

PREVISÃO DE SÉRIES TEMPORAIS ATRAVÉS DE UMA MODELAGEM HÍBRIDA PARA A SÉRIE CANADIAN LYNX

Samuel Bellido Rodrigues

Universidade Tecnológica Federal do Paraná - UTFPR
Av. Brasil, 4232 -Medianeira-PR
samuelb@utfpr.edu.br

Arinei Carlos Lindbeck da Silva

Universidade Federal do Paraná - UFPR
Centro Politécnico – Jardim das Américas -Curitiba-PR
arineicls@gmail.com

Luiz Albino Teixeira Júnior

Universidade Federal da Integração Latina-Americana - UNILA
Av. Tancredo Neves, 6731 – PTI - Foz do Iguaçu-PR
luiz.a.t.junior@gmail.com

Edgar Manuel Carreño Franco

Universidade Estadual do Oeste do Paraná – UNIOESTE
Av. Tancredo Neves, 6731, Foz do Iguaçu, PR. CEP 85.867-900
emfra.uinoeste@gmail.com

Rafael Morais Souza

Universidade Federal de Minas Gerais -UFMG
Avenida Antônio Carlos, 6627-Belo Horizonte, 31270-901, Brasil
rafael.morais.souza@gmail.com

RESUMO

Nas últimas décadas, a análise de séries temporais tem sido uma área de pesquisa bastante ativa. Diversos modelos de previsão têm sido com o objetivo de se obter um maior nível de acurácia nas predições. Recentemente, tanto a técnica Support Vector Regression (SVR) quanto a técnica Least Squares Support Vector Regression (LSSVR) vem sendo apresentadas como uma alternativa promissora para modelagem e previsão de séries temporais, em consonância aos métodos tradicionais. Vários trabalhos destacam o desempenho de tais técnicas e conclusões mistas em termos de acurácia de previsão são apresentadas. Assim sendo, este artigo propõe uma abordagem híbrida para previsão de séries temporais que combina linearmente as metodologias SVR e LSSVR. Para tal, é usada uma decomposição wavelet para o pré-processamento dos dados. Os resultados indicaram que o método proposto pode ser uma abordagem eficaz para prever séries temporais.

PALAVRAS-CHAVE: Previsão de Séries Temporais, Wavelet, Support Vector Machine.

ABSTRACT

In last decades the time series analysis has been a very active area of research. Several forecasting models have been developed with the aim of obtaining greater accuracy level in predictions. Recently, both the Support Vector Regression (SVR) technique as the Least Squares Support Vector Regression (LSSVR) technique are being presented as a promising alternative for modeling and time series forecasting, in consonance to the traditional methods. Several studies point out the performance of these techniques and mixed conclusions in terms of forecasting accuracy are presented. Thereby, this paper proposes a hybrid approach for prediction of time series that combines linearly the SVR and LSSVR methodologies. For this, it is used a wavelet decomposition for preprocessing of the data. The results have indicated that the proposed method can be an effective approach to forecast time series.

KEYWORDS: Time Series Forecast, Wavelet, Support Vector Machine.

1. Introdução

Um dos objetivos no estudo de séries temporais (estocásticas) consiste em modelá-las com a finalidade de se produzir previsões, com o maior nível de precisão possível. Não obstante, para prevê-las, é necessária a utilização de observações passadas que são analisadas com o intuito de desenvolver um modelo matemático que a descreve por meio de uma relação existente entre observações passadas e futuras, Hamilton (1994). A acurácia nas projeções temporais é de suma importância em muitos processos de tomada de decisão, o que motiva a realização de pesquisas, cada vez maiores, neste campo do conhecimento.

Nesta perspectiva, o método preditivo individual *support vector regression* (SVR), cuja versão inicial foi proposta por Vapnik (1995), consiste em uma metodologia preditiva baseada na teoria da aprendizagem estatística, a qual adere ao princípio da minimização do risco estrutural. Tal abordagem que vem sendo frequentemente utilizada, nos últimos anos, na projeção de séries temporais, principalmente aquelas exibem não linearidade em sua estrutura de autodependência, com extraordinário sucesso, como mostram Vapnik et al. (1996), Hong (2011), Anandhi e Chezian (2013). Derivado do SVR, o método *least squares support vector regression* (LSSVR), proposto por Suykens e Vandewalle (1999), possui as mesmas capacidades que o SVR, no tocante à capacidade de mapear estruturas de autodependência não-lineares, porém um de diferenciais reside no fato de que utiliza restrições de igualdade, ao invés de restrições de desigualdade. Além disso, adota o sistema de mínimos quadrados lineares como a função objetivo a ser otimizado que o torna, do ponto de vista computacional, atraente.

Da Análise Wavelet, Mallat (2009), obtém-se uma decomposição ortogonal de uma série temporal em termos de *componentes wavelet* (CWs), que consistem em subséries temporais com frequências espectrais diferentes e constantes, com o intuito de gerar subséries temporais com melhores padrões de comportamento que a original. Diversos artigos utilizaram a decomposição wavelet como métodos auxiliares para pré-processamento (no caso, decomposição) dos sinais temporais e obtiveram ganhos preditivos relevantes, como, por exemplo, em Teixeira Jr et al. (2012). Em Teixeira Jr et al. (2012), é apresentada uma metodologia em que as CWs de uma série temporal constituem os padrões de entrada de uma *rede neural artificial multilayer perceptron* (RNA-MLP), Zhang (2003), cuja camada de saída neural fornece as suas previsão, dentro e fora da amostra. Os autores mostram que os resultados preditivos da RNA-MLP integrada com a decomposição wavelet alcançou resultados muito melhores aos de uma RNA-MLP convencional (ou seja, sem realizar o pré-processamento dos padrões de entrada via a decomposição wavelet).

