Modelo Autorregressivo de Integração Adaptativa*

Arthur Ronald[1], **Rebecca Salles**[1], **Kele Belloze**[1], **Dayse Pastore**[1], **Eduardo Ogasawara**[1]

[1]CEFET/RJ - Centro Federal de Educação Tecnológica Celso Suckow da Fonseca

{arthur.garcia,rebecca.salles}@eic.cefet-rj.br

{kele.belloze,dayse.pastore}@cefet-rj.br, eogasawara@ieee.org

Abstract. *Several preprocessing techniques combined with time series models have been used to predict non-stationary time series. The study of the mathematical and statistical properties of the data and the preprocessing techniques can help in the adjustment of machine learning models. Such a study, however, may not be easily obtained. Linear models enable the interpretation of such properties. This article introduces and analyzes, based on proof of concept, a new linear model applied to stationary time series that are built by using adaptive normalization. The model allows the use of autoregressive models with sliding windows of data that preserve the properties of the original series, and allow the observation of its inertia. The model was able to present superior prediction performance to other linear models consolidated in the literature, specially in short-term horizons.*

Resumo. *Diversas técnicas de preprocessamento combinadas a modelos de séries temporais vêm sendo utilizadas para previsão de séries temporais não-estacionárias. O estudo das propriedades matemáticas e estatísticas dos dados e das técnicas de preprocessamento pode auxiliar no ajustamento de modelos de aprendizado de máquina. Tal estudo, entretanto, muitas vezes não é facilmente obtido. Modelos lineares, por sua vez, possibilitam a interpretação de tais propriedades. Este artigo introduz e analisa, por meio de prova de conceito, um novo modelo linear aplicado a séries estacionárias construídas com base em normalização adaptativa. O modelo viabiliza o uso de modelos autorregressivos em cenários de janelas deslizantes que preservam as propriedades da série original, e permitem acompanhar a sua inércia. O modelo foi capaz de apresentar desempenho de previsão superior a outros modelos lineares consolidados na literatura, principalmente em horizontes de curto-prazo.*

1. Introdução

No âmbito de previsões de séries temporais (ou, simplificadamente, séries), diversos métodos para modelagem requerem que a série seja estacionária, *i.e.*, possua média, variância e covariância constantes. Entretanto, para a maioria delas, a condição de não-estacionariedade é a regra enquanto a estacionariedade é a exceção. O trabalho seminal para modelagem integrada de séries resilientes à não-estacionariedade veio a partir do processo *Autoregressive Integrated Moving Average* ($ARIMA$), que usa a diferenciação como forma de obter a estacionariedade [Box et al., 2008].

*Os autores agradecem à FAPERJ, à CAPES (código 001) e ao CNPq pelo financiamento do projeto.

Posteriormente, outras abordagens de transformação de séries não-estacionárias em estacionárias surgiram e vêm sendo frequentemente associadas ao estado-da-arte em modelos de aprendizado de máquina [Salles et al., 2019]. No entanto, o uso direto dessas novas abordagens aplicadas a modelos de aprendizado de máquina pode ocultar as propriedades matemáticas e estatísticas dos dados, em virtude do ajustamento por hiperparâmetros, comumente aplicado no processo de construção de tais modelos. Em contraste, essas propriedades são bem estabelecidas e podem ser melhor observadas e interpretadas no contexto de modelos lineares. Este fato favorece o seu uso quando a análise do comportamento das séries temporais é um fator de interesse.

Neste contexto, este artigo introduz um novo modelo linear, denominado $ARAI$ (do inglês, *Autoregressive Adaptive Integration*), que combina os modelos autorregressivos com a normalização adaptativa [Ogasawara et al., 2010]. O modelo inova ao viabilizar o uso do $ARIMA$ em cenários de janelas deslizantes, comumente adotados no contexto de aprendizado de máquina. Cada janela preserva as propriedades da série original, e permite acompanhar o seu aspecto inercial. A análise deste modelo pode também contribuir para o emprego adequado da normalização adaptativa no cenário de aprendizado de máquina. O modelo foi avaliado com base na previsão de séries temporais socieconômicas e foi capaz de apresentar desempenho superior quando comparado a outros modelos lineares consolidados na literatura, como AR, MA e $ARIMA$, principalmente em horizontes de curto-prazo.

