

Previsão de Carga Baseada em Ensemble de Modelos Inteligentes

Domingos S. de O. Santos Júnior*, João F. L. de Oliveira†, Hugo Siqueira§, Manoel Marinho†, Francisco Madeiro†, Marcos de Almeida Leone Filho¶, Paulo S. G. de Mattos Neto*

* Centro de Informática, Universidade Federal de Pernambuco, Recife-PE, Brasil

† Universidade de Pernambuco, Recife-PE, Brasil

§ Universidade Tecnológica Federal do Paraná - Ponta Grossa-PR, Brasil

¶ Venidera Pesquisa e Desenvolvimento - Campinas-SP, Brasil

E-mails: dsosj@cin.ufpe.br, jflo@poli.br, hugosiqueira@utfpr.edu.br

marinho75@poli.br, madeiro@poli.br, marcos@venidera.com, psgmn@cin.ufpe.br

Resumo—A previsão de demanda de carga é uma etapa fundamental no planejamento e operação do sistema elétrico nacional, em virtude da minimização dos custos que ela pode trazer. Isto é particularmente importante tendo em vista que outras etapas também são dependentes de tal previsão, como análise do fluxo de potência, despacho econômico e estabilidade do sistema. Tradicionalmente, modelos únicos são utilizados para solução deste problema. Entretanto, esta abordagem pode alcançar uma configuração que leve a um baixo desempenho na modelagem de séries temporais reais. Neste contexto, neste trabalho é proposta a aplicação de um *ensemble* evolucionário de preditores com o objetivo de maximizar o desempenho dos preditores individuais. O método proposto lança mão de um Algoritmo Genético (AG) para treinar automaticamente os componentes do *ensemble*, utilizando um processo de validação cruzada para geração de diversidade. São discutidas duas versões que empregam, respectivamente, redes neurais do tipo Perceptron de Múltiplas Camadas (MLP) e Máquina de Vetores de Regressão (SVR) na aplicação de quatro séries temporais, as quais correspondem à demanda de carga diária dos subsistemas brasileiros. A análise experimental considerando três diferentes horizontes de previsão mostra que o *ensemble* evolucionário alcançou maior acurácia que modelos tradicionais da Estatística e da Computação Inteligente.

Index Terms—Previsão de carga, redes neurais artificiais, *ensemble*.

I. INTRODUÇÃO

A previsão de demanda de carga é de grande relevância para os agentes da operação de sistemas elétricos de potência. A operação é fortemente dependente desta informação para o correto funcionamento de geradores, subestações e linhas [1], [2]. Através desta tarefa, os agentes do setor elétrico podem estimar a compra de energia baseada em demandas futuras e preços, minimizando a diferença entre a quantidade comprada e consumida [3]. Portanto, o aumento da acurácia de previsão está ligada à redução dos custos da operação [4].

A previsão supracitada é objeto de intensa investigação devido aos seus impactos financeiros e ambientais no planejamento dos agentes e governos [1]. Ademais, estimativas futuras precisas podem ser usadas para evitar interrupções no fornecimento de energia e para o cálculo do fluxo de potência [1].

Tanto a superestimação quanto a subestimação da previsão de carga podem prejudicar o planejamento da transmissão e da

distribuição de sistemas elétricos, o que leva à necessidade de modelos de previsão precisos [5], [6]. Nesse sentido, diversas pesquisas têm abordado dados de carga elétrica utilizando modelos lineares (da Estatística) e não-lineares (da Computação Inteligente) para previsão de séries temporais com o objetivo de desenvolver sistemas mais confiáveis [7]–[11].

Métodos lineares como regressão linear, suavização exponencial de Holt-Winters, filtros de Kalman e os modelos da família Box & Jenkins, como o Autorregressivo (AR) e o Autorregressivo Integrado de Médias Móveis (ARIMA) são fortemente utilizados neste contexto [12], [13]. Os parâmetros destes modelos são determinados a partir de dados históricos de cargas para realizar as estimações futuras para novas entradas [1]. Contudo, estes estão limitados a modelar apenas relações lineares nos dados temporais, o que pode ser insuficiente do ponto de vista da acurácia [14].

Em contrapartida, com o desenvolvimento nas últimas décadas da inteligência computacional, modelos não-lineares como as redes neurais artificiais (RNAs) também têm sido considerados para previsão de cargas [15]. Tais métodos têm apresentando resultados satisfatórios em diversos contextos [2], [8], [16]. Em [8] os autores propõem a utilização de uma rede recorrente de Elman otimizada pelo algoritmo Otimização por Enxame de Partículas (*Particle Swarm Optimization* – PSO) para a tarefa. Abreu *et al.* abordam o modelo *Fuzzy ARTMAP* com treinamento continuado [1]. Além disso, na literatura também são encontrados diversos sistemas híbridos, como forma de melhorar a qualidade das previsões dos modelos individuais [3], [7], [16]. Singh e Dwivedi [16] desenvolveram um método composto por algoritmos multi-objetivos e RNA para simultaneamente minimizar o erro de previsão e a variabilidade estatística das previsões.

O *ensemble* é um método que combina as saídas de diversos preditores com vistas a melhorar o desempenho geral. Neste estudo, um *ensemble* evolucionário é proposto para a previsão da demanda elétrica para os quatro sub-sistemas brasileiros. O modelo é desenvolvido com o objetivo de alcançar previsões com maior acurácia, visto que: (i) realiza o ajuste automático dos modelos inteligentes que o compõem utilizando um Algoritmo Genético (AG); (ii) utiliza partições de dados

independentes para treinar os modelos gerando diversidade; (iii) combina as previsões desses modelos usando a média.