Por sua vez, é bem conhecido na literatura de séries temporais que a combinação linear de previsões pode acarretar ganhos de acurácia relevantes, no processo preditivo. Em Granger (1989), por exemplo, podem ser verificadas na modelagem de diversas séries temporais, oriundas de diferentes áreas, sendo que a combinação linear de previsões incorreu em ganhos de acurácia preditiva. Faria e Mubwandarikwa (2008), por outro lado, salientam que a previsão combinada tende a ser mais informativa, uma vez que pode ser visualizada como uma agregadora de informações oriundas de diferentes fontes (métodos preditivos individuais, a saber). Em Wallis (2011) e em Mancuso e Werner (2013), encontram-se revisões históricas, nas quais são referidos e comentados muitos dos principais artigos que tratam de combinações lineares de previsões. Recentemente, várias outras contribuições atestam e incentivam à utilização de combinações de previsões, como, por exemplo em: Babu e Reddy (2014); Kosanan e Kantanantha (2014); Rodrigues et al. (2014); Teixeira et al. (2014); Rodrigues et al. (2015) e Teixeira et al. (2015).

Dado o exposto, este artigo propõe uma metodologia preditiva (aplicável a quaisquer séries temporais que exiba estrutura de autodependência), a qual integra as seguintes abordagens: decomposição wavelet, SVR, LSSVR e combinação linear de previsões. Desse modo, este artigo é dividido, como segue: Seção 1, é feita uma contextualização; na Seção 2, realiza-se uma sucinta

revisão da literatura; na Seção 3, a metodologia proposta é descrita; na Seção 4, são apresentados os principais resultados numéricos, com aplicações à série temporal Canadian lynx, Zhang (2003), e, finalmente, na Seção 5, têm-se as principais conclusões.

2. REVISÃO DA LITERATURA

2.1 Decomposição Wavelet

Na Análise de Fourier Canônica, uma série temporal com T observações, denotada por $\mathbb{X}(\cdot) := (\mathbb{X}(1), \mathbb{X}(2), \dots, \mathbb{X}(T))$, pode ser ortogonalmente decomposta em termos de senos e cossenos definidos em diversas bandas de frequências, Morettin e Tolo (2006), enquanto que, na Análise Wavelet, $\mathbb{X}(\cdot)$ pode, também de forma ortogonal, ser decomposta, simultaneamente, a partir de informações de frequência e de tempo, ambas capturadas a partir de um conjunto de *funções wavelet* $\{\mathbb{W}_{\mathbb{m}, \mathbb{n}}(\cdot)\}_{(\mathbb{m}, \mathbb{n}) \in \mathbb{Z} \times \mathbb{Z}}$, Teixeira Júnior et al. (2015), onde \mathbb{Z} é o conjunto dos números inteiros, sendo que cada mapa $\mathbb{W}_{\mathbb{m}, \mathbb{n}}(\cdot)$ é tal que $\mathbb{W}_{\mathbb{m}, \mathbb{n}}(\cdot) = 2^{\frac{\mathbb{m}}{2}} \mathbb{W}(2^{\mathbb{m}}(\cdot) - \mathbb{n})$, o qual é gerado por meio de convoluções de dilatação (associada ao parâmetro m) e de translação (ao parâmetro n) de uma função wavelet $\mathbb{W}(\cdot)$ original, conforme Mallat (2009). Ainda de acordo com Mallat (2009), o parâmetro m é chamado de parâmetro de escala diádica e n , de parâmetro de translação unitária.

Em Teixeira Jr et al. (2012), verifica-se, empiricamente, que cada função wavelet $\mathbb{W}_{\mathbb{m}, \mathbb{n}}(\cdot)$ está associada ao mapeamento das componentes wavelet (CWs) de detalhe (que são as CW de alta frequência) da série temporal $\mathbb{X}(\cdot)$; e, teoricamente, que a CW de aproximação (que é a CW de baixa frequência) da série temporal $\mathbb{X}(\cdot)$ está associada ao conjunto de *funções escala* $\{\mathbb{V}_{\mathbb{m}, \mathbb{n}}(\cdot)\}_{\mathbb{m} \in \mathbb{Z}}$, Teixeira Júnior et al. (2015). Cada mapa $\mathbb{V}_{\mathbb{m}, \mathbb{n}}(\cdot) = 2^{\frac{\mathbb{m}}{2}} \mathbb{V}(2^{\mathbb{m}}(\cdot) - \mathbb{n})$, onde \mathbb{V}_0 é um número inteiro fixo, é gerado por meio da convolução de translação unitária n de uma função escala original $\mathbb{V}(\cdot)$, Mallat (2009).

De acordo com Mallat (2009), as famílias wavelet, em geral, são apresentadas em pares constituídos de uma função wavelet $\mathbb{W}(\cdot)$ e uma função escala $\mathbb{V}(\cdot)$ (originais). As famílias de Haar, de Daubechies, de Coiflet e de Symlet são exemplos que, comumente, são vastamente utilizadas na decomposição de sinais temporais, antes de sua efetiva modelagem, Mallat (2009).

Baseado em Kubrusly e Levan (2006) e Teixeira Júnior et al. (2015), tem-se que a *decomposição wavelet* de uma série temporal $\mathbb{X}(\cdot)$ é dada, genericamente, em (1).

$$\mathbb{X}(\cdot) = \mathbb{V}_{\mathbb{m}, \mathbb{n}}(\cdot) + \sum_{\mathbb{m}=\mathbb{V}_0}^{\infty} \mathbb{W}_{\mathbb{m}, \mathbb{n}}(\cdot) \quad (1)$$

