependência do domínio, de acordo com a abordagem utilizada. A dependência do domínio faz com que uma determinada abordagem possua bons resultados para um certo tipo de problema, mas não necessariamente resulte em bons resultados para outros problemas.

Diversas pesquisas em previsão de séries temporais vêm avaliando a adequação de algoritmos baseados no paradigma da Lógica Nebulosa (Vale, 2001) (Tanscheit et al, 2005) (Wang e Mendel, 1992) (Omar et al, 2007), comparando-os a algoritmos clássicos como Média Móvel e Suavização Exponencial. O presente trabalho procura complementar os demais descrevendo a utilização de algoritmos baseados redes neurais artificiais na tarefa de previsão de séries temporais. Para tanto, foram realizados experimentos utilizando Redes Neurais MLP Back-Propagation, RBF e combinações destas redes em comitês de aprendizado. Os resultados foram analisados e comparados diretamente com os obtidos em (Omar et al, 2007), mostrando um bom desempenho dos algoritmos conexionistas. A facilidade de adaptação a novos domínios, a capacidade de aprender a partir de dados históricos e de lidar com problemas não lineares foram os principais fatores que motivaram a escolha das redes neurais artificiais nesse estudo comparativo.

Este texto está estruturado em mais cinco seções. A segunda seção apresenta as bases teóricas em redes neurais artificiais e comitês de aprendizado para elaboração do trabalho. A abordagem proposta encontra-se descrita na seção três. Na quarta seção, são apresentados detalhes da implementação da abordagem proposta, os testes e os resultados obtidos. A quinta e última seção comenta os resultados obtidos e apresenta perspectivas de trabalhos futuros.

2 Fundamentação Teórica

A presente seção fornece uma pequena introdução às áreas de previsão de séries temporais, descrevendo alguns de seus algoritmos clássicos, e de redes neurais artificiais, apresentando as arquiteturas MultiLayer Perceptron (MLP) e Radial Basis Function (RBF), além do conceito de comitês de aprendizado.

2.1 Séries Temporais

De acordo com a definição estatística, uma série temporal é uma coleção de dados sequenciais obtidos ao longo do tempo (Ehlers, 1999). A ordem das observações pode ser irrelevante para uma regressão linear, mas é fundamental para a análise de séries temporais. Seu objetivo consiste em identificar alguma dependência entre as T observações vizinhas de uma série $Z_T = (Z_1, Z_2, \dots, Z_T)$ e construir um modelo matemático a partir do qual seja possível prever os valores futuros para a série.

É necessário definir não só o horizonte de previsão mas também a janela a ser utilizada. O primeiro relaciona-se à quantidade de valores subseqüentes que serão previstos; a segunda regulariza a quantidade de elementos anteriores ao que será previsto (Ehlers, 1999). O horizonte de previsão pode ser classificado como de curto, médio e longo prazo. Quanto maior o horizonte de previsão maior a chance de erros, logo, previsões de curto prazo possibilitam uma boa visão do futuro próximo com um baixo erro.

Algumas séries são denominadas estacionárias, ou seja, ao longo do tempo, a série mantém-se ao redor de uma média constante. Outras apresentam tendências, podendo ter uma inclinação positiva ou negativa, chamada de tendência linear.

Entre alguns métodos de previsão de séries temporais podem ser citados a Média Móvel Simples, a Média Móvel Ponderada, a Média Móvel Exponencial, e a Previsão de Séries Temporais sujeitas a fenômenos sazonais (Nogueira, 2005). A seguir, encontram-se detalhados os métodos utilizados no estudo comparativo do presente projeto.

2.1.1 Média Móvel Simples

A média móvel simples (*Simple Moving Average - SMA*) é a média aritmética simples dos n valores anteriores, onde n é a janela de previsão. Sua fórmula é:

$$MM_{t+1} = \frac{\sum_{i=t-n+1}^t x_i}{n},$$

onde n é o tamanho da janela de previsão, t é o ponto da série (posição) e x os valores da variável de previsão.

2.1.2 Média Móvel Ponderada

A média móvel ponderada (*Weighted moving average - WMA*) é calculada aplicando-se fatores aos valores da janela de previsão. Utilizam-se pesos maiores para os valores mais recentes por serem mais significativos à previsão.

Assim, a expressão para cálculo da média móvel ponderada dá-se da seguinte

forma:
$$MMP_{t+1} = \frac{\sum_{i=t-n+1}^t x_i w_i}{n}$$
, onde n é o tamanho da janela de previsão, t é o ponto da série (posição) e w o fator (peso) a ser multiplicado à variável x , o valor a ser somado.

2.1.3 Média Móvel Exponencial

Na média móvel exponencial (*Exponential moving average* – EMA), também conhecida como suavização exponencial, a determinação do valor previsto é definida a partir da fórmula $VP_{t+1} = P_t + \alpha(R_t - P_t)$ onde, α ($0 < \alpha < 1$) é uma constante de suavização, VP_{t+1} é o valor a ser previsto, P_t é a previsão de valor do elemento anterior e X_t é o valor real do elemento anterior.

A constante α define a influência dos dados da série. Quanto maior for seu valor, mais significativos serão os dados mais recentes e quanto menor, mais significativo serão os dados mais antigos.

Define-se a previsão por suavização exponencial como uma soma ponderada da última observação X_t e da previsão VP_{t+1} (Nogueira, 2005). Devido à relação recursiva entre VP_{t+1} e X_t , VP_{t+1} , este método pode ser representado pela fórmula abaixo que fornece maior peso para o valor X_t , decrescendo o peso para as observações anteriores:

$$VP_{t+1} = \alpha x_t + \alpha(1 - \alpha)x_{t-1} + \alpha(1 - \alpha)^2 x_{t-2} + \dots$$

É possível obter uma medida de eficiência deste método considerando que o processo é completamente estável, e que assim X_1, X_2, \dots, X_n são variáveis randômicas independentes e identicamente distribuídas com variância σ^2 (Nogueira, 2005). Portanto é conveniente descrevê-la através da expressão: $VP_{t+1} \approx \frac{\alpha\sigma^2}{2 - \alpha} = \frac{\sigma^2}{(2 - \alpha)/\alpha}$

Podemos considerar o método de suavização exponencial como filtro estatístico cujas entradas são os dados de um processo estocástico e as saídas são estimativas suavizadas de uma média que varia no decorrer do tempo (Nogueira, 2005).

Maiores detalhes sobre previsão de séries temporais podem ser obtidos em (Statsoft, 2004), (Morettin e Toloi, 2004) e (Souza, 1989).

2.2 Redes Neurais

Redes Neurais são modelos matemáticos inspirados em neurônios biológicos, propostas por McCulloch e Pitts em 1943 que procuram simular o funcionamento do cérebro (Haykin, 1999). Os neurônios biológicos emitem um impulso elétrico quando a concentração de uma substância química alcança um certo nível no ambiente. Esta ação denomina-se ativação do neurônio.

