



Universidade Federal de Pernambuco
Centro de Informática

Graduação em Engenharia da Computação

**Sistema inteligente de seleção dinâmica
baseado em janelas temporais mais
próximas para previsão de séries
temporais**

Gabriel Mendes Matos

Trabalho de Graduação

Recife
27 de abril de 2023

Universidade Federal de Pernambuco
Centro de Informática

Gabriel Mendes Matos

Sistema inteligente de seleção dinâmica baseado em janelas temporais mais próximas para previsão de séries temporais

Trabalho apresentado ao Programa de Graduação em Engenharia da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação.

Orientador: *Prof. Paulo S. G. de Mattos Neto*

Recife
27 de abril de 2023

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Matos, Gabriel Mendes.

Sistema inteligente de seleção dinâmica baseado em janelas temporais mais próximas para previsão de séries temporais / Gabriel Mendes Matos. - Recife, 2023.

42 : il., tab.

Orientador(a): Paulo Salgado Gomes de Mattos Neto

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Informática, Engenharia da Computação - Bacharelado, 2023.

1. Inteligência Artificial. 2. Previsão de Séries Temporais. 3. Seleção Dinâmica. 4. Redes Neurais. 5. Aprendizagem de Máquina. I. de Mattos Neto, Paulo Salgado Gomes. (Orientação). II. Título.

000 CDD (22.ed.)

Agradecimentos

Gostaria de expressar minha sincera gratidão a todas as pessoas que me ajudaram e me apoiaram durante a minha jornada acadêmica.

Em primeiro lugar, gostaria de agradecer ao meu professor orientador, Paulo Salgado, por sua orientação, incentivo e paciência ao longo deste trabalho. Suas contribuições foram inestimáveis para a conclusão deste projeto e para o meu crescimento acadêmico.

Gostaria também de agradecer aos meus amigos Jefferson, Rafael, Ferreira, Mariana, Jucival, Pedro, Filipe, Mitterrand, Mateus e Nathan. Obrigado por me apoiarem em momentos de dúvida e por me ajudarem a superar desafios durante a universidade. Sua amizade e companheirismo foram fundamentais para tornar essa jornada mais leve e divertida.

Não posso deixar de agradecer meus pais e minha irmã, que sempre estiveram ao meu lado, me dando suporte e encorajamento durante esse tempo. Sem o amor e apoio de vocês, essa conquista não seria possível.

Por fim, agradeço a todos aqueles que de alguma forma contribuíram para o meu sucesso acadêmico, seja através de um simples gesto ou de um apoio mais significativo. Vocês foram fundamentais para que eu chegasse até aqui.

Obrigado!

Resumo

As séries temporais de aplicações reais apresentam padrões que mudam ao longo do tempo, tornando-as difíceis de prever usando apenas um modelo de previsão. Dessa forma, abordagens de seleção dinâmica têm se destacado na literatura devido à sua acurácia e capacidade de modelar diferentes padrões locais. Essas abordagens selecionam um ou mais modelos de um *pool* (ou comitê) para prever cada padrão de teste. Essa seleção é realizada com base no desempenho do *pool* em uma Região de Competência (RoC), um conjunto de amostras mais semelhante a um padrão de teste. A definição da RoC, a criação do *pool*, o número de modelos selecionados e a função de combinação são questões críticas para as abordagens de seleção dinâmica, uma vez que sua precisão está intimamente relacionada a elas. Esse trabalho propõe um sistema de seleção dinâmica baseado em um *pool* heterogêneo que realiza uma escolha baseada em dados para determinar: (i) o melhor tamanho da RoC, (ii) o conjunto dos modelos de previsão mais competentes e (iii) a função de combinação mais adequada. A seleção usa uma RoC composta pelas janelas antecedentes mais próximas a um padrão de teste. A proposta emprega um *pool* heterogêneo composto por seis modelos de previsão, dois estatísticos (modelos *Autoregressive Integrated Moving Average* (ARIMA) e Theta) e quatro modelos de aprendizagem de máquina (*Support Vector Regression* (SVR), *Multilayer Perceptron* (MLP), *Extreme Learning Machine* (ELM) e *Long Short-Term Memory* (LSTM)). Uma análise experimental realizada com sete conjuntos de dados conhecidos mostrou que a proposta superou as abordagens de comitês e de modelos individuais da literatura, indicando que é capaz de realizar uma seleção dinâmica melhor.