2. Referencial Teórico

Uma série pode ser compreendida como um processo estocástico composto por uma coleção y de variáveis randômicas y_t, tal que $|y| = n$ e y_n corresponde à observação mais recente. A série é classificada como estritamente estacionária se, para qualquer $(t_1, t_2, ..., t_k)$ e para qualquer h, a distribuição $(y_{t_1}, y_{t_2}, ..., y_{t_k})$ for idêntica à $(y_{t_{1+h}}, y_{t_{2+h}}, ..., y_{t_{k+h}})$ [Brockwell and Davis, 2016].

Essa condição é rígida e geralmente difícil de ser observada matematicamente, haja vista que, para a maioria das aplicações, as respectivas distribuições são desconhecidas a priori [Woodward et al., 2017]. Por causa disso, uma definição menos restritiva, denominada covariância-estacionária, foi estabelecida, em que uma série é considerada estacionária se possuir média, variância e covariância constantes.

2.1. ARIMA

Modelos $ARIMA(p, d, q)$ são compostos pelos processos de modelagem autorregressivo (AR) e média móvel (MA) (respectivamente representados por p e q). Além disso, tais modelos contam com um processo de diferenciação preliminar (I) (representado por d), desenvolvidos para lidar com a não estacionariedade. Dentro da família $ARIMA$, o processo autorregressivo $AR(p)$ estabelece que o valor corrente de uma série é função de seus p valores precedentes. Tal processo é descrito pela Equação 1, em que δ corresponde ao intercepto, p à ordem do modelo AR e ε_t ao erro, que idealmente deve se comportar como um ruído branco [Brockwell and Davis, 2016].

$$\mathrm{y}_t = \delta + \sum_{j=1}^{p} \phi_j \mathrm{y}_{t-j} + \varepsilon_t \tag{1}$$

Tradicionalmente, a metodologia usada para descrever o processo $ARIMA(p, d, q)$ é a Box-Jenkins, composta pelas fases de identificação, estimação e validação. A ordem d estabelece o nível de diferenciação a ser aplicado a priori na série temporal. Os valores das ordens p e q, quando aplicáveis, baseiam-se na interpretação dos gráficos das funções de autocorrelação e autocorrelação parcial [**?**].

No caso proposto por Brockwell and Davis [2016], o critério sugerido é a combinação p e q que minimize o valor da função AICc [Hurvich and Tsai, 1989]. Essa função é uma versão aprimorada do critério de informação de Akaike (AIC) [Akaike, 1992], que penaliza o resultado da função original a fim de evitar introdução de viés. Idealmente, quanto menor o valor da função AICc, melhor. Dessa maneira, o modelo se ajusta cada vez mais à série à medida que o valor da função se aproxima de $-\infty$ [Enders, 2015]. Assim, para cada combinação p e q, os coeficientes do processo $ARIMA$ são estimados de forma a minimizar o valor da referida função. Diante disso, a fim de auxiliar na automação do processo, Hyndman and Khandakar [2008] desenvolveram um algoritmo que otimiza a escolha dos parâmetros p, d e q do modelo $ARIMA(p, d, q)$.

2.2. Mapeamento do Processo AR como Regressão Múltipla

Uma subsequência é uma amostra contínua de uma série. A i-ésima subsequência de tamanho p em uma série y, representada por $seq_{p,i}(y)$, é uma sequência ordenada de valores $(y_i, y_{i+1}, \ldots, y_{i+p-1})$, onde $|seq_{p,i}(y)| = p$ e $1 \leq i \leq |y| - p$.

As janelas deslizantes de tamanho p consistem em explorar todas as subsequências de tamanho p de uma série. Formalmente, podem ser representadas por $sw_p(y)$, que corresponde a uma matriz A de tamanho $(|y| - p + 1) \times p$. Cada vetor-linha a_i em A é a i-ésima subsequência de tamanho p em y. Dado $A = sw_p(y)$, $\forall a_i \in A$, $a_i = seq_{p,i}(y)$. Vale observar que as janelas deslizantes organizam as colunas da matriz A de forma que a j-ésima coluna corresponda a uma defasagem da série original y por $(p - j)$ valores precedentes.