Com base em McCulloch e Pitts, o modelo geral de um neurônio artificial, representado na figura 1, recebe valores de entrada combinados por meio do produto escalar de $w_i x_i$ (analogia matemática que simula a reação química), aplica a função φ , para produzir seu estado de ativação. Em seguida, aplica o valor resultante da etapa anterior à função η para produzir o valor de saída do neurônio.

A soma ponderada das entradas do neurônio k produz o valor efetivo de entrada

$$u: u_k = \sum_{i=1}^n w_i x_i .$$

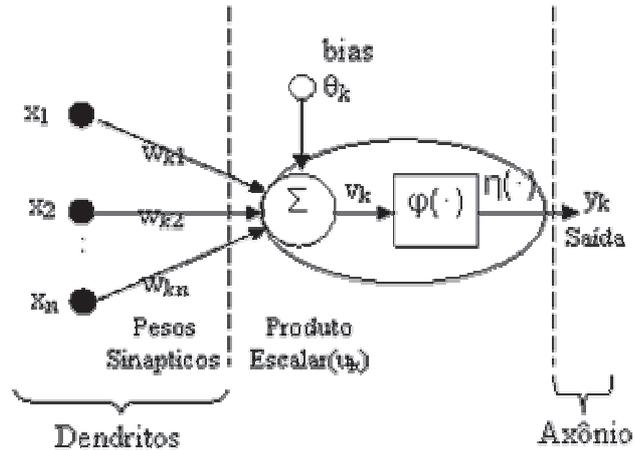


Figura 1 Modelo de McCulloch e Pitts

É dado o nome de confluência à combinação ponderada das entradas de um neurônio (Giacomini, 2003). O uso do valor auxiliar θ_k (bias – o nível de concentração da substância química ao ponto em que os neurônios são excitados) tem o efeito de ajustar a saída u_k do combinador linear, como demonstrado a seguir:

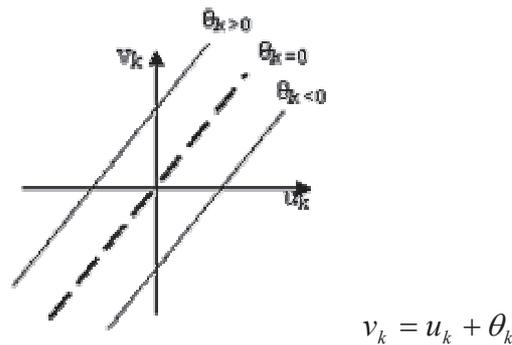


Figura 2 Ajuste de Bias

Na figura 1 é possível notar que as conexões sinápticas são consideradas externas no modelo do neurônio artificial assim como ocorre no sistema nervoso biológico. Desta

forma é possível representar a matriz de conexões sinápticas (matriz de conectividade) como a matriz de pesos de um grafo representativo da rede neural.

Aprendizado

As RNAs possuem a capacidade de aprender por exemplos e generalizar a partir do que aprenderam. O aprendizado é o ato de definir-se um comportamento a um estímulo externo baseado na informação adquirida de estímulos anteriores.

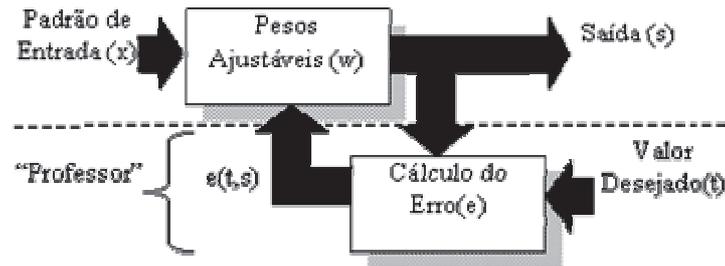


Figura 3 Aprendizado supervisionado

Função de Erro

Para avaliar a qualidade da saída obtida pela RNA em relação à saída correta esperada, deve-se comparar tais valores como demonstrado na figura 3. Para este propósito

utiliza-se a função de erro definida por $Q(w) = \sum_{i=1}^k e(x_i)$ onde $e(x_i) = \begin{cases} 1 & \text{se } y(x_i) - t_i = 0 \\ 0 & \text{se } y(x_i) - t_i \neq 0 \end{cases}$.

Regra Delta

A regra delta é baseada na lei de Hebb¹ e define que o valor de modificação dos pesos sinápticos deveria variar com o tempo, e que os neurônios vizinhos deveriam competir entre si, influenciando no valor de modificação. Pode ser expressa como: $\Delta w_{ij} = \eta x_i (o_j - w_{ij})$, onde Δw_{ij} é o acréscimo da intensidade da conexão entre os neurônios i e j , η o parâmetro definindo a intensidade da correção chamado taxa de aprendizado, x_i o estado de ativação do neurônio i , o_j saída do neurônio j , w_{ij} a intensidade da conexão entre os neurônios i e j e o_j a saída do neurônio j .

Para a tarefa de aprendizado, define-se um estado inicial randômico para os pesos w , um conjunto de treinamento $T = \{(x_i, t_i)\}, i = 1, \dots, k$, uma taxa de treinamento η e um contador $t = 0$ para controle das iterações.

¹ "O ajuste da intensidade da conexão entre os neurônios X e Y é proporcional ao valor da ativação simultânea dos dois neurônios. Caso o neurônio X tente excitar o neurônio Y e não consiga, a conexão deve ser enfraquecida". (Haykin, 1999)

Se o elemento i está previsto corretamente, então não se ajustam os pesos: $x_i \in C_1, w^T x_i \geq 0 \Rightarrow w_{(t+1)} = w_{(t)}$ e $x_i \in C_2, w^T x_i < 0 \Rightarrow w_{(t+1)} = w_{(t)}$. Se o elemento i não está previsto corretamente, são ajustados os pesos de acordo com $x_i \in C_1, w^T x_i < 0 \Rightarrow w_{(t+1)} = w_{(t)} + \eta_{(t)} x_i$ e $x_i \in C_2, w^T x_i \geq 0 \Rightarrow w_{(t+1)} = w_{(t)} - \eta_{(t)} x_i$ até que todos os i elementos sejam previstos corretamente ou alcançadas as t iterações (Haykin, 2003).

Retropropagação (Backpropagation) – Modelo de Rede Neural Artificial que consiste em comparar o resultado de saída gerado pelo neurônio com um exemplo de saída esperada, resultando em um valor de erro (Haykin, 1999). Em seguida, calcula-se o gradiente do erro em relação aos valores sinápticos do neurônio de saída. Calcula-se o erro da saída da penúltima camada e suas anteriores, propagando o valor de erro as camadas seguintes. Repete-se este processo até que o erro seja menor que o valor de tolerância desejado.

Funções de ativação - A função de ativação $\varphi(x)$ define o grau de excitação do neurônio a partir do valor induzido x . Existem várias formas de funções de ativação (Haykin, 1999):

a) Função linear: $\varphi(x) = ax + b$

b) Função sigmóide: $\varphi(v) = \frac{1}{1 + \exp(-av)}$

O parâmetro a define o valor de inclinação da função sigmóide. A aproximação deste parâmetro ao infinito $c \rightarrow \infty$ transforma a função sigmóide em uma função de limiar. A função sigmóide possui valores entre 0 e 1. Para situações onde são necessários valores entre -1 e 1, usamos a *função tangente hiperbólica*.

c) Função tangente hiperbólica: $\varphi(v) = \tanh\left(\frac{cv}{2}\right) = \frac{1 - e^{(-cv)}}{1 + e^{(-cv)}}$.