Duas versões da abordagem são construídas utilizando redes neurais do tipo Perceptron de Múltiplas Camadas (*Multilayer Perceptron* – MLP) e Máquina de Vetores de Regressão (*Support Vector Regression* – SVR) como componentes do *ensemble*. A metodologia proposta é avaliada em três horizontes de previsão (um, dois e três dias à frente) utilizando duas métricas de desempenho bem conhecidas na literatura. Os resultados obtidos são comparados aos dos modelos únicos pertencentes a literatura de previsão de séries temporais: ARIMA [12], MLP [10], máquinas de aprendizado extremo (*Extreme Learning Machine* – ELM) [17] e SVR [18].

Os dados reais consistem na demanda diária dos 4 subsistemas delimitados pelo Operador Nacional do Sistema Elétrico (ONS): Centro-Oeste/Sudeste, Nordeste, Norte e Sul. Este é um problema de difícil tratamento e solução, já que o comportamento de demanda de carga é influenciado por diversos fatores, tais como questões regionais, economia, clima, dias de consumo atípicos como feriados e até efeitos aleatórios [3].

O restante do artigo está organizado como segue: a Seção II apresenta os modelos utilizados para realizar a tarefa de previsão; o método proposto é apresentado na Seção III; a Seção IV descreve o protocolo experimental e os resultados das simulações; por fim, na Seção V as conclusões são apresentadas e possíveis trabalhos futuros são mencionados.

II. BACKGROUND

Nesta Seção descrevem-se as metodologias aplicadas ao problema.

A. Modelo Autorregressivo e Médias Móveis Integrado

Modelos lineares são largamente aplicados à modelagem e previsão das mais variadas séries temporais. São relativamente fáceis de implementar e seu custo computacional é baixo se comparado a métodos não-lineares.

Um dos mais conhecidos é o Autorregressivo (AR(p)), que é definido como a combinação linear dos p atrasos (*lags*) da observação x_t , ou seja, x_{t-p} sendo $p = 1, \dots, P$, com a adição de um ruído branco gaussiano a_t , como mostra a Equação 1 [12]:

$$\tilde{x}_t = \phi_1 \tilde{x}_{t-1} + \dots + \phi_p \tilde{x}_{t-p} + \dots + \phi_P \tilde{x}_{t-P} + a_t, \quad (1)$$

em que $\tilde{x}_t = x_t - \mu$, sendo μ a média amostral e ϕ_p é o coeficiente de ponderação do atraso p .

Observa-se que o termo a_t é considerado o erro inerente ao processo de regressão [19].

O ajuste e otimização do AR consiste, na verdade, em encontrar os melhores valores dos coeficientes ϕ_p que minimizem o erro entre os dados observados e a saída oferecida pelo modelo. Neste caso, tal tarefa possui solução determinística no sentido do erro quadrático médio, as conhecidas Equações de Yule-Walker [15].

Já o modelo de Médias Móveis, ao contrário do AR, combina sinais de ruído branco [12]. Um modelo MA de ordem q - (MA(q)) é dado pela Expressão 2:

$$\tilde{x}_t = -\theta_1 a_{t-1} - \dots - \theta_q a_{t-q} - \dots - \theta_Q a_{t-Q} + a_t, \quad (2)$$

sendo θ_q os parâmetros livres e $q \in \{1, 2, \dots, Q\}$,

Entretanto, diferentemente do AR, para o ajuste e otimização do MA é necessária a solução de um sistema não-linear de equações, não havendo fórmula fechada para seu cálculo. A literatura é vasta em mostrar que o uso da função de Máxima Verossimilhança é o caminho mais usual para isso [12], [15].

Por fim, o modelo Autorregressivo e de Médias Móveis de ordem p, q (ARMA(p, q)) é dado pela união entre o AR e o MA. Para realizar previsões utilizando o ARMA um termo de diferenciação d para retirada da componente de tendência pode ser introduzido ao mesmo, gerando o Autorregressivo de Médias Móveis Integrado (ARIMA (p, d, q)) [12].

Na aplicação direta do ARIMA, é possível modelar os choques aleatórios do MA por meio do erro da previsão do passo anterior a cada período de tempo, ou seja, $\epsilon_t = a_t$. A Equação 3 exhibe a formulação matemática do mesmo:

$$\tilde{x}_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} + \epsilon_t \quad (3)$$

sendo os parâmetros do modelo similares aos previamente discutidos.

Neste trabalho, são definidas as ordens do ARIMA pela metodologia de Hyndman e Khandakar [20].

B. Perceptron de Múltiplas Camadas

O Perceptron de Múltiplas Camadas (MLP) é, sem dúvidas, a arquitetura de redes neurais artificiais (RNA) mais difundida da literatura [21], sendo largamente aplicada em problemas como aproximação de funções, previsão, reconhecimento de padrões, classificação e otimização de sistemas [22]. Trata-se de um grupo de neurônios artificiais disposto em múltiplas camadas, de forma que os sinais são propagados da entrada para saída [22], [23].

O treinamento de uma rede MLP consiste no ajuste sistemático dos pesos, de forma que a minimização do erro seja alcançada. O algoritmo do gradiente descendente [24], [25] é um método tradicional para atualização dos pesos da rede, dado por 4:

$$w_{ij}^m(t+1) = w_{ij}^m(t) - \alpha \frac{\partial E(t)}{\partial w_{ij}^m(t)} \quad (4)$$

em que $w_{ij}^m(t)$ é o peso sináptico da entrada i para o neurônio j da camada m na iteração t e $\partial E(t)$ é a derivada parcial em relação ao erro, geralmente o erro quadrático médio.

O cálculo da derivada é realizado pelo conhecido algoritmo *backpropagation* [22], [24], [25].