Palavras-chave: Previsão de Séries Temporais, Seleção Dinâmica, Comitê Heterogêneo, Aprendizagem de Máquina, Redes Neurais

Abstract

Real-world time series present patterns that change over time, making them difficult to forecast using only one forecasting model. In this way, dynamic selection approaches have been highlighted in literature due to their accuracy and ability to model different local patterns. These approaches select one or more models from a pool (or ensemble) to forecast each test pattern. This selection is performed based on the pool's performance in a Region of Competence (RoC), a set of samples most similar to a test pattern. The RoC definition, the pool creation, the number of selected models, and the combination function are critical issues for the dynamic selection approaches once their accuracy is closely related to them. In this paper, we propose a dynamic selection system based on a heterogeneous pool that performs a data-driven choice to determine: (i) the best RoC size, (ii) the set of the most competent forecasting models, and (iii) the most suitable combination function. The selection uses an RoC composed of the nearest antecedent windows to a test pattern. The proposal employs a heterogeneous pool comprising six forecasting models, two statistical (Autoregressive Integrated Moving Average (ARIMA) and Theta models), and four Machine Learning models (Support Vector Regression (SVR), Multilayer Perceptron (MLP), Extreme Learning Machine (ELM) and Long Short-Term Memory (LSTM)). An experimental analysis performed using seven well-known data sets showed that the proposal overcame literature single and ensemble approaches, indicating that it is able to perform a better dynamic selection.

Keywords: Time Series Forecasting, Dynamic Selection, Heterogeneous Ensemble, Machine Learning, Neural Networks

Sumário

| | | |
|----------|---|-----------|
| 1 | Introdução | 1 |
| 1.1 | Motivação | 1 |
| 1.2 | Objetivos | 2 |
| 1.3 | Estrutura do Trabalho | 2 |
| 2 | Fundamentação | 4 |
| 3 | Trabalhos Relacionados | 7 |
| 4 | Sistema Proposto | 8 |
| 4.1 | Visão Geral | 8 |
| 4.2 | Modelos do Comitê | 9 |
| 4.2.1 | <i>Autoregressive Integrated Moving Average (ARIMA)</i> | 9 |
| 4.2.2 | Modelo Theta | 11 |
| 4.2.3 | <i>Support Vector Regression (SVR)</i> | 11 |
| 4.2.4 | <i>Multilayer Perceptron (MLP)</i> | 12 |
| 4.2.5 | <i>Extreme Learning Machine (ELM)</i> | 12 |
| 4.2.6 | <i>Long Short-Term Memory (LSTM)</i> | 12 |
| 5 | Protocolo Experimental | 14 |
| 6 | Resultados | 20 |
| 7 | Conclusão | 30 |

Lista de Figuras

| | | |
|-----|--|----|
| 2.1 | Padrões encontrados em séries temporais. | 4 |
| 2.2 | Série temporal e respectivo <i>plot</i> ACF. | 5 |
| 2.3 | Séries temporais resultantes de uma transformação por diferenciação. | 6 |
| 4.1 | Visão geral do sistema proposto. | 9 |
| 4.2 | Etapas de Seleção Dinâmica e Combinação. | 10 |
| 5.1 | <i>Plot</i> das séries temporais utilizadas no experimento. | 15 |
| 5.2 | <i>Plot</i> ACF das séries temporais do experimento. | 16 |
| 6.1 | Razão de performance entre as abordagens de seleção dinâmica e seus oráculos em termos da Equação 5.5. | 23 |
| 6.2 | Tempos de execução em escala logarítmica. | 25 |
| 6.3 | <i>Plot</i> de previsões para a série temporal <i>Electricity</i> . | 26 |
| 6.4 | <i>Plot</i> de previsões para a série temporal <i>Goldman</i> . | 27 |
| 6.5 | <i>Plot</i> de previsões para a série temporal <i>Pollution</i> . | 27 |
| 6.6 | <i>Plot</i> de previsões para a série temporal <i>Star</i> . | 28 |
| 6.7 | <i>Plot</i> de previsões para a série temporal <i>Sunspot</i> . | 28 |
| 6.8 | <i>Plot</i> de previsões para a série temporal <i>Vehicle</i> . | 29 |
| 6.9 | <i>Plot</i> de previsões para a série temporal <i>Wine</i> . | 29 |