O parâmetro a define o valor de inclinação da função sigmóide. A aproximação deste parâmetro ao infinito $c \rightarrow \infty$ transforma a função sigmóide em uma função de limiar bipolar.

d) Função Gaussiana: $\varphi(v) = \exp(-v^2 / a)$.

O parâmetro a define o valor da largura da base da distribuição normal.

2.3 Arquiteturas de Rede Adotadas

No presente estudo, foram utilizadas duas arquiteturas de rede: uma rede MPL e uma rede RBF que foram baseadas nos trabalhos de Charalambous e Martzoukos(2001), Kutsurelis(1998) e Castro(2001). Outras arquiteturas podem ser utilizadas, talvez

alcançando melhores resultados que as sugeridas a seguir. Objetiva-se comparar seus resultados e indicar qual é a melhor solução para o problema de previsão.

Arquitetura Multi-Layer Perceptron (MLP) - As redes neurais onde os neurônios das camadas ocultas e de saída utilizam um tipo de função de ativação senoidal são chamadas de Multi Layer Perceptrons (Perceptrons de múltiplas camadas). $MPL : \mathcal{R}^n \rightarrow \mathcal{R}^m$, onde n é a quantidade de neurônios de entrada e m de saída.

Charalambous e Martzoukos(2001) apresentam uma arquitetura MLP convencional para a precificação de derivativos e opções. Seu trabalho demonstra a implementação de uma arquitetura que utiliza funções de ativação sigmóide e tan-sigmóide. De acordo com suas pesquisas, a função tan-sigmóide apresenta melhores resultados para análise e precificação de ativos financeiros em relação à função sigmóide.

Arquitetura RBF – As redes neurais podem ser classificadas como Radial Basis Function (RBF) quando a função de ativação $\varphi(x)$ usada nesta arquitetura possua a forma de uma função simétrica radial. Geralmente é usada a função gaussiana (ou alguma função quádrada). São chamadas de redes $RBF : \mathcal{R}^n \rightarrow \mathcal{R}^m$ quando possuem n neurônios como entrada, m neurônios como saída e, obrigatoriamente, r neurônios RBF na camada oculta.

Kutsurelis(1998) examinou as habilidades de uma rede neural na predição de ganhos futuros no mercado de ações americano. Em seus experimentos, Kutsurelis(1998) buscava a maior precisão em suas previsões ajustando um modelo de rede RBF. Baseado no livro de Edward Gately chamado *Neural Networks for Financial Forecasting*, seu modelo alcançou 93.3 % de probabilidade de prever uma alta e 88.07 % de prever uma baixa do índice S&P500.

2.4 Comitês de Aprendizado

O comitê de aprendizado (ou comitê de redes neurais) é uma combinação de redes neurais de tamanhos e/ou características diferentes que, trabalhando em paralelo, procuram melhorar o desempenho do mecanismo de previsão (Hofman, 2001). A resposta final de um problema é obtida a partir de um procedimento de combinação de resultados envolvendo os resultados gerados pelos membros do sistema.

Dentre as vantagens da utilização de comitês está a redução na variância do erro, podendo-se ponderar a contribuição de cada rede para a resposta final de um problema. Os comitês também podem ser utilizados para transformar a rede em um sistema modular mais fácil de manter e entender.

Os comitês baseiam-se no conceito de *dividir e conquistar*. Tarefas computacionalmente complexas podem ser divididas em tarefas menores e simples proporcionando o mesmo resultado.

A tarefa é dividida entre um número de *redes especialistas*, que divide o vetor de entrada em um conjunto de subespaços. A combinação de especialistas constitui uma *máquina de comitê*.(Haykin, 1999) A combinação de redes especialistas pode ser classificada como um exemplo de redes modulares.

As máquinas de comitê são aproximadores universais e podem ser classificadas em:

- Estruturas estáticas – Quando o resultado de várias redes especialistas é combinado sem envolver o sinal de entrada original, representado na figura 4.

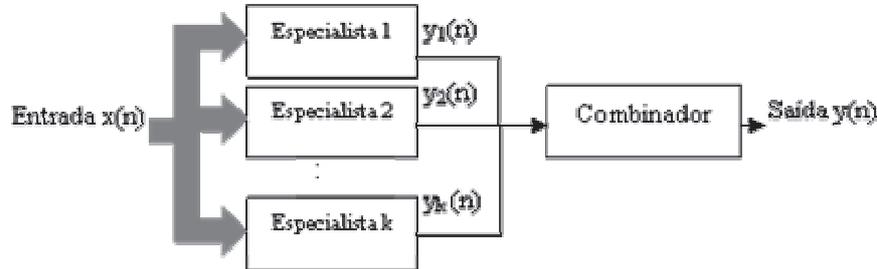


Figura 4 Diagrama de uma máquina de comitê utilizando estruturas estáticas

- Estruturas dinâmicas – Nesta categoria, o sinal de entrada atua e integra as saídas dos especialistas formando uma saída global, representado na figura 5.

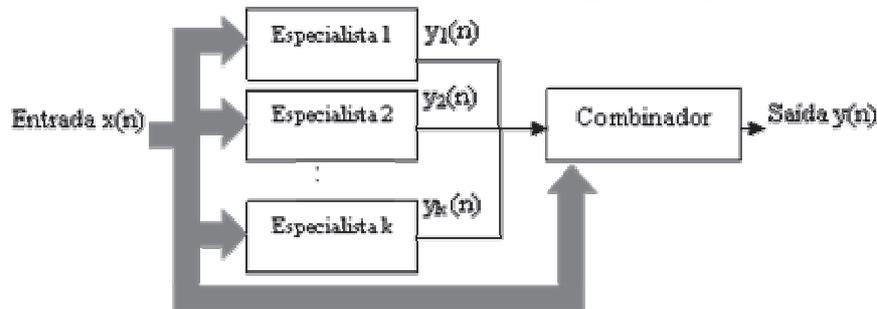


Figura 5 Diagrama de uma máquina de comitê utilizando estruturas dinâmicas

3. Abordagem Proposta

A abordagem conexionista proposta adotada no presente trabalho seguiu o modelo de comitê de aprendizado da figura 5. Como especialistas foram utilizadas redes neurais artificiais MLP e RBF. Como combinador foi adotada uma rede MLP.