Onde: (i) $\mathbb{V}_{\mathbb{m}, \mathbb{n}}(\cdot) = \sum_{\mathbb{m} \in \mathbb{Z}} \mathbb{V}_{\mathbb{m}, \mathbb{n}} \mathbb{V}_{\mathbb{m}, \mathbb{n}}(\cdot)$ é a CW de aproximação de nível \mathbb{V}_0 de $\mathbb{X}(\cdot)$; $\mathbb{W}_{\mathbb{m}, \mathbb{n}}(\cdot) = \sum_{\mathbb{m} \in \mathbb{Z}} \mathbb{W}_{\mathbb{m}, \mathbb{n}} \mathbb{W}_{\mathbb{m}, \mathbb{n}}(\cdot)$ é a componente de detalhe de nível \mathbb{V} de $\mathbb{X}(\cdot)$; (ii) $\mathbb{W}_{\mathbb{m}, \mathbb{n}}$ e $\mathbb{V}_{\mathbb{m}, \mathbb{n}}$ são, respectivamente, os coeficientes de detalhe e de aproximação e são calculados, ainda nesta ordem, a partir do produto interno usual entre $\mathbb{X}(\cdot)$ e $\mathbb{W}_{\mathbb{m}, \mathbb{n}}(\cdot)$ (isto é, $\mathbb{W}_{\mathbb{m}, \mathbb{n}} = \sum_{\mathbb{m}=1}^{\mathbb{V}} \mathbb{X}(\mathbb{m}) \mathbb{W}_{\mathbb{m}, \mathbb{n}}(\mathbb{m})$) e entre $\mathbb{X}(\cdot)$ e $\mathbb{V}_{\mathbb{m}, \mathbb{n}}(\cdot)$ (isto é, $\mathbb{V}_{\mathbb{m}, \mathbb{n}} = \sum_{\mathbb{m}=1}^{\mathbb{V}} \mathbb{X}(\mathbb{m}) \mathbb{V}_{\mathbb{m}, \mathbb{n}}(\mathbb{m})$), para todo $\mathbb{V}_0, \mathbb{V}, \mathbb{m} \in \mathbb{Z}$.

2.2 Support Vector Regression

O método preditivo individual *support vector regression* (SVR) é fundamentado nas *support vector machines* (SVMs), propostas por Vapnik (1995) Vapnik (1995). O modelo produzido pelo SVR depende apenas da própria série temporal e de um objetivo chamado de *função custo* (o qual é utilizado para a construção do SVR).

As descrições acerca do SVR são baseadas no tutorial editado por Smola e Schölkopf (1998), no qual mais detalhes teóricos e práticos podem ser encontrados. Assim, assumamos que

$\{(x_1, y_1), \dots, (x_l, y_l)\} \subset X \times \mathbb{R}$ consiste em uma amostra de dados temporais de treinamento, onde X indica o espaço dos padrões de entrada. Aqui, o objetivo é encontrar uma função $f(x)$ que

apresente no máximo um desvio ε dos alvos y_i ($i=1,\dots,l$) de treinamento. A função linear de aproximação $f(x)$ é definida por: $f(x) = \langle w, x \rangle + b$ com $w \in X, b \in \mathbb{R}$, onde $\langle \dots \rangle$ denota o produto interno sobre o espaço X . Não obstante, nem sempre é possível garantir a viabilidade do problema, uma vez que existem pontos que violam as restrições. Por conta disso, a formulação em (2), proposta por Vapnik (1995), utiliza variáveis de folga ξ_i e ξ_i^* .

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2} \|w\|^2 + C \left(\sum_{i=1}^l \xi_i + \xi_i^* \right) \\ \text{sujeito a:} \quad & \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \quad \forall i = 1, \dots, l \end{cases} \end{aligned} \quad (2)$$

Para a construção da formulação dual do problema descrito em (2), são introduzidos alguns multiplicadores de Lagrange (α_i, α_i^*) não negativos, conforme Vapnik (1995), de modo que o mesmo é convertido em um problema de maximização, na sua forma dual, dado em (3).

$$\begin{aligned} \text{Maximizar} \quad & \left(-\frac{1}{2} \right) \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + y_i \sum_{i=1}^l (\alpha_i - \alpha_i^*) \\ \text{Sujeito a} \quad & \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ & \alpha_i, \alpha_i^* \in [0, C], \quad \forall i = 1, \dots, l \end{aligned} \quad (3)$$

Assim, a função linear de aproximação é reescrita tal como:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b$$

A formulação dual do problema do SVR fornece como alternativa fazer modelagens em um espaço de alta dimensionalidade. Adicionalmente, pode-se realizar um mapeamento não-linear de dados temporais de entrada para espaços de dimensão maior, nos quais a regressão linear torna-se possível, desde que sejam utilizadas abordagens do SVR baseadas em *funções Kernel* (não-lineares), as quais são denotadas por $K(x, x') := \langle \phi(x), \phi'(x') \rangle$, Smola e Schölkopf (1998). Consequentemente, o problema descrito em (3) pode é reeditado conforme a formulação em (4).

$$\begin{aligned} \text{Maximizar} \quad & -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + y_i \sum_{i=1}^l (\alpha_i - \alpha_i^*) \\ \text{Sujeito a} \quad & \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ & \alpha_i, \alpha_i^* \in [0, C], \quad \forall i = 1, \dots, l \end{aligned} \quad (4)$$

Com efeito, a função de aproximação da SVR consiste em um modelo não-linear e é

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x) + b$$

reeditada tal como segue:

Dentre as funções Kernel não-lineares utilizadas na modelagem SVR, destacam-se: o linear, o polinomial, a sigmoideal e a gaussiana, cujas definições podem ser encontradas em Smola e Schölkopf (1998).

2.3 Least Squares Support Vector Regression

O método *least squares support vector regression* (LSSVR) é uma variante do SVR e foi proposto por Suykens e Vandewalle (1999). De acordo com Suykens et al. (2002), a principal modificação do LSSVR, em relação ao SVR de Vapnik (1995), ocorre em dois pontos: (1º) em vez de restrições de desigualdade, são consideradas restrições de igualdade, ou seja, no LSSVR encontra-se a solução através da resolução de um conjunto de equações lineares ao invés de um problema quadrático, conforme equação 2; e (2º) a função de perda quadrática é tomada para a variável e_i (erro de aproximação) que desempenham um papel semelhante ao das variáveis de folga ξ_i e ξ_i^* na formulação do SVR.