O treinamento consiste em duas fases. Inicialmente, os dados de entrada são propagados pela rede até ser encontrada uma saída. Estes valores são comparados aos desejados (observados) para obtenção do erro. Num segundo momento, o caminho inverso é percorrido, da saída para a entrada. Neste caso, todos os pesos são atualizados de acordo com a equação

4, de modo que uma determinada métrica de erro seja reduzida a cada iteração [22].

C. Máquinas de Vetores de Regressão

As Máquinas de Vetores de Regressão (*Support Vector Regression - SVR*) [18] possuem como característica a convexidade do problema de otimização durante seu treinamento, gerando um único mínimo global. O objetivo é encontrar uma função na forma de 5:

$$\{f|f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b, w \in \mathbb{R}^d, b \in \mathbb{R}\} \quad (5)$$

em que w é um vetor de pesos estimado através da otimização da função de risco regularizado apresentado na Equação 6 e b é o bias.

Na Equação 6, $C > 0$ é um fator de regularização e $L(\dots)$ é uma função de custo:

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l L(y_i, f(\mathbf{x}_i)). \quad (6)$$

Em problemas de regressão, são produzidas saídas contínuas para os dados de treinamento e é desejado que desviem no máximo de ε do valor esperado. Portanto, a função de custo L pode ser definida como mostrado na Equação 7.

$$L(y, f(\mathbf{x})) = \begin{cases} 0, & |f(\mathbf{x}) - y| < \varepsilon \\ |f(\mathbf{x}) - y| - \varepsilon, & \text{Caso contrário.} \end{cases} \quad (7)$$

É possível observar que a função de custo não é computada para valores que estejam dentro da região formada por ε . Portanto, a SVR pode ser formulada de acordo com a Expressão 8, na qual ξ e ξ^* são variáveis de folga utilizadas para avaliar erros de valores estimados que ultrapassam a região delimitada por ε .

$$\begin{aligned} \text{Minimizar} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i - \xi_i^*) \\ \text{Restrições} \quad & \begin{cases} \mathbf{w}^T \mathbf{x}_i + b - y_i \leq \varepsilon + \xi_i \\ y_i - \mathbf{w}^T \mathbf{x}_i - b \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 0 \dots l \end{cases} \end{aligned} \quad (8)$$

A utilização de funções de kernel k possibilita o mapeamento não-linear dos dados para um espaço de características no qual uma função linear é encontrada. Dessa forma, a regressão é realizada através da Equação 9, em que α_i e α_i^* são as variáveis de Lagrange que possuem valor nulo para exemplos que se encontram dentro da região delimitada por ε .

$$f(\mathbf{x}) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) k(\mathbf{x}_i, \mathbf{x}_j) + b. \quad (9)$$

As principais diferenças do SVR em relação a RNAs estão na convexidade do problema de otimização citado anteriormente, já que as RNAs, em geral, possuem vários mínimos

loais na função objetivo otimizada. É a utilização de funções de kernel que possibilita o mapeamento não-linear dos dados para um espaço de características, no qual uma função linear é encontrada. A sensibilidade na seleção de parâmetros também apresenta diferenças importantes nestes casos [26].

D. Máquinas de Aprendizado Extremo

Máquinas de Aprendizado Extremo (ELM) [17] são arquiteturas de redes neurais *feedforward*, com uma camada intermediária. A principal diferença entre esta rede e a tradicional MLP é que os pesos sinápticos da camada intermediária são definidos de forma aleatória e permanecem sem ajuste durante o treinamento [17].

Dessa maneira, calibrar uma ELM consiste em determinar os pesos da matriz que representa camada de saída \mathbf{W}^{out} , de modo a gerar o menor erro quadrático médio (MSE) entre a saída da rede e o sinal desejado \mathbf{d} . O ponto importante é que esta tarefa pode ser resolvida por uma solução de mínimos quadrados, sobretudo com a aplicação do operador de pseudo-inversão de Moore-Penrose [27]. Esta proposta garante o mínimo MSE e tem solução determinística, o que confere à rede pequeno esforço computacional relativo. Tal solução é dada por

$$\mathbf{W}^{\text{out}} = (\mathbf{X}_{\text{hid}}^T \mathbf{X}_{\text{hid}})^{-1} \mathbf{X}_{\text{hid}}^T \mathbf{d}, \quad (10)$$

em que \mathbf{X}_{hid} é a matriz contendo as saídas dos neurônios da camada escondida relativo ao conjunto de treinamento e m é o número de neurônios na camada de saída.

Este operador garante também a norma mínima dos pesos da camada de saída, o que tende a elevar o poder de generalização da rede [17].

III. MÉTODO PROPOSTO

Na Figura 1 é apresentado o esquemático do *ensemble* evolucionário proposto para a previsão de cargas. O desempenho de um *ensemble* depende da acurácia individual dos modelos e também da sua diversidade [28]. Assim, tal método tem como objetivo otimizar os hiper-parâmetros dos modelos assim como preservar a diversidade entre eles. A otimização de parâmetros está ligada à minimização dos erros de previsão e é realizada através de um algoritmo genético (AG). Por outro lado, a diversidade é alcançada através do treinamento de modelos em diferentes subconjuntos dos dados.

O método proposto é dividido em duas fases: treinamento e teste. A primeira pode ser dividida em dois passos: divisão em subconjuntos e busca e treinamento dos parâmetros dos componentes do *ensemble* pelo AG. A fase de teste é, também, dividida em duas etapas: previsão de novos exemplos (conjunto de teste) utilizando os n modelos treinados e a posterior combinação das saídas dos modelos, gerando a previsão final.