Para as redes neurais adotadas, foi necessário definir os passos para o algoritmo de aprendizado. As camadas de entrada precisavam variar de acordo com o tamanho da janela a ser utilizada no aprendizado. As arquiteturas apresentadas definem valores padrões de entradas que foram utilizados como ponto de partida para a análise. Utilizando com exemplo a arquitetura MLP proposta, a janela deve possuir 13 valores de amostragem e seu próximo valor que será a saída. Assim, o conjunto de dados de entrada-saída pode ser definido como:

$$\left(x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, \dots, x_{13}^{(1)}; y^{(1)}\right), \left(x_1^{(2)}, x_2^{(2)}, x_3^{(2)}, \dots, x_{13}^{(2)}; y^{(2)}\right), \dots$$

Onde os valores X_1, X_2, \dots, X_n são as entradas e y sua respectiva saída. A abordagem proposta estipula que cada série a ser processada seja normalizada antes de ser submetida ao treinamento ou predição. Tal restrição segue uma recomendação disponível

em (Nogueira, 2005) que estabelece que cada série seja transformada em uma série estacionária para que a rede possa aprender a relação entre os dados sem haver distorções. A figura 6 representa um exemplo de série original com tendência positiva e a figura 7 a mesma série após a normalização.

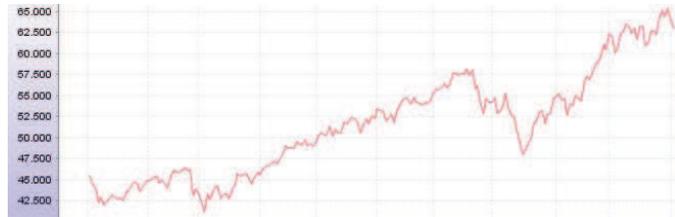


Figura 6 Série Índice Bovespa



Figura 7 Série normalizada do Índice Bovespa

Deve-se realizar o aprendizado (treinamento da rede) do conjunto de dados de entrada-saída várias vezes para que a rede possa se aproximar ao máximo da resolução do problema. Definem-se como épocas o número de vezes no qual a rede é submetida aos exemplos de aprendizado.

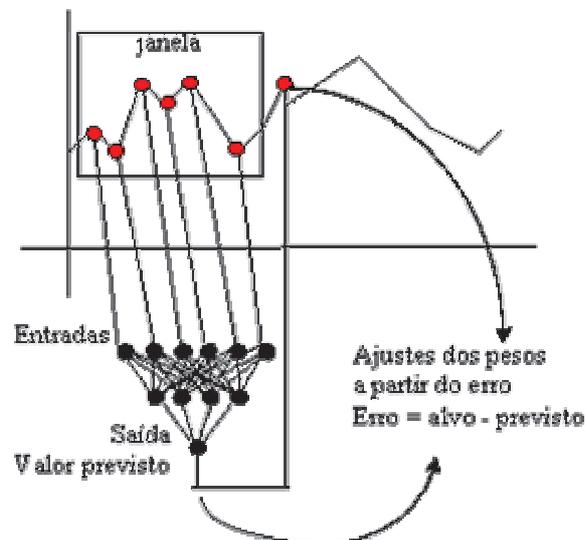


Figura 8 Aprendizado da janela da série temporal

Para cada n elementos do conjunto entrada-saída, verificar o erro em relação ao valor calculado e à saída alvo e ajustar os pesos dos neurônios, como demonstrado na figura 8. A equação da regra delta calcula o valor a ser acrescentado (ou subtraído) ao valor do peso sináptico: $\Delta w_{ij} = \eta x_i (o_j - w_{ij})$. Em seguida, realiza-se o mesmo cálculo de ajuste para os elementos seguintes do conjunto de entrada-saída.

Ao final do aprendizado a rede deve ser capaz de encontrar os próximos pontos em relação aos últimos valores da série, conforme representado na figura 9.

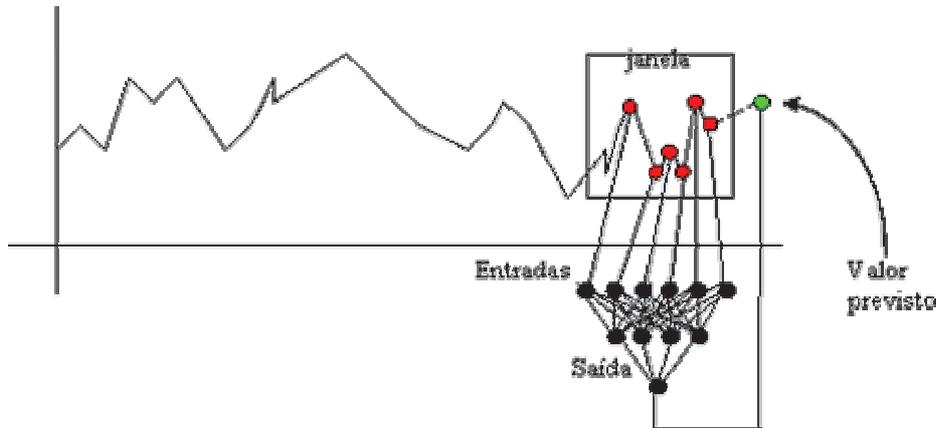


Figura 9 Previsão do próximo ponto da série

4 Protótipo, Experimentos e Resultados

Esta seção tem por objetivos fornecer detalhes sobre a implementação do método proposto, descrever a metodologia de teste adotada, além de apresentar os resultados obtidos e uma análise dos experimentos realizados.

4.1 Protótipo

A implementação do método de previsão de séries temporais foi desenvolvida na linguagem Java, utilizando banco de dados MySQL. O método encontra-se acoplado ao *framework JavaBramining* (Ribeiro et al, 2005), desenvolvido pelo Núcleo de Projetos e Pesquisas em Aplicações Computacionais do Centro Universitário da Cidade do Rio de Janeiro. O *JavaBramining* é uma ferramenta voltada ao processo operacional de descoberta de conhecimento e mineração de dados que incorpora diversos métodos aplicáveis nesse processo.

No método implementado existe a possibilidade de se escolher qual atributo do banco de dados deverá ser previsto, configurar a rede neural (camadas, quantidade de neurônios e funções de ativação), a taxa de aprendizado, *momentum*, número de épocas, o tamanho da janela e o intervalo de registros para treinamento. Para a previsão dos valores, são informados o intervalo de registros sobre o qual será realizada a previsão. As figuras 10 e 11 apresentam algumas telas do protótipo a título ilustrativo.