Considere uma matriz A formada por janelas deslizantes de tamanho $p + 1$ ($A = sw_{p+1}(y)$). Considere o mapeamento da matriz A por meio de uma relação $\mathcal{R}(A)$, na qual, os vetores-colunas a_j de A podem ser mapeados como uma dimensão x_j $(1 \leq j \leq p+1)$ em $\mathcal{R}(A)$. A dimensão x_{p+1} corresponde às observações da série enquanto as dimensões x_j $(\forall j \in [1, p])$ aos seus valores defasados. A Equação 1 pode ser reescrita por meio da Equação 2, de modo que o modelo autorregressivo AR pode ser interpretado como um problema de regressão múltipla, *i.e.*, os coeficientes $\phi_1, \cdots \phi_p$ podem ser aproximados pelos métodos de mínimos quadrados ou verosimilhança adotados no cálculo de regressão múltipla. O δ e ϵ_{p+1} são análogos aos da Equação 1, onde neste caso, ϵ_{p+1} corresponde ao conjunto de erros observados em toda a série em x_{p+1}, ou $\epsilon_{p+1} = (\varepsilon_{p+1}, \cdots, \varepsilon_{|y|})$ [Tsay, 2010].

$$x_{p+1} = \delta + \phi_1 x_1 + \cdots + \phi_p x_p + \epsilon_{p+1} \tag{2}$$

3. ARAI

O processo de construção do modelo ARAI é composto por três fases sequenciais: transformação da série temporal, remoção de *outliers* e modelagem.

Seja a matriz A formada por janelas deslizantes de tamanho $p+1$ sobre uma série y de tamanho n. Seja $a_i = (a_{i_1}, \cdots, a_{i_{p+1}})$ o *i*-ésimo vetor-linha ($1 \leq i \leq (|y|-p+1)$). A primeira fase consiste em transformar a matriz A em uma matriz B normalizada adaptativamente. Para cada $a_i \in A$, define-se a operação básica do $ARAI$ como uma diferenciação a qual se aplica uma função f computada a partir dos termos endógenos $(a_{i_1}, \cdots, a_{i_p})$, como descrito na Equação 3. A função f, descrita na Equação 4, é uma função de centralidade (média) sobre um vetor v. A função traz um aspecto inercial de acompanhamento da série. A função f aplicada sobre todas as linhas a_i produz uma matriz B normalizada adaptativamente [Ogasawara et al., 2010]. Ao final deste processo, a matriz B possui média 0 e variância σ^2. Tais propriedades foram observadas experimentalmente.

$$b_{i_j} = a_{i_j} - f(a_{i_1}, \cdots, a_{i_p}), \forall j \in [1, p+1] \tag{3}$$

$$f(v) = \frac{\sum_{j=1}^{|v|}(v_j)}{|v|} \tag{4}$$

Logo após a primeira fase, os *outliers* são removidos, desconsiderando os vetores-linha da matriz B nos quais um ou mais valores estejam fora do intervalo $[Q_1 - 1.5(IQR), Q3 + 1.5(IQR)]$. Este critério de identificação de *outliers* é comumente utilizado por *boxplots*, em que IQR corresponde ao intervalo interquartil.

A partir deste momento, tem-se a terceira fase (modelagem), onde encontram-se os coeficientes $\phi_1, \cdots, \phi_p$ que ajustam a Equação 2 ao se converter a matriz B em uma relação $\mathcal{R}(B)$. Uma vez previsto o valor $b_{i_{p+1}}$ proveniente das observações $b_{i_1}, \cdots, b_{i_p}$, basta somar o valor de $f(a_{i_1}, \cdots, a_{i_p})$ para se obter o valor $a_{i_{p+1}}$.

4. Avaliação Experimental

Para a avaliação do modelo $ARAI$, foram obtidas as quatro séries mensais mais usadas do IPEAData (http://www.ipeadata.gov.br), conforme descritas pela Tabela 1. A ordem p do $ARAI(p)$ é definida variando-se de 2 a 12 (do menor tamanho de janela possível ao número de meses que compõem um ano), definindo como modelo final aquele que obtiver o menor valor para o $AICc$ para um conjunto de treinamento. Os modelos foram avaliados usando como métrica o erro quadrático médio (MSE) para horizontes de 1 a 12 (um ano de previsão).