Desse modo, seja $\{(x_1, y_1), \dots, (x_l, y_l)\} \subset X \times \mathbb{R}$ uma amostra de dados temporais de treinamento, onde X representa o espaço de padrões de entrada (por exemplo, $X = \mathbb{R}^d$). Aqui, o objetivo é ajustar numericamente um modelo $f(x) = \langle w, x \rangle + b$ com $w \in X, b \in \mathbb{R}$, onde $\langle \cdot, \cdot \rangle$ denota o produto interno em X , considerando o problema em (5).

$$\begin{aligned} \text{Minimizar} \quad & \frac{1}{2} \|w\|^2 + \frac{\gamma}{2} \sum_{i=1}^l e_i^2 \\ \text{s.a} \quad & (y_i - (\langle w, x_i \rangle + b)) = e_i, \quad i = 1, 2, \dots, l \end{aligned} \quad (5)$$

Onde: w e b são os vetores de pesos; γ é o parâmetro de regularização; e_i é o i -ésimo erro de aproximação. Para resolver o problema (5), a função de Lagrange é definida em (6).

$$L(w, b, e, \alpha) = \frac{1}{2} \|w\|^2 + \frac{\gamma}{2} \sum_{i=1}^l e_i^2 - \sum_{i=1}^l \alpha_i (\langle w, x_i \rangle + b + e_i - y_i) \quad (6)$$

Onde α_i ($i=1, \dots, l$) são os multiplicadores de Lagrange, (que podem ser positivos ou negativos devido às restrições de igualdade como decorre das condições de Kuhn-Tucker, Fletcher (1987)).

No tocante às condições de otimalidade da equação (6), as mesmas são dadas em (7).

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^l \alpha_i x_i \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^l \alpha_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \Rightarrow \alpha_i = \gamma e_i, \quad i = 1, 2, \dots, l \\ \frac{\partial L}{\partial \alpha_i} = 0 \Rightarrow w^T x_i + b + e_i - y_i = 0 \quad i = 1, 2, \dots, l \end{cases} \quad (7)$$

Onde, eliminando-se os vetores w e e , obtém-se um sistema de equações lineares, descrito em (8), ao invés de um problema de programação quadrática (descrito em (2)).

$$\begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 & I_v^T \\ I_v & K + \gamma^{-1} I \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (8)$$

Onde: $y := [y_1, y_2, \dots, y_l]^T$, $I_v := [1, 1, \dots, 1]^T$, $\alpha := [\alpha_1, \alpha_2, \dots, \alpha_l]^T$ e $K(x, x') := \langle \phi(x), \phi'(x) \rangle$ é a função Kernel, que satisfaz às condições de Mercer, Smola e Schölkopf (2004), que realiza um mapeamento não-linear dos dados de entrada para um espaço de dimensão maior, onde a regressão linear torna-se possível. Com efeito, o modelos LSSVR é definido em termos dos coeficientes α, b e da função Kernel (K) e é dado genericamente por

$$f(x) = \sum_{i=1}^l \alpha_i K(x, x_i) + b$$

2.4 Combinação Linear de Previsões

A combinação linear de previsões, de acordo com Clemen (1989), consiste em uma metodologia atraente na obtenção de previsões mais acuradas. Assim, ao invés escolher o melhor método preditivo individual, baseado em algum critério de seleção, o problema passa a ser quais os previsores individuais que podem ser utilizados na obtenção das melhores previsões linearmente combinadas.

De acordo com Teixeira Jr (2013), as previsões linearmente combinadas necessitam de algum mecanismo ponderador de forma à minimização da variância dos resíduos combinados ou outra função residual. Os pesos podem ser fixos ou variáveis e podem não necessariamente ser positivos ou somar uma unidade. Destaca ainda que as médias simples, ponderada ou harmônica, também podem ser utilizadas na combinação de previsões.

3. Metodologia Proposta

Seja $(y(t))_{t=1}^T$ uma série temporal de cardinalidade T a ser prevista, o método híbrido é realizado de acordo com os quatro passos seguintes.

Passo 1 - Decomposição wavelet da série original: uma decomposição wavelet de nível r , Teixeira Júnior et al. (2015), na série temporal $(y(t))_{t=1}^T$ é realizada, gerando-se $r+1$ CWs - isto é, uma CW de aproximação de nível m_0 , denotada por $(y_{A_{m_0}}(t))_{t=1}^T$, e r CWs de detalhe de níveis $m_0, m_0+1, \dots, m_0+(r-1)$, denotadas por $(y_{D_{m_0}}(t))_{t=1}^T, (y_{D_{m_0+1}}(t))_{t=1}^T, \dots, (y_{D_{m_0+(r-1)}}(t))_{t=1}^T$ respectivamente.

Na prática, usualmente se adota para o parâmetro m_0 valor igual ao nível de decomposição r . As CWs de aproximação e de detalhe são obtidas a partir do aplicativo computacional *Matlab 8.0*, que oferece várias opções de bases wavelet, tais como: as famílias de Haar, Daubechies, coiflets e symlets.

Passo 2 - Modelagem das componentes wavelet via SVR e LSSVR: as CW obtidas no passo 1 são modeladas através de um SVR e de um LSSVR.

Para a obtenção dos modelos individuais SVR, foi utilizada uma biblioteca do software *Matlab 8.0* denominada LIBSVM, versão 3.2, desenvolvida por Chang e Lin (2015), disponível no link <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.

Para modelagem do LSSVR foi utilizada uma biblioteca para *Matlab 8.0* denominada LS-SVMlab, versão 1.8, disponível no link <http://www.esat.kuleuven.be/sista/lssvmlab/>, a qual foi desenvolvida por Brabanter et al. (2011), baseado nos trabalhos de Suykens et al. (2002) e de Suykens e Vandewalle (1999).

Passo 3 - Combinação linear das previsões dos métodos preditivos SVR e LSSVR: Obtidas as previsões da componente wavelet de aproximação e as r previsões das componentes de detalhe, na etapa anterior, através dos métodos SVR e LSSVR, faz-se a combinação linear das

previsões, com o objetivo de se obter a previsão da componente de aproximação, e de cada umas das r componentes de detalhe.