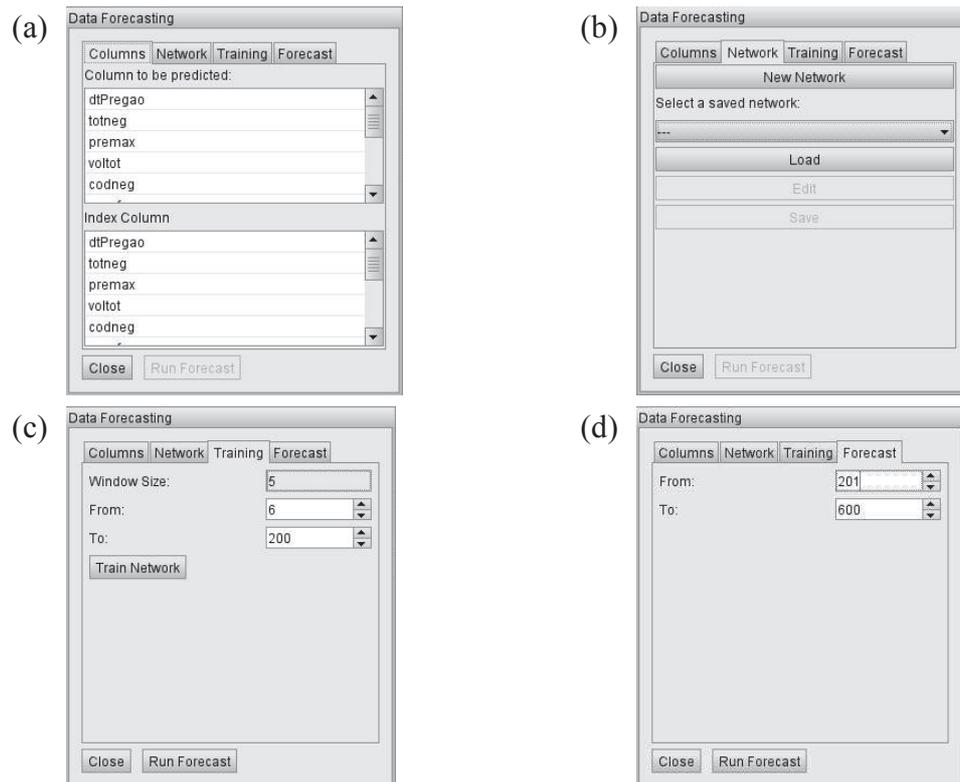
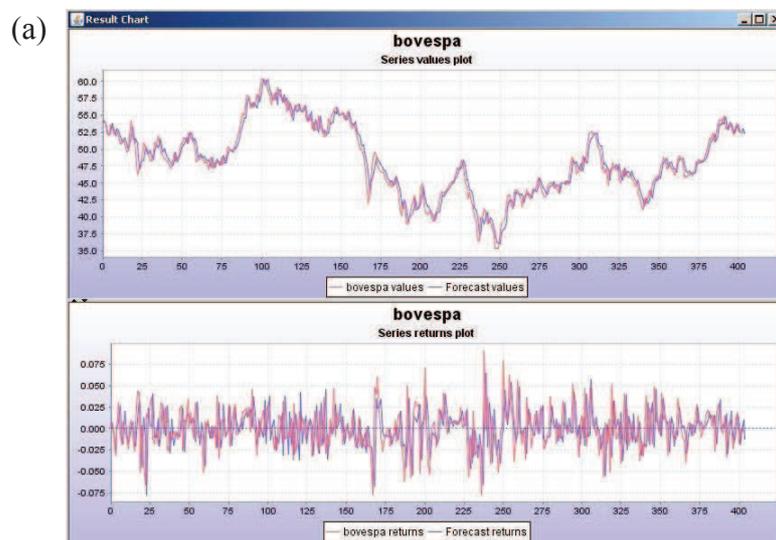


Figura 10 Telas do método de previsão de dados. (a) Configuração dos atributos de previsão e de ordem dos dados; (b) Configuração das redes neurais; (c) Configuração dos dados de treinamento; (d) Configuração dos dados de previsão



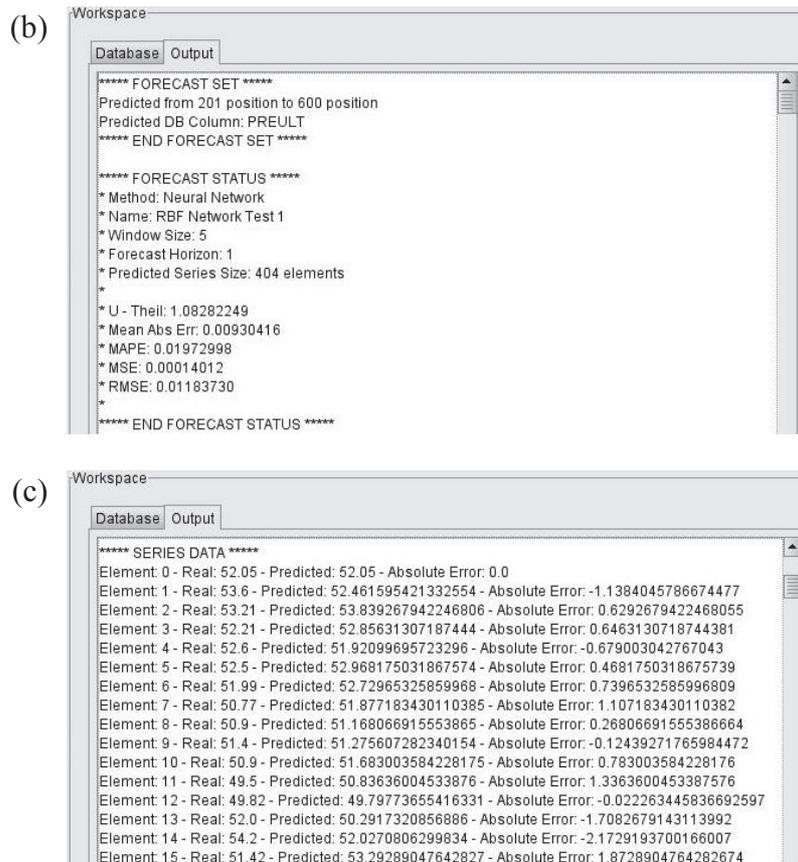


Figura 11 Telas do método de previsão de dados. (a) Gráfico superior representando a série temporal e a série prevista e gráfico inferior representando as série normalizada; (b) Resumo dos dados utilizados na previsão; (c) Resultado da previsão dos valores

4.2 Metodologia de Teste

De forma análoga a (Omar et al., 2007), os testes e o estudo comparativo foram feitos com cinco séries temporais diferentes, procurando variar a natureza (comercial ou científica), além do volume e da diversidade dos dados. A seguir encontra-se um resumo das séries adotadas nos experimentos:

- O preço de abertura dos valores das ações PETR4F da empresa Petrobras (COTAÇÕES HISTÓRICAS, 2006), representada na figura 12 como uma das representantes das bases de dados comerciais;
- Série anual de manchas solares (SUNSPOT DATA, 2006), representada na figura 13 e muito utilizada pela comunidade científica mundial;
- Série Mackey-Glass (WAN, 2006), representada na figura 14 e também bastante utilizada em trabalhos de pesquisa científica;
- Série Dow Jones (WESSA, 2006), representada na figura 15, também de caráter comercial, representando valores de ações nas bolsas de valores;

e) Série Laser (WAN, 2006), cujos primeiros três mil elementos estão representados na figura 16. Utilizada pela comunidade científica, foi adotada nos experimentos devido ao seu tamanho: mais de dez mil registros.