Tabela 1. Séries analisadas pelo processo ARAI

Série - periodicidade mensal	Treino	Teste
Salário mínimo real	07/1994 - 03/2018	04/2018 - 03/2019
Índice de Preços ao Consumidor Amplo (IPCA)	07/1994 - 03/2018	04/2018 - 03/2019
Índice de desemprego	12/1984 - 02/2018	03/2018 - 02/2019
Produto Interno Bruto (PIB)	07/1994 - 03/2018	04/2018 - 03/2019

A Figura 1 apresenta o MSE das previsões feitas por ($ARAI$, $ARIMA$, AR e MA) para as quatro séries analisadas por horizonte de previsão. No caso do salário mínimo real (Figura 1.a), o $ARAI$ apresentou acurácia superior ao $ARIMA$ até o horizonte

9. A partir deste ponto, a acurácia degenerou ficando próxima aos modelos AR e MA. De modo análogo, no caso do índice de desemprego (Figura 1.b), o $ARAI$ apresentou melhor acurácia até horizonte 8. A partir do mês 9, a acurácia do $ARAI$ ficou semelhante à dos processos AR, MA e $ARIMA$.

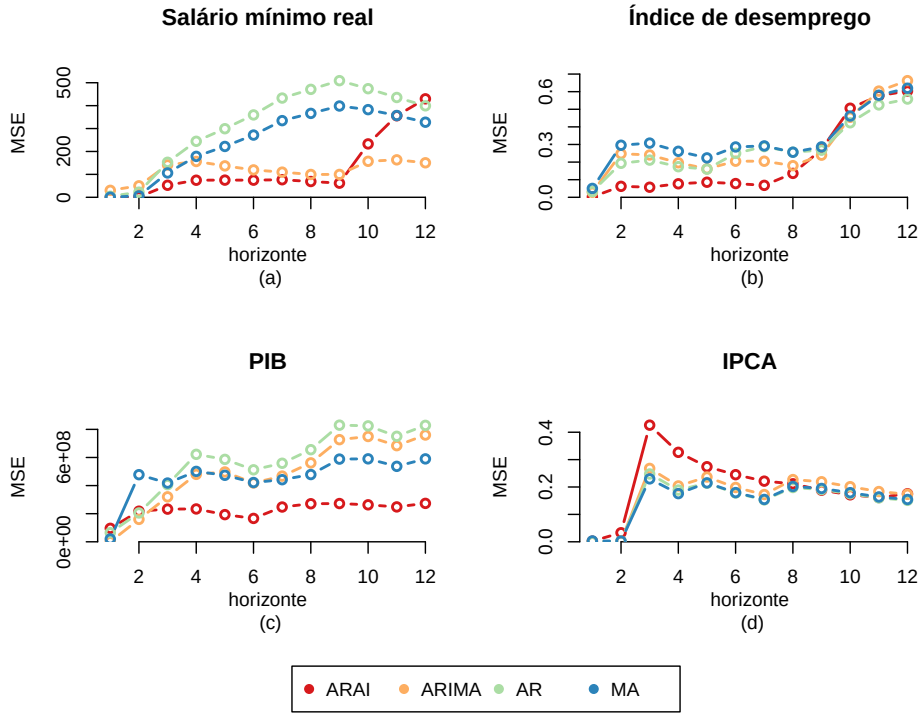
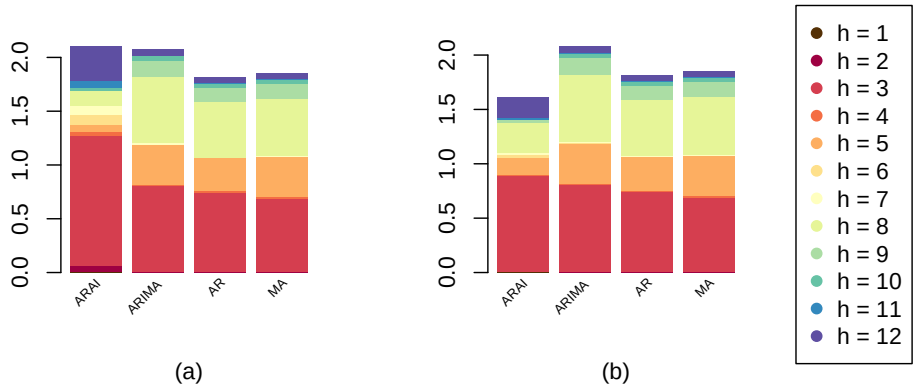


Figura 1. Séries avaliadas com os respectivos MSE

No caso do PIB (Figura 1.c), o $ARAI$ foi dominante, apresentando melhor acurácia para quase todos os horizontes estudados. Em contrapartida, no caso do IPCA (Figura 1.d) para o horizonte de tamanho 3, o $ARAI$ teve uma acurácia inferior. Conforme o horizonte foi aumentando, as perdas provenientes da observação 3 são recuperadas, aproximando-se dos demais modelos.