Para cada uma das CW a combinação é dada pela equação 9.

$$\left(CL_{CW_{SVR-LSSVR}} \right)_{t=T^*}^T = \rho_1 \times \left(y_{CW,SVR_{CW}}(t) \right)_{t=T^*}^T + \rho_2 \times \left(y_{CW,LSSVR_{CW}}(t) \right)_{t=T^*}^T \quad (9)$$

Onde os pesos ρ_1 e ρ_2 , para cada uma das combinações é determinado através de um problema de programação linear cuja função objetivo consiste na minimização do erro quadrático médio (MSE), conforme 10.

$$\begin{aligned} \text{Min } MSE &= \frac{1}{T^*} \sum_{t=T^*}^T \left(y_{CW} - CL_{CW_{SVR-LSSVR}} \right)^2 \\ \text{s.a. } &\rho_1 \text{ e } \rho_2 \text{ são irrestritas} \end{aligned} \quad (10)$$

Passo 4 – Soma das combinações lineares: Neste passo soma-se para cada t as previsões linearmente combinadas obtidas no passo anterior, conforme 11.

$$\left(\hat{y}_H(t) \right)_{t=T^*}^{T+h} = CL_{A_{m_0, SVR-LSSVR}} + CL_{D_{m_0, SVR-LSSVR}} + CL_{D_{m_0+1, SVR-LSSVR}} + \dots + CL_{D_{m_0+(r-1), SVR-LSSVR}} \quad (11)$$

Nesta etapa soma-se para cada t as previsões linearmente combinadas, gerando desta forma a previsão do modelo híbrido dentro e fora da amostra, que será denotado por $\left(\hat{y}_H(t) \right)_{t=T^*}^{T+h}$.

Os passos do método híbrido proposto podem ser visualizados no fluxograma da Figura 1.

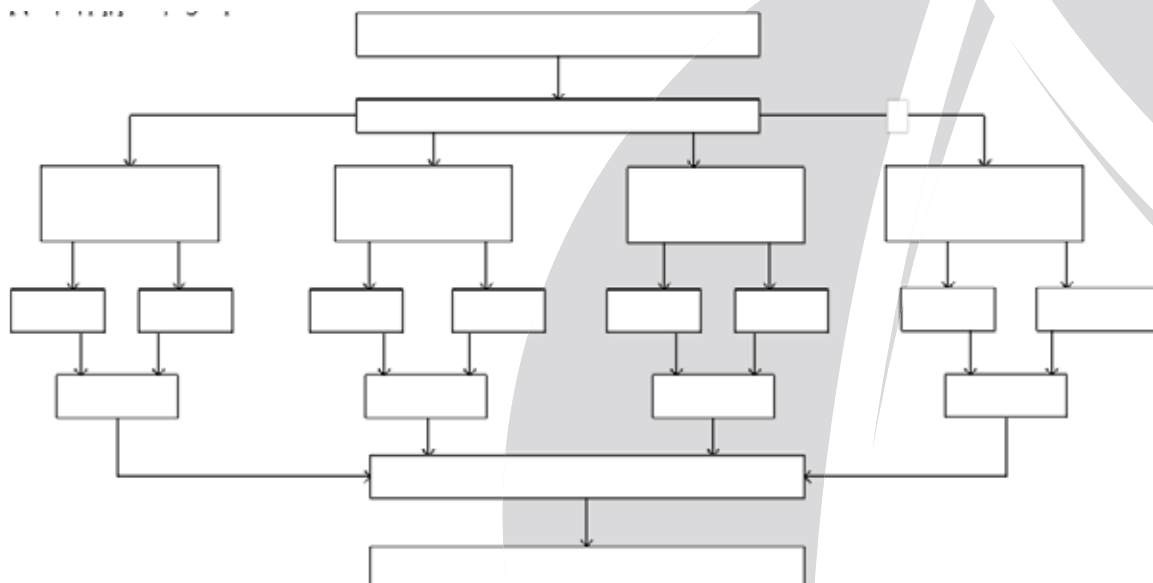


Figura 1 - Fluxograma do modelo híbrido.

4. Resultados Numéricos

Nesta seção, a série temporal anual do Canadian Lynx (lince canadense) foi usada para expor a eficácia e o poder do método proposto para gerar previsões. A série temporal refere-se ao número de lince preso por ano no distrito de Rio Mackenzie do Norte do Canadá para o período 1821-1934. Tal série pode ser obtida no seguinte endereço eletrônico: <https://datamarket.com/>. Existem na literatura vários trabalhos que fazem análise e previsão da série de tempo Canadian Lynx, credita-se a Moran (1953), como sendo o primeiro trabalho de análise desta. O gráfico da

série Canadian Lynx (com 114 observações no período: 1821-1934) é dado na Figura 2, onde se pode observar uma periodicidade de aproximadamente 10 anos, Zhang (2003).

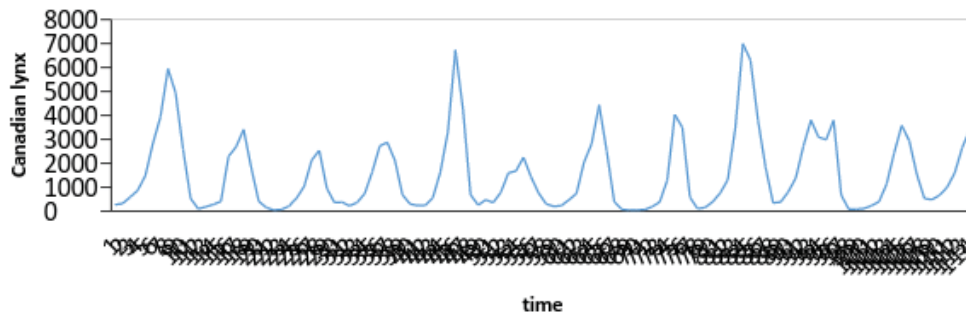


Figura 2: Série Canadian Lynx (1821-1934).

Neste experimento, foi considerado a previsão um passo à frente, com um horizonte de previsão de 14 passos à frente (ou seja, $h = 14$), já que o presente método é comparado aos resultados de outros autores, que referem-se exclusivamente a este tipo de projeção. É aplicado a série temporal do Canadian Lynx o log na base 10, tal série possui cardinalidade $T=114$, donde os primeiros 100 valores Treino e os 14 últimos valores para teste. Para a série Canadian Lynx foi aplicado a decomposição wavelet de nível $r=2$, gerando uma subsérie denominada de componente de aproximação e duas subséries denominadas de componentes de detalhe.