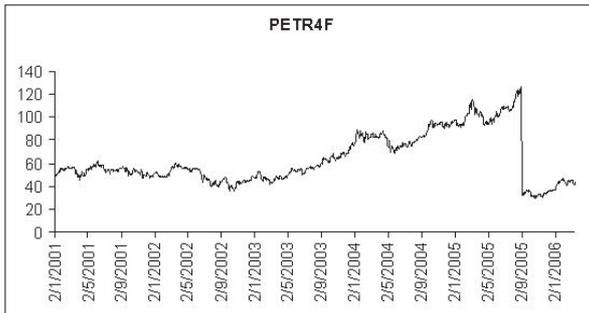


Figura 12 – Série da ação PETR4F

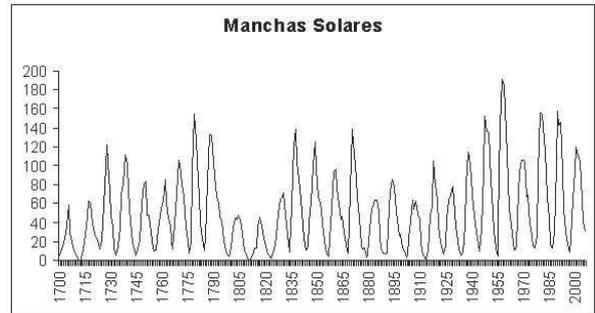


Figura 13 – Série Manchas Solares

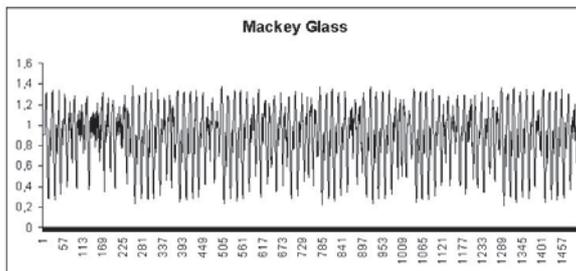


Figura 14 – Série Mackey-Glass

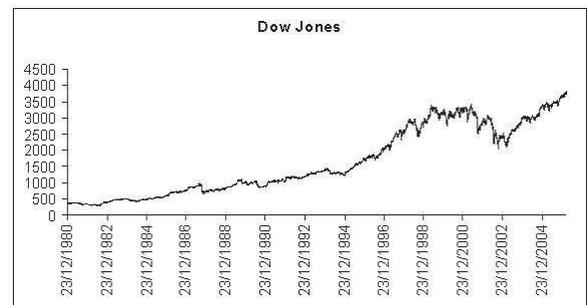


Figura 15 – Série Dow Jones

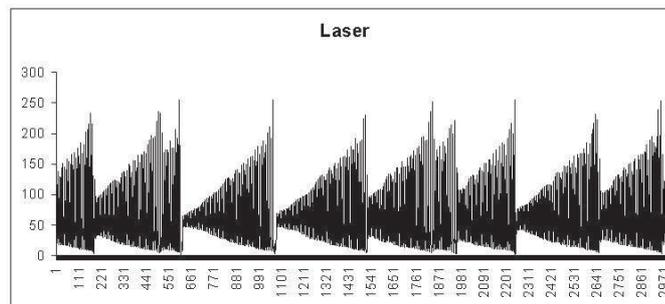


Figura 16 – Série Laser - 3000 elementos

Pode-se perceber uma nítida diferença entre as cinco séries escolhidas para os experimentos. Enquanto as séries **a** e **d** apresentam uma tendência linear positiva, as séries **b**, **c** e **e** apresentam uma certa sazonalidade.

A tabela 1 apresenta um resumo estatístico sobre as séries adotadas nos experimentos.

Série	Quantidade total de registros	Menor valor	Maior valor
PETR4F/Preabe	1293	30,1	126,9
Manchas Solares	306	0	190,2
Mackey-Glass	1500	0,212559300	1,378507200
Dow Jones	6369	299,21	3833,97
Laser	10093	0	255

Tabela 1 – Detalhes das Séries

A maneira encontrada para avaliar o desempenho dos algoritmos na previsão das séries foi a utilização do cálculo do coeficiente U-Theil (VEIGA, 2005). Esse coeficiente avalia o desempenho da previsão com relação à previsão trivial ou ingênua, já descrita anteriormente.

Índice de desigualdade de Theil
(Theil's inequality index)

$$UTheil_n = \frac{\sqrt{\sum_{k=1}^n (a_k - y_k)^2}}{\sqrt{\sum_{k=1}^n (a_k - a_{k-1})^2}}$$

a_k é o valor real da série na posição k e y_k é o valor previsto na posição k . Quanto mais próximo de zero for o valor encontrado, melhor desempenho obteve a previsão.

Para definir a melhor previsão, deve-se verificar o valor do $UTheil$ calculado. Para $UTheil > 1$, o erro da previsão é maior do que o da previsão ingênua. Para $UTheil < 1$, o erro da previsão é menor do que o da previsão ingênua.

Portanto, quanto mais próximo de zero for o valor de $UTheil$, melhor desempenho obteve o algoritmo responsável pela previsão.

Outro cálculo de desempenho adotado nos experimentos do presente trabalho é feito através do erro médio absoluto (MAE), representado pela fórmula:

Erro Médio Absoluto
(Mean Absolute Error)

$$MAE_n = \frac{\sum_{k=1}^n |a_k - y_k|}{n}$$

onde a_k é o valor real da série, y_k é o valor previsto da série e n é a quantidade de elementos a serem previstos.

Uma terceira medida de desempenho nos testes foi o erro percentual médio absoluto (MAPE), calculada através da fórmula:

$$\text{Erro Porcentual Médio Absoluto} \\ (\text{Mean Absolute Percentage Error}) \quad \text{MAPE}_n = \frac{\sum_{k=1}^n \left| \frac{a_k - y_k}{a_k} \right|}{n}$$

onde a_k é o valor real da série, y_k é o valor previsto da série e n é a quantidade de elementos a serem previstos.

Cada base de dados foi separada em dois conjuntos: treino e teste. O subconjunto de treino, que variou de 50% a 89,79% das bases de dados originais, foi utilizado para geração das bases de conhecimento das redes MLP e RBF e do comitê de previsão. O subconjunto de teste, que consistiu do restante dos dados originais, foi utilizado para avaliar o desempenho de cada um dos métodos aplicados.

A Rede MLP utilizada foi constituída de quatro camadas:

Entradas: Linear [-1,1] 13 neurônios;
 1ª Camada oculta: Tanh 16 neurônios;
 2ª Camada oculta: Tanh 16 neurônios;
 Saída: Tanh, 1 Neurônio.

A Rede RBF utilizada foi constituída de cinco camadas:

Entradas: Linear [-1,1] 13 neurônios;
 1ª Camada oculta: Gaussiana, 16 neurônios;
 2ª Camada oculta: Tanh, 16 neurônios;
 3ª Camada oculta: Gaussiana, 16 neurônios;
 Saída: Tanh, 1 Neurônio.

Convém ressaltar que a adoção das configurações acima foi resultado de diversos testes preliminares em busca de configurações que apresentassem bom desempenho.

O comitê de previsão utilizado neste trabalho está representado na figura 17. O método de previsão utilizando o comitê proporciona o aprendizado a partir da janela de previsão e dos resultados previstos pelas redes especialistas. Antes, é necessário proporcionar o aprendizado de cada rede especialista e realizar sua previsão. Em seguida, com os valores de cada rede especialista, é possível elaborar o aprendizado da rede comitê.