A. Divisão em subconjuntos

Como mencionado, um dos aspectos importantes no desenvolvimento de sistemas de múltiplos preditores é a geração de modelos diversos. Nesse sentido, o AG ajusta cada modelo componente do *ensemble* utilizando um subconjunto único

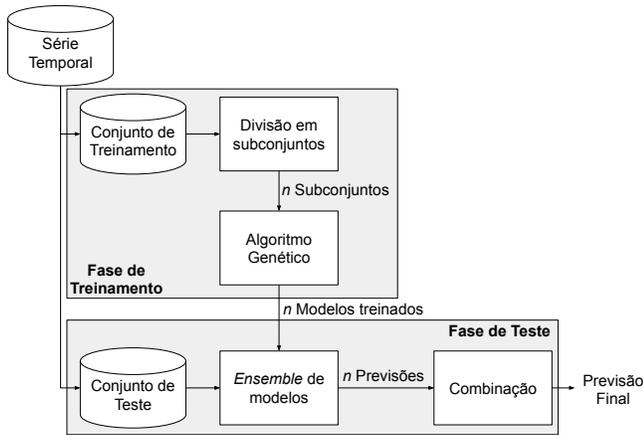


Figura 1. Método proposto.

proveniente dos dados de treinamento. Estes subconjuntos são gerados a partir da validação cruzada ordenada no tempo [29], como exemplificado na Figura 2, a qual usa quatro *folds* ($n = 4$) [22]. Neste caso, *fit* é a variável relativa ao *fitness* dos agentes do algoritmo genético, a ser apresentado na próxima seção.

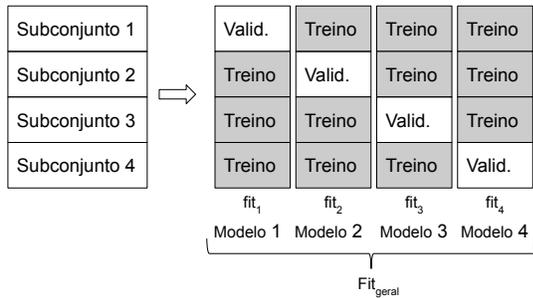


Figura 2. Exemplo da validação cruzada com 4 folds.

A abordagem de validação cruzada divide o conjunto de treinamento em n partições mutuamente excludentes, de modo que uma partição é usada para validação e o restante para treinamento do modelo. Assim, n possibilidades para divisão entre os conjuntos de treinamento e validação são geradas.

B. Algoritmo Genético

O desempenho de modelos inteligentes na previsão de séries temporais depende do correto ajuste de seus parâmetros. Porém, esta tarefa é custosa já que possui um espaço de busca grande para a maioria dos parâmetros. Nesse sentido, um Algoritmo Genético é utilizado com o objetivo de encontrar os coeficientes mais adequados para prever uma dada série temporal de interesse. Através de operadores de seleção, cruzamento e mutação a população inicial é evoluída até que um critério de parada seja atingido [30].

Neste trabalho foram adotados dois modelos de combinadores não-lineares para o *ensemble*, uma MLP e uma SVR. Na Tabela I podem ser vistos os hiper-parâmetros otimizados pelo

AG para ambos modelos. Para o primeiro caso (MLP_{ESB}) são otimizados: taxa de aprendizado (α), número de neurônios escondidos e a semente utilizada para gerar os pesos iniciais. No segundo (SVR_{ESB}), são determinados: o fator de regularização C , o parâmetro do *kernel* RBF γ , a margem de tolerância ε e a tolerância de erro como critério de parada.

Tabela I
LIMITE MÁXIMO E MÍNIMO DOS HIPER-PARÂMETROS OTIMIZADOS PELO ALGORITMO GENÉTICO

Modelo	Parâmetro	Limite Mínimo	Limite Máximo
MLP_{ESB}	Taxa de aprendizado	0,01	1
	Neurônios	1	100
	Semente	0	9,0E+10
SVR_{ESB}	C	0,1	10000
	γ	0,0001	0,1
	ε	0,001	0,9
	Tolerância	0,00001	0,01

A capacidade de generalização de cada modelo treinado é medida a partir da função de avaliação (*fitness*) considerando o respectivo Erro Percentual Absoluto Médio (MAPE) da validação. Cada modelo é treinado com $n - 1$ *folds* e avaliado com o *fold* de validação. Dessa forma, cada modelo do *ensemble* é treinado com diferentes partes dos dados, visando gerar modelos diversos.

A função de *fitness* parte da suposição de que na tarefa de previsão de séries temporais os padrões mais recentes têm mais relevância que padrões mais antigos [29]. Baseado nesta suposição, a função de *fitness* geral do *ensemble* é dada pelas Equações 11 e 12

$$fitness_{geral} = \sum_{i=1}^n w_i fitness_i, \quad (11)$$

$$w_j = \begin{cases} 1 - \sum_{i=2}^n w_i, & j = 1 \\ \frac{1}{2^{n+1-j}}, & j \neq 1, \end{cases} \quad (12)$$

em que $fitness_i$ é a função de *fitness* (MAPE da validação) do modelo i e w_i é o seu respectivo peso.

O *fitness* do *ensemble*, denotado por $fitness_{geral}$, é calculado a partir dos $fitness_i$ dos n modelos, sendo definido, neste caso, para um problema de minimização.

C. Combinação dos Modelos

Terminada a fase de seleção de hiper-parâmetros realizada pelo AG, o *ensemble* que obteve o menor $fitness_{geral}$ é selecionado para prever o conjunto de teste. Para a combinação das previsões de teste dos n modelos do *ensemble* é utilizada a média. A média foi escolhida para essa fase, pois estratégias como softmax, mediana, média e regressão linear possuem no geral um desempenho similar como reportado em [29].