A base escolhida para decomposição foi a base Daubechies de momento nulo igual a 8 (db 8), gerando uma série de baixa frequência e duas séries de alta frequência, conforme pode ser visto na Figura 3.

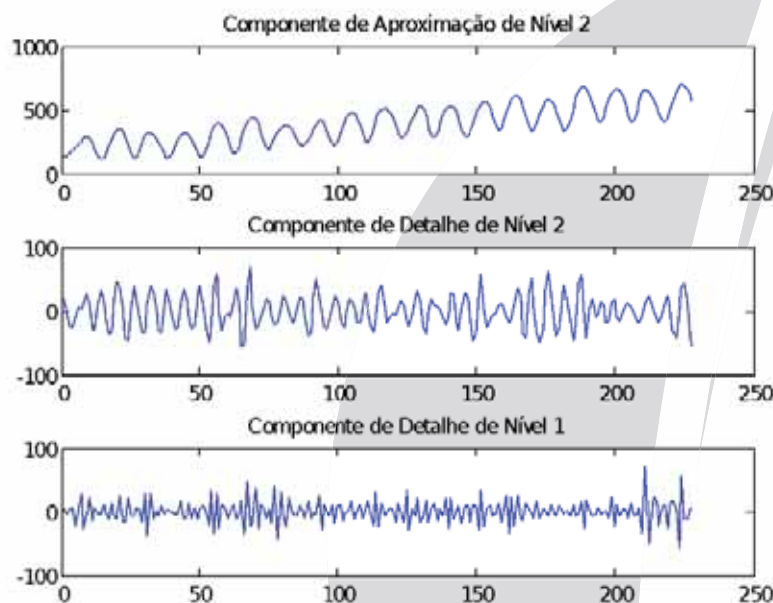


Figura 3: Séries procedentes da decomposição wavelet de base Daubechies 8

Quanto à modelagem SVR, foi utilizado na análise o Kernel polinomial, dado por $(\gamma * \langle u, v \rangle + \text{coef0})^{\text{degree}}$ e o Kernel RBF, dado por $\exp(-\gamma * \|u - v\|^2)$. É conhecido que o desempenho de generalização de um modelo SVR depende de um bom ajuste do parâmetro de regularização (C) e do parâmetro épsilon da função perda, bem como dos parâmetros relacionados ao Kernel, escolhido, Smola e Schölkopf (1998). A biblioteca, que foi utilizada, permitem que os parâmetros sejam definidos pelo usuário. A fim de melhor avaliar o

desempenho da abordagem proposta, para cada subsérie a ser modelada os parâmetros C , ϵ e os parâmetros do Kernel, foi configurado uma busca por grade (grid search), Hsu et al. (2010), e selecionado aquele que apresentou um melhor desempenho de validação cruzada 10-fold no conjunto de treino.

Para o LSSVR, é necessário o ajuste de dois parâmetros, que são o parâmetro de regularização (γ) e do parâmetro do Kernel (σ^2) , sendo que nesse caso somente o Kernel

$$\exp\left(-\frac{\|u-v\|^2}{2\sigma^2}\right)$$

gaussiano foi utilizado, e é dado por . A fim de se determinar a melhor configuração, foi programado um algoritmo de busca por grade e selecionado aquele que apresentou um melhor desempenho de validação cruzada (10-fold) no conjunto de treinamento.

Segue na Tabela 1 os resultados selecionados após o treino com base no menor Cross Validation Mean squared error (CVMSE) na modelagem SVR e na Tabela 2 para a modelagem LSSVR.

Tabela 1- Resultados dos parâmetros selecionados SVR.

Série	Kernel	Gamma	Degree	coef0	Cost	epsilon	janela
A2	Polinomial	0.36	3	3	121	0.001	10
D2	Polinomial	0.01	3	4	1094	0.01	10
D3	Polinomial	0.21	3	4	73	0.01	10

Tabela 2- Resultados dos parâmetros selecionados LSSVR.

Série	Kernel	γ	σ^2	janela
A2	Rbf	47345	40.76	4
D1	Rbf	5157098	21742.47	10
D2	Rbf	8201281283894		10
		4	1865591127	

Os resultados obtidos pelo método proposto são comparados, na Tabela 3, com as seguintes abordagens: (i) Zhang (2003), que comparou resultados obtidos com os modelos ARIMA e RNA com o híbrido ARIMA-RNA; (ii) Kajitani et al. (2005), que utilizaram na modelagem da Canadian Lynx via uma RNA Feed-Forward (FFNN - Feed-Forward Neuron Networks) e modelo SETAR (Self-Exciting Threshold Auto-Regression); (iii) Aladag et al. (2009), que realizaram a combinação de RNAs recorrentes Elman e um modelo ARIMA; (iv) Zheng e Zhong (2011), que apresentaram uma metodologia que combina a Função de Base Radial (RBF – radial basis function) de uma rede neural artificial baseado em alisamento binomial (BS-RBF) e uma abordagem que combina a função de base radial (RBF) de uma rede neural com modelo auto-regressivo (AR), utilizando a técnica alisamento binomial (BS), denominado BS-RBFAR; (v) Khashei e Bijari (2011), os quais propuseram um modelo híbrido RNA/ARIMA e o compararam com a modelagem ANN (p,d,q), (vi) Khashei e Bijari (2012), que sugeriram metodologias híbridas com base em um classificador (redes neurais probabilísticas (PNNs)), gerando os previsores ARIMA/PNN e ANN/PNN, (vii) Karnabopathy e Venkatesan (2012), que utilizaram em sua modelagem o modelo Full Range Auto-regressive (FRAR); (viii) Adhikari e Agrawal (2013), os quais propuseram uma metodologia combinando vários algoritmos de treinamento denominado (Ensemble of Multiple ANN Training Methods) e a comparou com as modelagens ARIMA e Máquinas de Vetores Suporte (SVM) para Regressão; e (ix) Ismail e Shabri (2014), que usaram Máquinas de Vetores Suporte (SVM) para Regressão e SVM com mínimos quadrados (LSSMV).