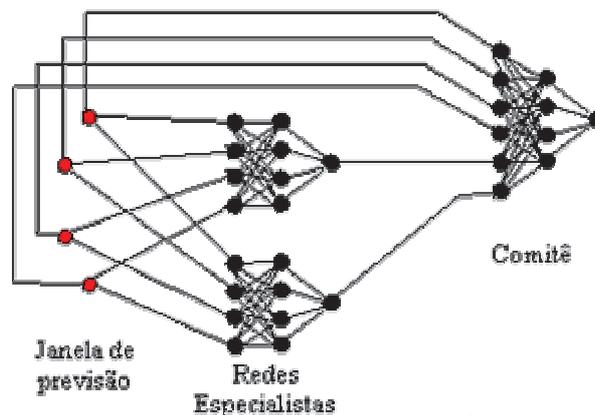


Figura 17 Previsão utilizando comitê de previsão

Calculam-se os resultados de cada especialista (rede neural ou método estatístico) para, em seguida, utilizá-los junto aos valores da janela de previsão como entrada da rede comitê.

A tabela 2 indica os parâmetros adotados para cada algoritmo nas respectivas bases de dados. A escolha desses valores foi decorrente de uma experimentação prévia dos métodos nas referidas séries temporais. Cabe ressaltar que tantos os parâmetros utilizados quanto os resultados obtidos com os demais algoritmos são os mesmos relatados em (Omar et al., 2007).

Base Método	Parâmetros	PETR4F	Manchas Solares	Mackey-Glass	Dow Jones	Laser
Wang-Mendel (Wang e Mendel, 1992)	Nº de Conjuntos	25	51	13	13	51
	Tamanho da janela	5	7	9	10	10
	Registros de Treino	89,79%	84,97%	50%	87,93%	89,17%
	Registros de Previsão	10,21%	15,03%	50%	12,07%	10,83%
Wang-Mendel Adaptado (Omar et al, 2007)	Nº de Conjuntos	25	51	13	13	51
	Tamanho da janela	5	7	9	10	10
	Registros de Treino	89,79%	84,97%	50%	87,93%	89,17%
	Registros de Previsão	10,21%	15,03%	50%	12,07%	10,83%
Média Móvel	Nº Pontos para média (n)	5	7	9	10	10
	Registros de Previsão	10,21%	15,03%	50%	12,07%	10,83%
Suavização Exponencial	α	0,9	0,9	0,9	0,9	0,9
	Registros de Previsão	10,21%	15,03%	50%	12,07%	10,83%
Rede MLP	Tamanho da janela	5	7	9	10	10
	Registros de Treino	89,79%	84,97%	50%	87,93%	89,17%
	Registros de Previsão	10,21%	15,03%	50%	12,07%	10,83%

Rede RBF	Tamanho da janela	5	7	9	10	10
	Registros de Treino	89,79%	84,97%	50%	87,93%	89,17%
	Registros de Previsão	10,21%	15,03%	50%	12,07%	10,83%
Comitê de Previsão	Tamanho da janela	5	7	9	10	10
	Registros de Treino	89,79%	84,97%	50%	87,93%	89,17%
	Registros de Previsão	10,21%	15,03%	50%	12,07%	10,83%

Tabela 2 – Parâmetros adotados nos Métodos de Previsão de Séries Temporais

4.3 Avaliação dos Resultados

A tabela 3 apresenta um resumo dos testes efetuados. Convém observar que em todos os casos, a rede RBF apresentou melhor desempenho do que a rede MLP (U-Theil menor, mais próximo de zero), e que o uso do comitê parece contribuir para a melhoria da precisão das redes especialistas.

Base \ Método	PETR4F		Manchas Solares		Mackey-Glass		Dow Jones		Laser	
	U-Theil	MAPE	U-Theil	MAPE	U-Theil	MAPE	U-Theil	MAPE	U-Theil	MAPE
Wang-Mendel	1,90	7,60	2,66	36,67	0,44	5,15	52,03	13,22	0,19	1,64
Wang-Mendel Adaptado	1,33	4,21	0,91	11,60	0,40	4,29	9,15	3,66	0,14	0,92
Média Móvel	1,47	1,59	1,93	19,16	1,87	23,19	1,40	0,74	1,06	14,0 3
Suavização Exponencial ($\alpha = 0,9$)	1,01	1,08	1,06	9,53	1,05	12,98	1,00	0,52	1,03	12,3 3
Rede Neural MLP	0,99	3,40	0,25	24,57	0,06	0,95	1,36	0,81	0,06	0,00 1
Rede Neural RBF	1,00	1,40	0,15	17,79	0,05	0,88	1,07	0,64	0,08	5,13
Comitê	1,36	8,80	0,08	10,69	0,06	1,04	1,12	0,68	0,07	0,00 1

Tabela 3 – Comparação entre os Métodos de Previsão de Séries Temporais

Pode-se perceber ainda que, na maioria dos casos, comparando com os métodos de previsão tradicionais de Média Móvel e Suavização Exponencial, tanto as Redes MLP e RBF quanto o comitê de previsão obtiveram melhor desempenho na previsão de séries que

apresentavam ciclos de repetição, e bons resultados em séries que não apresentavam tendência de repetição de comportamento.

As redes MPL demonstraram melhores desempenhos de aprendizado e previsão quando aplicadas a séries com repetições, que possuam movimentos senoidais, enquanto as redes RBF foram capazes de reproduzir o serrilhado das séries estocásticas.

Com exceção da série Dow Jones (que é uma série altamente estocástica), onde a rede RBF obteve um resultado próximo porém não superior ao da suavização exponencial, a previsão utilizando redes neurais superou o desempenho dos métodos tradicionais e dos algoritmos Wang-Mendel e Wang-Mendel adaptado.

O comitê de previsão obteve uma redução de eficiência na previsão de séries estocásticas devido a influência da rede MLP, se comparado aos resultados de cada rede individualmente, porém, a sua predição de séries cíclicas ou sazonais obteve resultados muito bons. Neste caso, quando não superou os resultados das outras redes, obteve valores muito próximos, com um erro médio percentual absoluto menor.

Em geral, os testes demonstraram que as redes neurais são capazes de obter excelentes resultados em séries estáveis, cíclicas ou com uma certa sazonalidade. Em séries estocásticas, como as de preços de ações na bolsa de valores, seus resultados foram próximos aos de métodos convencionais, podendo alcançar melhores resultados em outras tentativas de aprendizado

5 Conclusões

A proposta deste trabalho foi analisar o desempenho de redes neurais artificiais aplicadas ao problema de previsão de Séries Temporais e compará-lo com os resultados obtidos em (Omar et al, 2007).

Nos experimentos realizados, as redes neurais demonstraram bastante adaptabilidade a qualquer tipo de série temporal, sendo capazes de efetuar previsões com bom desempenho a partir do aprendizado baseado nos dados históricos das séries.

Também foi possível observar que, nos testes realizados, as redes MLP se adaptaram melhor a movimentos senoidais, como as curvas das séries cíclicas, e as redes RBF apresentaram melhores resultados em séries semelhantes a “ondas dente de serra”.