IV. PROTOCOLO EXPERIMENTAL E RESULTADOS

Neste trabalho são utilizados dados de demandas de carga diária dos 4 subsistemas brasileiros: Centro-Oeste/Sudeste, Nordeste, Norte e Sul. Os dados estão disponíveis no site do Operador Nacional do Sistema Elétrico (ONS) [31]. As

amostras correspondem aos registros diários de 1 de fevereiro de 2017 a 09 de julho de 2019, totalizando 889 dados para cada subsistema. Na Tabela II estão descritos os valores de média e desvio padrão para cada série.

Tabela II
MÉDIA E DESVIO PADRÃO DA DEMANDA DE CARGA PARA CADA SUBSISTEMA

Subsistema	Média (MWh)	Desvio padrão
Centro-Oeste/Sudeste	8.7745E+05	9.0126E+04
Norte	1.2927E+05	6.6126E+03
Nordeste	2.4677E+05	1.7238E+04
Sul	2.6659E+05	3.5026E+04

Os primeiros 709 valores são usados para treinamento e validação dos modelos e os últimos 180 registros (aproximadamente 6 meses) na fase de teste. Os resultados das duas versões do método proposto (MLP_{ESB} e SVR_{ESB}) são comparados a modelos comumente utilizados na literatura de previsão: ARIMA, ELM, MLP e SVR. Foram escolhidas as últimas 180 amostras de treinamento para validação da ELM, MLP e SVR.

A seleção dos retardos temporais foi realizada por meio de uma busca exaustiva de até 20 entradas. Destas, as 14 primeiras foram selecionadas para todos os modelos e para os quatro subsistemas estudados.

A avaliação da acurácia dos modelos foi realizada utilizando o erro quadrático médio (*mean square error* - MSE) e o erro absoluto percentual médio (*mean absolute percentage error* - MAPE). Para ambas as métricas, quanto menor o valor, melhor é o desempenho do modelo de previsão. Trabalhos correlatos mostram que nem sempre métricas distintas levam a conclusões semelhantes em problemas de previsão, sendo sempre relevante comparar mais de uma [15], [27].

O *ensemble* evolucionário proposto é avaliado em três horizontes de previsão distintos ($P = 1, 2$ e 3 passos à frente), totalizando 12 casos de estudo. A previsão de múltiplos passos é realizada pelo método direto [15], ou seja os modelos são treinados diretamente no horizonte P desejado e as previsões também são feitas de acordo com o horizonte selecionado [32].

O AG utiliza uma população de 50 indivíduos inicializados aleatoriamente. Dois pais são selecionados, gerando dois filhos por cruzamento linear. A mutação é feita por uma perturbação gaussiana, com probabilidade de 20%. A seleção de sobreviventes é feita utilizando um torneio.

Os modelos MLP_{ESB} e SVR_{ESB} foram implementados utilizando o sklearn [33] e o DEAP [34] foi usado para a implementação do AG. Os modelos únicos foram elaborados em Matlab.

A. Resultados e discussão

Nas Tabelas III, IV, V e VI estão sumarizados os desempenhos em termos de MSE e MAPE para os quatro subsistemas estudados. Os menores erros estão destacados em negrito.

Inicialmente, nota-se que em 3 casos, as métricas de desempenho foram divergentes em apontar qual preditor foi o melhor. Este comportamento não é raro em problemas de previsão [27]. Assim, este trabalho adotará o MAPE como

medida absoluta, já que é a de maior interesse do sistema elétrico brasileiro.

De forma geral, o modelo ARIMA apresentou maiores valores de erro na maior parte dos experimentos em relação aos modelos não-lineares devido à sua limitação em mapear apenas padrões lineares [14].

Na Tabela III podem ser vistos os resultados das métricas para o subsistema Centro-Oeste/Sudeste. A MLP obteve o melhor MSE para todos os horizontes de previsão estudados, enquanto a SVR_{ESB} obteve o melhor MAPE para $P = 1$ e 3 passos adiante.

Nas Tabelas IV e V podem ser analisados os desempenhos para os subsistemas Nordeste e Norte, respectivamente. Em ambos, o método proposto alcançou a melhor acurácia. A versão MLP_{ESB} foi superior para 1 passo à frente enquanto a SVR_{ESB} obteve o menores erros para $P = 2$ e 3 .

Os resultados das métricas para o subsistema Sul são exibidos na Tabela VI. A MLP_{ESB} alcançou menores erros para 1 passo à frente. Para $P = 2$, a MLP_{ESB} e a SVR_{ESB} obtiveram os melhores MSE e MAPE, respectivamente. O modelo único SVR chegou a melhores valores para 3 passos adiante.

Tabela III
RESULTADOS PARA O SUBSISTEMA CENTRO-OESTE/SUDESTE

Modelo	$P = 1$		$P = 2$		$P = 3$	
	MSE	MAPE	MSE	MAPE	MSE	MAPE
ARIMA	1.4324E+9	3,078	2.8163E+9	4,631	3.5874E+9	5,214
MLP	9,9351E+8	2,510	2,2339E+9	4,056	3,0814E+9	4,848
SVR	1.1004E+9	2,530	3.3286E+9	4,740	3.6990E+9	5,004
ELM	1.4907E+9	3,016	2.7215E+9	4,405	3.5705E+9	5,085
MLP_{ESB}	1.0802E+9	2,558	2.9118E+9	4,452	4.5885E+9	5,403
SVR_{ESB}	1.0749E+9	2,499	2.5056E+9	4,247	3.1124E+9	4,836