Tabela 3 - Comparativo entre métodos de previsão aplicados à série temporal log Canadian Lynx, utilizando as estatísticas de aderência MSE e MAE.

Autores	Métodos	h=14	
		MSE	MAE
Zhang (2003)	ARIMA	0,020486	0,112255
	ANN	0,020466	0,112109
	HYBRID	0,017233	0,103972
Kajitani et al. (2005)	FFNN	0,0090	-
	SETAR	0,0140	-
Aladag et al. (2009)	HYBRID	0,00900	-
Zheng; Zhong (2011)	BS-RBF	0,002809	-
	BS-RBFAR	0,002199	-
Khashei; Bijari (2011)	ANN(p,d,q)	0,013609	0,089625
	ANNs/ARIMA	0,00999	0,085055
Khashei; Bijari (2012)	ARIMA/PNN	0,011461	0,084381
	ANN/PNN	0,014872	0,079628
Karnaboopathy; Venkatesan (2012)	FRAR	0,00455	-
Adhikari; Agrawal (2013)	ARIMA	0,01285	-
	SVR	0,05267	-
	ENSAMBLE	0,00715	-
Ismail; Shabri (2014)	SVR	0,00850	0,07460
	LSSVM	0,00300	0,04180
	Método Proposto	0,00123	0,02626

Quando comparado com os resultados de outros métodos nas mesmas condições, nossa proposta híbrida se mostrou melhor, com relação as medidas de acurácia MSE e MAE. A Figura 4 mostra os valores de previsão para as últimas 14 observações, utilizadas neste estudo. A linha continua representa os dados reais da série log Lynx, enquanto a linha tracejada representa os valores de previsão, pelo modelo híbrido proposto.

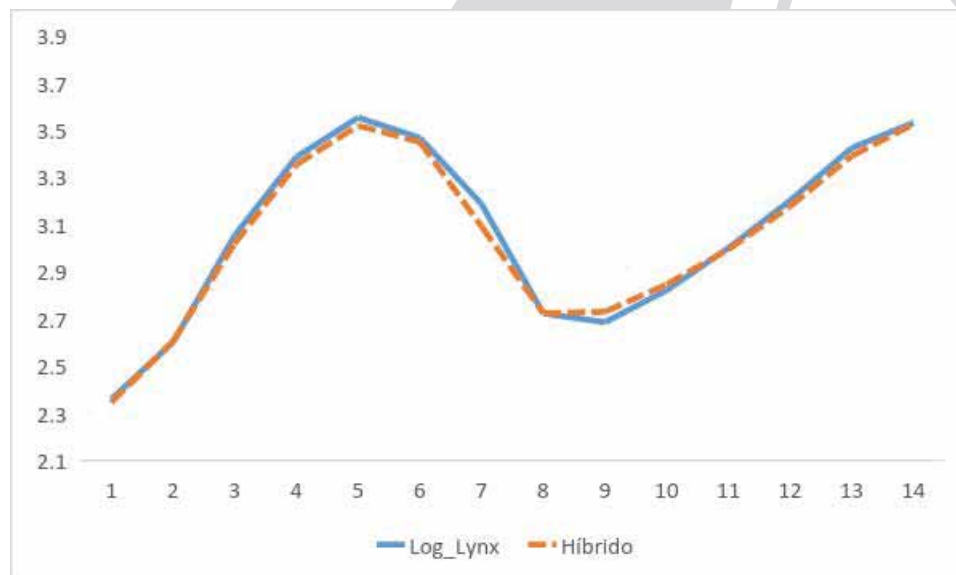


Figura 4: Série original versus modelo híbrido

7. Conclusão

Nesse trabalho, foi proposta um método híbrido para previsão de séries temporais usando a decomposição Wavelet, a modelagem via Support Vector Regression (SVR) e Least

Square Support Vector Regression (LSSVR) e a combinação de previsões. A eficácia do método foi verificada por meio da aplicação na modelagem da série temporal Canadian Lynx, e comparada com diversas outras modelagens, nas mesmas condições, a qual mostrou-se superior quando comparadas as medidas de acurácia MSE e MAE.

Bibliografia

ADHIKARI, R.; AGRAWAL, R. K. A Homogeneous Ensemble of Artificial Neural Networks for Time Series Forecasting. **International Journal of Computer Applications**, v. 32, n. 7, p. 8, 2013.

ALADAG, C. H.; EGRIOGLU, E.; KADILAR, C. Forecasting nonlinear time series with a hybrid methodology. **Applied Mathematics Letters**, v. 22, n. 9, p. 1467–1470, 2009. Elsevier Ltd.

ANANDHI, V.; CHEZIAN, R. M. Support Vector Regression to Forecast the Demand and Supply of Pulpwood. **International Journal of Future Computer and Communication**, v. 2, n. 3, p. 1–4, 2013.

BABU, C. N.; REDDY, B. E. A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. **Applied Soft Computing**, v. 23, p. 27–38, 2014. Elsevier B.V.

BRABANTER, K. DE; KARSMAKERS, P.; OJEDA, F.; et al. **LS-SVMlab Toolbox User ' s Guide**. 2011.

CHANG, C.-C.; LIN, C.-J. LIBSVM -- A Library for Support Vector Machines. Disponível em: <<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>>. Acesso em: 2/1/2015.

CLEMEN, R. T. Combining forecasts: A review and annotated bibliography. **International Journal of Forecasting**, 1989.

FARIA, A. E.; MUBWANDARIKWA, E. The geometric combination of bayesian forecasting models. **Journal of Forecasting**, v. 27, n. 1, p. 519–535, 2008.

FLETCHER, R. Practical methods of optimization; (2nd ed.). , 1987. Wiley-Interscience.

GRANGER, C. W. J. Combining forecasts – twenty years later. **Journal of Forecasting**, v. 8, p. 167–73, 1989.

HAMILTON, J. D. **Time Series Analysis**. Princeton University Press, 1994.

HONG, W.-C. Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. **Energy**, v. 36, n. 9, p. 5568–5578, 2011.