Em função deste comportamento no IPCA, foi feita uma análise adicional. Observou-se que, no mês de junho de 2018 (observação 3), a inflação foi 1,26%, impactada pela greve dos caminhoneiros. Nos doze meses precedentes, a inflação média foi 0,235% com desvio padrão de 0,17%, sendo a mínima e a máxima de -0,23% e 0,44%, respectivamente. A Figura 2.a apresenta a contribuição individual do erro ao quadrado para cada previsão. Nota-se que para todos os modelos, a acurácia da previsão foi afetada pela observação 3, sendo mais impactante no $ARAI$. Entretanto, o $ARAI$ produziu previsões competitivas para as observações de 4 a 9. Nos demais modelos, as previsões tiveram acurácia inferior para as observações deste intervalo.

Estudando-se mais profundamente a previsão da observação 3, identificou-se que o *outlier* afetou o valor inercial usado para normalização. Por conta disto, avaliou-se em vez da média a mediana como função inercial (Equação 4). Neste caso, o impacto do *outlier* é minimizado, alcançando-se previsões próximas àquelas registradas pelos processos AR, MA e $ARIMA$ (Figura 2.b), sem deteriorar as previsões entre 4 e 9, o que fez esta adaptação no $ARAI$ apresentar melhor desempenho em relação aos demais modelos.



(a) (b)

Figura 2. Erro ao quadrado associado ao respectivo horizonte para a série IPCA

5. Conclusão

Neste artigo foi apresentado um modelo linear Autorregressivo de Integração Adaptativa. Uma característica interessante desta abordagem é que, diferentemente dos processos clássicos de diferenciação no qual a série transformada é bem diferente da série original, o formato da série em cada janela preserva semelhanças em relação à série original, uma vez que a série é subtraída em relação à sua função inercial.

O modelo abre espaço para o estudo das propriedades estatísticas da normalização adaptativa no contexto de modelos lineares, o que pode auxiliar no desenvolvimento do seu emprego no cenário de aprendizado de máquina. Como trabalhos futuros, tem-se o estudo da acurácia das previsões do $ARAI$ em horizontes maiores e a prova de que os erros ϵ_{p+1} do modelo $ARAI$ se caracterizam como ruído branco.

Referências

Akaike, H. (1992). Information Theory and an Extension of the Maximum Likelihood Principle. In Kotz, S. and Johnson, N. L., editors, *Breakthroughs in Statistics: Foundations and Basic Theory*, Springer Series in Statistics, pages 610–624. Springer New York, New York, NY.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control*. Wiley, Hoboken, N.J, 4 edition.

Brockwell, P. J. and Davis, R. A. (2016). *Introduction to Time Series and Forecasting*. Springer International Publishing, Cham, 3 edition.

Enders, W. (2015). *Applied Econometric Time Series*. Wiley, 4 edition.

Hurvich, C. and Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307.

Hyndman, R. and Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3):1–22.

Ogasawara, E., Martinez, L., De Oliveira, D., Zimbrão, G., Pappa, G., and Mattoso, M. (2010). Adaptive Normalization: A novel data normalization approach for non-stationary time series. In *Proceedings of the International Joint Conference on Neural Networks*.

Salles, R., Belloze, K., Porto, F., Gonzalez, P., and Ogasawara, E. (2019). Nonstationary time series transformation methods: An experimental review. *Knowledge-Based Systems*, 164:274–291.

Tsay, R. S. (2010). *Analysis of Financial Time Series*. Wiley, Cambridge, Mass, 3 edition.

Woodward, W. A., Gray, H. L., and Elliott, A. C. (2017). *Applied Time Series Analysis with R*. CRC Press, Boca Raton, 2 edition.