Tabela IV
RESULTADOS PARA O SUBSISTEMA NORDESTE

Modelo	$P = 1$		$P = 2$		$P = 3$	
	MSE	MAPE	MSE	MAPE	MSE	MAPE
ARIMA	7.4513E+7	2,424	1.2011E+8	3,208	1.3845E+8	3,507
MLP	6.8288E+7	2,427	1.3576E+8	3,548	1.3079E+8	3,446
SVR	5.6070E+7	2,097	1.3075E+8	3,420	1.5937E+8	3,816
ELM	7.1563E+7	2,456	1.0841E+8	3,125	1.2742E+8	3,450
MLP_{ESB}	5,2684E+7	2,045	1.1806E+8	3,245	1.2406E+8	3,392
SVR_{ESB}	5.3505E+7	2,030	9,8329E+7	2,847	1,1465E+8	3,142

Tabela V
RESULTADOS PARA O SUBSISTEMA NORTE

Modelo	$P = 1$		$P = 2$		$P = 3$	
	MSE	MAPE	MSE	MAPE	MSE	MAPE
ARIMA	1.4068E+7	2,234	1.9508E+7	2,599	2.1283E+7	2,679
MLP	1.3527E+7	2,137	1.7688E+7	2,432	1.9795E+7	2,600
SVR	1.2691E+7	2,046	1.7334E+7	2,407	1.9302E+7	2,579
ELM	1.2287E+7	2,038	1.6875E+7	2,411	1.8672E+7	2,470
MLP_{ESB}	1,0923E+7	1,914	1.7197E+7	2,397	2.0120E+7	2,633
SVR_{ESB}	1.1745E+7	1,962	1,6638E+7	2,350	1,7802E+7	2,497

A Tabela VII mostra o número de vezes em que o melhor desempenho foi alcançado por cada modelo em termos do MAPE. Na maioria dos cenários, o *ensemble* evolucionário proposto alcançou os menores erros. Dos 12 cenários, a versão SVR_{ESB} obteve o melhor resultado em 7 e a MLP_{ESB} em 3. Quando 1 passo à frente é considerado, a MLP_{ESB} obtém o melhor desempenho geral, e nos demais (2 e 3) a SVR_{ESB} .

Tabela VI
RESULTADOS PARA O SUBSISTEMA SUL

Modelo	$P = 1$		$P = 2$		$P = 3$	
	MSE	MAPE	MSE	MAPE	MSE	MAPE
ARIMA	6,4201E0+8	7,711	8,7886E0+8	8,675	8,9737E0+8	8,419
MLP	1,7293E0+8	3,273	3,9265E0+8	5,464	4,6492E0+8	5,904
SVR	1,9581E0+8	3,386	4,0019E0+8	5,154	4,2065E0+8	5,526
ELM	2,4054E0+8	3,822	4,2294E0+8	5,465	4,7310E0+8	5,962
MLP _{ESB}	1,4605E0+8	2,954	3,5079E0+8	5,094	5,0134E0+8	6,146
SVR _{ESB}	1,7744E0+8	3,408	3,5493E0+8	5,069	5,3816E0+8	6,244

Tabela VII
RANKING DOS MELHORES DESEMPENHOS EM TERMOS DE MAPE

Model	P = 1	P = 2	P = 3	Total
MLP	—	1	—	1
SVR	—	—	1	1
MLP _{ESB}	3	—	—	3
SVR _{ESB}	1	3	3	7

Nas Figuras 3, 4, 5 e 6 podem ser vistas as previsões do *ensemble* evolucionário proposto para 1 passo adiante, considerando cada subsistema. Apesar dos resultados discutidos anteriormente, observa-se que em determinados dias as previsões obtidas pelo método proposto são menos precisas, como por exemplo entre os dias 10 e 20 na Figura 3.

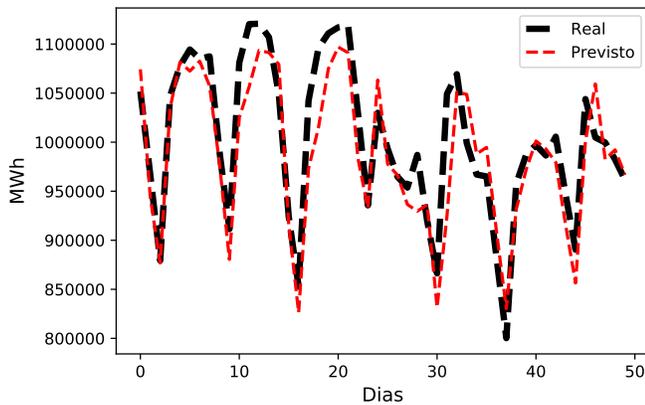


Figura 3. Previsão da SVR_{ESB} nos primeiros 50 pontos de teste no subsistema Centro-Oeste/Sudeste ($P = 1$).

V. CONCLUSÃO

A previsão de demanda de energia elétrica constitui-se em um problema importante no planejamento energético e possui impactos econômicos e ecológicos no planejamento governamental. Uma forma de colaborar nesta difícil tarefa é desenvolver sistemas de previsão precisos, de forma que os custos de operação sejam minimizados.