Como trabalhos futuros são sugeridas algumas melhorias no protótipo implementado. Por exemplo, o desenvolvimento de componentes de pré-processamento no ambiente auxiliaria na normalização dos dados e poderia ajustar a série temporal de acordo com técnicas utilizadas em previsão como a correção de séries com efeitos sazonais.

O processo de aprendizado dos comitês é demorado e, devido à complexidade da arquitetura do protótipo, pode demandar muitas horas de processamento. Analisar muitas séries e utilizar muitos especialistas em comitês pode demorar dias de processamento. Recomenda-se, portanto, a utilização de *frameworks* de computação em grade, pois a mesma permitiria a distribuição do processamento das redes entre diversos computadores.

Os resultados aqui apresentados, foram obtidos utilizando arquiteturas específicas que não necessariamente são as melhores para este problema. Os comitês de aprendizado podem ser formados por diversos métodos estatísticos e outras arquiteturas de redes neurais que não foram exploradas por este trabalho.

A experimentação com outras bases, de diversos domínios de conhecimento, é importante para o aperfeiçoamento do modelo. As arquiteturas de redes neurais adotadas podem obter melhores resultados de previsão em outras séries não exploradas no presente artigo. Estudos neste sentido também se mostram extremamente oportunos como trabalhos futuros.

Referências

BOJADZIEV, G., BOJADZIEV, M., Fuzzy Logic for Business, Finance, and Management. World Scientific, 1997.

CASTRO, Maria Cristina Felippetto de; Março 2001; Predição não-linear de séries temporais usando redes neurais RBF por decomposição em componentes principais

CHARALAMBOUS, Christakis; MARTZOUKOS, Spiros H.; May 2001; **Artificial Neural Networks for Valuation of Financial Derivatives and Customized Option Embedded Contracts**

BOVESPA, 2006, **Cotações Históricas**, Bolsa de Valores de São Paulo. Disponível em <http://www.bovespa.com.br/Mercado/RendaVariavel/SeriesHistoricas/FormSeriesHistoricas.asp>. Acesso em 03/05/2006.

COX, E., 1994, **The Fuzzy Systems Handbook**. AP Professional.

DUMONT, Danilo Mozeli; RIBEIRO, José Araújo; Rodrigues, Luiz Alberto; Ed. Revan, 2006; **Inteligência Pública na Era do Conhecimento**.

EHLERS, Ricardo S.; 1999; **Análise de Séries Temporais**, Dissertação de Graduação - Universidade Federal do Paraná.

GIACOMINI, Enzo; December 2003; **Neural Networks in Quantitative Finance**

GOLDSCHMIDT, R. R., PASSOS, E., 2005, **Data Mining: um Guia Prático**. 1ª Edição. Rio de Janeiro, Editora Campus.

HAYKIN, Simon; 1999; Neural Networks - A Comprehensive Foundation, Editora Bookman.

KUTSURELIS, Jason E.; September 1998; Forecasting Financial Markets Using Neural Networks: An Analysis Of Methods And Accuracy.

JÚDICE, P. A., 2005, **Previsão de Demanda na Gestão Logística de um Produto Perecível Vendido por Máquina Automática**. Tese de Mestrado. Disponível em http://www.maxwell.lambda.ele.puc-rio.br/cgi-bin/db2www/PRG_0651.D2W/SHOW?Cont=6719:pt&Mat=&Sys=&Nr=&Fun=&CdLinPrg=pt. PUC, Rio de Janeiro, RJ, Brasil. Acesso em 16/10/2005.

- MORETTIN, P. A., TOLOI, C. M. C., 2004, **Análise de Séries Temporais**. 1ª Edição. São Paulo, Editora Edgard Blücher.
- NOGUEIRA, F. M. de A., 2005, **Modelos de Previsão**. Juiz de Fora. Disponível em <http://www.engprod.ufjf.br/fernando/epd042/previs%E3o.pdf>. Acesso em 16/10/2005.
- PEDRYCZ, W., GOMIDE, F., **An Introduction to Fuzzy Sets Analysis and Design**. MIT, 1998.
- REZENDE, S. O., 2003, **Sistemas Inteligentes: Fundamentos e Aplicações**. São Paulo: Manole.
- RIBEIRO, R. C., GOLDSCHMIDT, R., FERLIN, C., SOARES, J. A., 2005, **Apresentação Técnica do JavaBramining**, Relatório Técnico, Centro Universitário do Rio de Janeiro, Rio de Janeiro, RJ, Brasil.
- SANDRI, S., CORREA C., 1999, **Lógica Nebulosa**, ITA, São José dos Campos, SP.
- SILVA, P. O. M. P., GOLDSCHMIDT, R. R., SOARES, J. A., FERLIN, C., **Previsão de Séries Temporais Utilizando Lógica Nebulosa**, artigo apresentado no 4º CONTECSI - Universidade de São Paulo, Junho de 2007.
- SOUZA, R. C., 1989, **Modelos Estruturais para Previsão de Séries Temporais: Abordagens Clássica e Bayesiana**. 17º Colóquio Brasileiro de Matemática. IMPA, Rio de Janeiro.
- STATSOFT, 2004, **Electronic Statistics Textbook**, StatSoft, Inc., Tulsa, OK: StatSoft. Disponível em <http://www.statsoft.com/textbook/stathome.html>. Acesso em 27/02/2005.
- SUNSPOT DATA, **Solar Influences Data Analysis Center**, 2006. Disponível em <http://sidc.oma.be/sunspot-data>. Acesso em 03/05/2006.
- TANSCHKEIT, R., GOLDSCHMIDT, R. R., SILVA, A. T. R. da, 2005, **Extração de Regras Fuzzy e Sistemas de Inferência para a Previsão de Séries Temporais**, PUC e IME, Rio de Janeiro, RJ.
- VALE, M. N. do, 2001, **Sistema de Inferência Fuzzy com Geração Automática de Regras**. Trabalho de Conclusão de Curso. UERJ, Rio de Janeiro, RJ, Brasil.
- VEIGA, J., 2005, **Modelagem Temporal das Medidas de Vazão de Drenos na Barragem de Funil (RJ) Utilizando Redes Neurais e Métodos Estatísticos**, Tese de Doutorado. Disponível em http://www.maxwell.lambda.ele.puc-rio.br/cgi-bin/db2www/PRG_0651.D2W/META2?CdLinPrg=pt&CdLin=pt&NrConSis=7060&CdIdeAut=JVEIGA&CdNatCon=TE. PUC, Rio de Janeiro, RJ, Brasil. Acesso em 16/03/2006.

WAN, E. A, **Time Series Data**, Oregon Health & Science University. Disponível em <http://www.cse.ogi.edu/~ericwan/data.html>. Acesso em 03/06/2006.

WANG, L., MENDEL, J. M., 1992, **Generating Fuzzy Rules by Learning from Example**. IEEE Transactions on System, Man and Cybernetics, Vol. 22, N. 6, Novembro/Dezembro.

WESSA. P., **Free Statistics Software**, Office for Research Development and Education, versão 1.1.18. Disponível em <http://www.wessa.net/finmardata.wasp?B1=1>. Acesso em 18/03/2006.