HSU, C.-W.; CHANG, C.-C.; LIN, C.-J. **A Practical Guide to Support Vector Classification**. 2010.

ISMAIL, S.; SHABRI, A. Time Series Forecasting using Least Square Support Vector Machine for Canadian Lynx Data. **Jurnal Teknologi**, v. 1, p. 11–15, 2014.

KAJITANI, Y.; HIPEL, K. W.; MCLEOD, A. I. Forecasting nonlinear time series with feed-forward neural networks: A case study of Canadian lynx data. **Journal of Forecasting**, v. 24, p. 105–117, 2005.

KARNABOOPATHY, R.; VENKATESAN, D. Data mining in canadian lynx time series. **Journal of Reliability and Statistical Studies**, v. 5, n. 1, p. 1–06, 2012.

KHASHEI, M.; BIJARI, M. Which Methodology is Better for Combining Linear and Nonlinear Models for Time Series Forecasting ? , v. 4, n. 4, p. 265–285, 2011.

KHASHEI, M.; BIJARI, M. A new class of hybrid models for time series forecasting. **Expert Systems with Applications**, v. 39, n. 4, p. 4344–4357, 2012. Elsevier Ltd.

KOSANAN, O.; KANTANANTHA, N. Thailand ' s Para Rubber Production Forecasting Comparison. Proceedings of the International MultiConference of Engineers and Computer Scientists 2014. **Anais... . v. II**, 2014. Hong Kong.

KUBRUSLY, C. S.; LEVAN, H. Abstract wavelets generated by Hilbert space shift operators. **Advances in mathematical and applications**, v. 16, p. 643–660, 2006.

MALLAT, S. **A Wavelet Tour of Signal Processing: The Sparse Way**. Third Edit ed.2009.

MANCUSO, A. C. B.; WERNER, L. Review of Combining Forecasts Approaches. **Independent Journal of Management & Production**, v. 4, n. June, p. 248–277, 2013.

- MORAN, P. The statistical analysis of the Canadian Lynx cycle. **Australian Journal of Zoology**, v. 1, n. 3, p. 291, 1953.
- MORETTIN, P. A.; TOLOI, C. M. DE C. **Análise de séries temporais**. Edgard Blucher, 2006.
- RODRIGUES, S. B.; TEIXEIRA, L. L.; CARLOS, A.; SIQUEIRA, P. H. Combinação Linear De Redes Neurais Artificiais E Máquinas De Vetores De Suporte Para Regressão Nas Previsões De Vazões Mensais No Posto 266-Itaipu. XVII Encontro de Modelagem Computacional V Encontro de Ciência e Tecnologia de Materiais Universidade Católica de Petrópolis (UCP). **Anais...**, 2014.
- RODRIGUES, S. B.; TEIXEIRA, L. L.; TEIXEIRA JÚNIOR, L. A.; et al. Forecasting of the relative displacements in block I11 of the Itaipu hydroelectric plant dam through a ARIMA-SVR hybrid approach. **Espaço Energia**, v. 22, n. 22, p. 28–37, 2015.
- SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. **Statistics and Computing**, 1998.
- SMOLA, A.; SCHÖLKOPF, B. A tutorial on support vector regression. **Statistics and Computing**, p. 199–222, 2004.
- SUYKENS, J. A. K.; GESTEL, T. VAN; BRABANTER, J. DE. **Least Squares Support Vector Machines**. 2002.
- SUYKENS, J. A. K.; VANDEWALLE, J. Least Squares Support Vector Machine Classifiers. **Neural Processing Letters**, v. 9, n. 3, p. 293–300, 1999. Kluwer Academic Publishers.
- TEIXEIRA JR, L. A. **Combinação SSA-Wavelet de Métodos Preditivos com Ajuste Numérico MINIMAX, na Geração de Previsões e de Cenários**, 2013. Puc-Rio.
- TEIXEIRA JR, L. A.; PESSANHA, J. F. M.; MENEZES, M. L. DE; CASSIANO, K. M.; SOUZA, R. C. Redes neurais artificiais e decomposição wavelet na previsão da radiação solar direta. Congresso Latino-Iberoamericano de Investigación Operativa e Simpósio Brasileiro de Pesquisa Operacional. **Anais...** p.1401–1412, 2012.
- TEIXEIRA JÚNIOR, L. A.; SOUZA, R. M.; MENEZES, L. M.; et al. Artificial Neural Network and Wavelet decomposition in the Forecast of Global Horizontal Solar Radiation. **Sobrapo**, v. 35, n. 1, p. 1–16, 2015.
- TEIXEIRA, L. L.; RODRIGUES, S. B.; SIQUEIRA, P. H.; SILVA, A. C. L. DA; TEIXEIRA JR, L. A. Comparativo Entre Arima, Redes Neurais Artificiais, Máquinas De Vetores Suporte Para Regressão E Combinação Linear De Previsões De Vazões Mensais No Posto 266- Itaipu. Simpósio de Engenharia de Produção. **Anais...** p.1–15, 2014. XXI SIMPEP.
- TEIXEIRA, L. L.; TEIXEIRA JUNIOR, L. A.; SIQUEIRA, P. H. Forecasts of monthly flows in the post-266 Itaipu using neural networks with wavelet shrinkage. **Espaço Energia**, v. 22, n. 22, p. 8–17, 2015.
- VAPNIK, V.; GOLOWICH, S. E.; SMOLA, A. Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. **Advances in Neural Information Processing Systems**, v. 9, p. 281–287, 1996.
- VAPNIK, V. N. **The Nature of Statistical Learning Theory**. 1995.
- WALLIS, K. F. Combining forecasts : forty years later. **Applied Financial Economics**, p. 33–41, 2011.
- ZHANG, G. P. Time series forecasting using a hybrid ARIMA and neural network model. **Neurocomputing**, v. 50, p. 159–175, 2003.
- ZHENG, F.; ZHONG, S. Time series forecasting using a hybrid RBF neural network and AR model based on binomial smoothing. **World Academy of Science Engineering and Technology**, p. 1125–1129, 2011.