Nesse contexto, neste trabalho são propostos *ensembles* evolucionários baseados na otimização de modelos ajustados em diferentes conjuntos dos dados para solução do problema. Neste caso, os parâmetros são otimizados por um algoritmo genético, enquanto que a diversidade entre eles é promovida pelo uso da validação cruzada tipo $k - fold$. O método proposto foi aplicado à previsão diária de carga nos 4 subsistemas brasileiros: Centro-oeste/Sudeste, Nordeste, Norte e Sul. Os

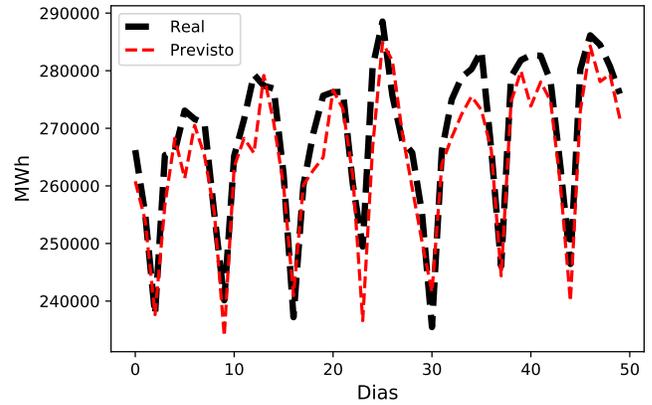


Figura 4. Previsão da MLP_{ESB} nos primeiros 50 pontos de teste no subsistema Nordeste ($P = 1$).

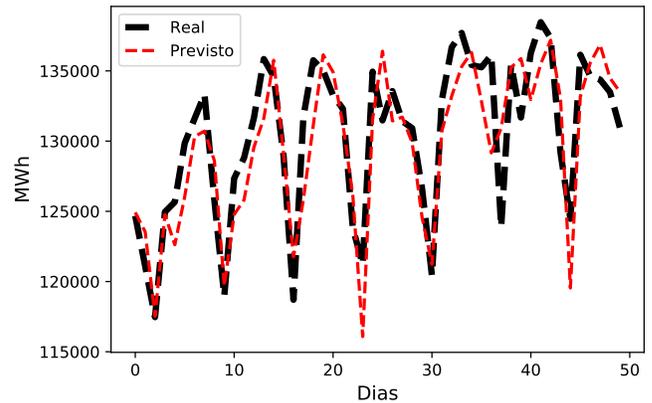


Figura 5. Previsão da MLP_{ESB} nos primeiros 50 pontos de teste no subsistema Norte ($P = 1$).

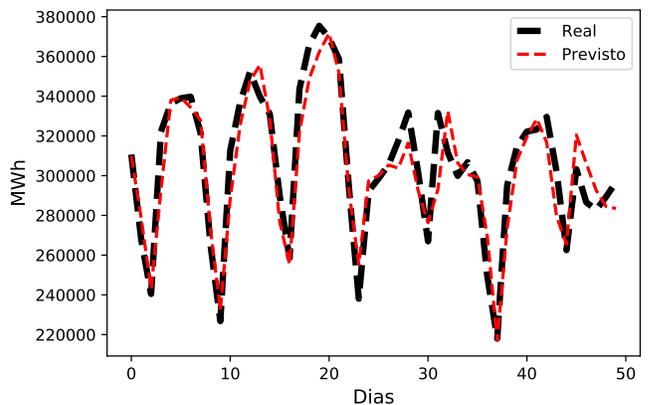


Figura 6. Previsão da MLP_{ESB} nos primeiros 50 pontos de teste no subsistema Sul ($P = 1$).

dados considerados para a simulação são disponibilizados pelo Operador Nacional do Sistema Elétrico (ONS). O desempenho foi avaliado em três horizontes de previsão.

Experimentos também foram realizados utilizando métodos únicos, como ARIMA, MLP, SVR e ELM. Os resultados mostram que o método proposto obteve resultados promissores, chegando aos melhores resultados em 10 dos 12 cenários estudados. Foi observado que os modelos não-lineares obtiveram melhores resultados em relação ao linear ARIMA.

Como trabalhos futuros, a previsão dos resíduos [23] gerados pelo *ensemble* pode ser explorada como método de correção para as previsões originais, com vistas à melhoria da precisão final do sistema. Além disso, técnicas de pré-processamento dos dados também poderão ser empregadas.

AGRADECIMENTO

Este trabalho recebeu financiamento e suporte técnico da AES e Companhias Associadas (do grupo CPFL, Brookfield e Globa) como parte do projeto ANEEL PD-0610-1004/2015, "IRIS - Integração de Renováveis Intermitentes: Um modelo de simulação da operação do sistema elétrico brasileiro para apoio ao planejamento, operação, comercialização e regulação", que é parte de um programa P&D regulado pela ANEEL, Brasil. Os autores também agradecem ao IATI - Instituto Avançado de Tecnologia e Inovação pelo suporte. Além disso, o mesmo também foi parcialmente financiado com recursos do Conselho Nacional de Desenvolvimento Científico e Tecnológico, projeto universal número 405580/2018-5.

REFERÊNCIAS

- [1] T. Abreu, C. S. Júnior, M. Lopes, A. D. Lotufo, and C. Minussi, "Previsão de cargas elétricas utilizando uma rede neural ARTMAP fuzzy com treinamento continuado," in *Anais XIII Brazilian Congress on Computational Intelligence (CBIC)*, 2017.
- [2] P. Bento, J. Pombo, M. Calado, and S. Mariano, "Optimization of neural network with wavelet transform and improved data selection using bat algorithm for short-term load forecasting," *Neurocomputing*, vol. 358, pp. 53–71, 2019.
- [3] G. C. Silva, A. C. Lisboa, D. A. Vieira, C. L. Castro, and R. R. Saldanha, "Modelos de previsão de carga: Abordagens imunoinspiradas baseadas em similaridade de padrões," in *Anais XIII Brazilian Congress on Computational Intelligence (CBIC)*, 2017.
- [4] B. F. Hobbs, S. Jitrapaikulsarn, S. Konda, V. Chankong, K. A. Loparo, and D. J. Maratukulam, "Analysis of the value for unit commitment of improved load forecasts," *IEEE Transactions on Power Systems*, vol. 14, no. 4, pp. 1342–1348, 1999.
- [5] K. Goswami, A. Ganguly, and A. K. Sil, "Day ahead forecasting and peak load management using multivariate auto regression technique," in *2018 IEEE Applied Signal Processing Conference (ASPCON)*. IEEE, 2018, pp. 279–282.
- [6] V. Dehalwar, A. Kalam, M. L. Kolhe, and A. Zayegh, "Electricity load forecasting for urban area using weather forecast information," in *2016 IEEE International Conference on Power and Renewable Energy (ICPRE)*. IEEE, 2016, pp. 355–359.
- [7] Y. Wang, Q. Chen, M. Sun, C. Kang, and Q. Xia, "An ensemble forecasting method for the aggregated load with subprofiles," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3906–3908, 2018.
- [8] K. Xie, H. Yi, G. Hu, L. Li, and Z. Fan, "Short-term power load forecasting based on Elman neural network with particle swarm optimization," *Neurocomputing*, 2019.
- [9] P. M. Macaira, R. C. Sousa, and F. L. C. Oliveira, "Forecasting Brazil's electricity consumption with pegels exponential smoothing techniques," *IEEE Latin America Transactions*, vol. 14, no. 3, pp. 1252–1258, 2016.
- [10] P. P. Balestrassi, E. Popova, A. Paiva, and J. M. Lima, "Design of experiments on neural network's training for nonlinear time series forecasting," *Neurocomputing*, vol. 72, no. 4-6, pp. 1160–1178, 2009.
- [11] A. Yang, W. Li, and X. Yang, "Short-term electricity load forecasting based on feature selection and least squares support vector machines," *Knowledge-Based Systems*, vol. 163, pp. 159–173, 2019.
- [12] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [13] T. M. O'Donovan, *Short term forecasting: An introduction to the Box-Jenkins approach*. John Wiley & Sons, Inc., 1983.
- [14] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.
- [15] H. Siqueira, L. Boccatto, R. Attux, and C. Lyra, "Unorganized machines for seasonal streamflow series forecasting," *International Journal of Neural Systems*, vol. 24, no. 03, p. 1430009, 2014.
- [16] P. Singh and P. Dwivedi, "A novel hybrid model based on neural network and multi-objective optimization for effective load forecast," *Energy*, vol. 182, pp. 606 – 622, 2019.
- [17] G. H. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics – PART B: Cybernetics*, vol. 42, no. 2, pp. 513–529, 2012.
- [18] V. Vapnik, *The nature of statistical learning theory*. Springer Science & Business Media, 2013.
- [19] S. O. Haykin, *Adaptive filter theory*. Pearson, 2013.
- [20] R. J. Hyndman and Y. Khandakar, *Automatic time series for forecasting: the forecast package for R*. Monash University, Department of Econometrics and Business Statistics, 2007, no. 6/07.
- [21] G. Polezer, Y. S. Tadano, H. V. Siqueira, A. F. Godoi, C. I. Yamamoto, P. A. de André, T. Pauliquevis, M. F. Andrade, A. Oliveira, P. H. Saldiva, P. Taylor, and R. Godoi, "Assessing the impact of pm2. 5 on respiratory disease using artificial neural networks," *Environmental pollution*, vol. 235, pp. 394–403, 2018.
- [22] S. S. Haykin, *Neural networks and learning machines*. New York: Prentice Hall, 2009.
- [23] P. S. de Mattos Neto, F. Madeiro, T. A. Ferreira, and G. D. Cavalcanti, "Hybrid intelligent system for air quality forecasting using phase adjustment," *Engineering Applications of Artificial Intelligence*, vol. 32, pp. 185 – 191, 2014.
- [24] P. Werbos, "Beyond regression: New tools for prediction and analysis in the behavioral sciences," *Ph. D. thesis, Harvard University*, 1974.
- [25] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Cognitive Modeling*, vol. 5, no. 3, p. 1, 1986.
- [26] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee, "Choosing multiple parameters for support vector machines," *Machine Learning*, vol. 46, no. 1-3, pp. 131–159, 2002.
- [27] H. Siqueira, L. Boccatto, I. Luna, R. Attux, and C. Lyra, "Performance analysis of unorganized machines in streamflow forecasting of Brazilian plants," *Applied Soft Computing*, vol. 68, pp. 494–506, 2018.
- [28] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Machine learning*, vol. 51, no. 2, pp. 181–207, 2003.
- [29] J. P. Donate, P. Cortez, G. G. Sánchez, and A. S. De Miguel, "Time series forecasting using a weighted cross-validation evolutionary artificial neural network ensemble," *Neurocomputing*, vol. 109, pp. 27–32, 2013.
- [30] H. Siqueira, R. Attux, and C. Lyra Filho, "Exploração de alternativas lineares para previsão de séries de vazões," *Mecânica Computacional XXIX*, vol. 2010, pp. 9629–9644, 2010.
- [31] ONS - Operador Nacional do Sistema Elétrico, "Carga horária por subsistema," 2019. [Online]. Available: <http://sdro.ons.org.br/SDRO/DIARIO/index.htm>
- [32] J. T. Belotti, I. Luna, and H. V. Siqueira, "Avaliação de abordagens alternativas de previsão no estudo de series de vazões," *Journal of Applied Instrumentation and Control*, vol. 6, no. 1, pp. 19–26, 2019.
- [33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [34] F.-A. Fortin, F.-M. De Rainville, M.-A. Gardner, M. Parizeau, and C. Gagné, "DEAP: Evolutionary algorithms made easy," *Journal of Machine Learning Research*, vol. 13, pp. 2171–2175, jul 2012.