Lista de Tabelas

| | | |
|-----|---|----|
| 5.1 | Frequências de sazonalidade estimadas para cada série temporal, medidas em passos de tempo. | 14 |
| 5.2 | Espaço de hiperparâmetros dos modelos que compõem o comitê do sistema proposto. | 17 |
| 5.3 | Parâmetros de Seleção Dinâmica para o sistema proposto definidos baseado na performance no conjunto de validação. | 18 |
| 6.1 | Resultados de performance para os métodos avaliados. Valores em negrito representam os melhores resultados para cada par de série temporal e métrica. | 21 |
| 6.2 | Ranking das abordagens avaliadas baseado nas métricas de erro. | 21 |
| 6.3 | Diferença percentual (% MSE) de acordo com a Equação 5.4. | 22 |
| 6.4 | Porcentagem dos padrões de teste onde um modelo do comitê foi escolhido para participar da previsão final (%), de acordo com a Equação 5.6. | 23 |
| 6.5 | Resultados de performance para os oráculos criados a partir de cada comitê. | 24 |

Introdução

1.1 Motivação

A tarefa de prever séries temporais está presente em muitas aplicações, como finanças [1], saúde [2], clima [3], engenharia [4] e astronomia [5]. Dessa forma, o desenvolvimento de sistemas de previsão precisos para modelagem de séries temporais do mundo real tem sido um tópico de pesquisa relevante [6, 7, 8]. A modelagem de séries temporais de aplicações reais é desafiadora porque os dados geralmente apresentam padrões diferentes ao longo do tempo [9]. A literatura [10, 11, 12] reforça essa afirmação e sugere que o emprego de um único modelo geralmente leva a uma precisão inferior. Isso ocorre porque é desafiador determinar os parâmetros apropriados de um modelo de previsão para todo o conjunto de dados, o que pode resultar em modelos mal especificados ou tendenciosos [13]. A área de sistemas de múltiplos preditores (MPS) tem se destacado neste contexto devido aos resultados teóricos e práticos. Esses sistemas empregam um comitê de previsores com o objetivo de modelar diferentes padrões temporais da série temporal [14]. O MPS consiste em três fases sequenciais: geração, seleção e combinação. Durante a fase de geração, um *pool* de modelos é criado a partir do conjunto de treinamento. Na fase de seleção, um ou mais modelos de previsão são escolhidos com base em critérios específicos. Na fase final, as saídas dos modelos selecionados são combinadas para gerar a previsão final.

O MPS pode ser classificado em duas categorias com base em sua etapa de seleção [11], estático e dinâmico. A primeira classe seleciona os modelos na fase de treinamento (offline), enquanto a segunda executa a escolha do modelo na etapa de teste (online). Na classe estática, um ou mais modelos são escolhidos para prever todo o conjunto de dados. Na classe dinâmica, um ou mais modelos são selecionados para prever cada padrão de teste. Assim, abordagens baseadas na seleção dinâmica escolhem um subconjunto de modelos do *pool* para prever cada ponto de teste. Esta estratégia de seleção é mais adequada para modelar o comportamento dinâmico comumente presente em séries temporais do mundo real [15].

A precisão de uma abordagem de seleção dinâmica está intimamente relacionada à qualidade do *pool*, à definição da Região de Competência (RoC), à quantidade de modelos selecionados para prever um novo padrão de teste e à função usada para combinar essas previsões [11, 15, 16]. Os previsores que compõem o *pool* devem ser capazes de modelar características complementares. Assim, o comitê deve ser preciso e diverso para modelar diferentes padrões locais [17]. Após a geração do *pool*, um ou mais modelos são selecionados com base em sua competência para prever um novo padrão de teste. A definição da RoC é uma etapa importante porque determina quantos e quais padrões de séries temporais serão usados para medir os modelos mais competentes do *pool* [15]. Essa região é composta pelos k padrões

temporais passados que são mais semelhantes a um determinado padrão de teste de acordo com algum critério, por exemplo, distância euclidiana. Após a criação da RoC, um ou mais modelos de previsão podem ser selecionados com base em sua precisão [11]. Se apenas um modelo ($n = 1$) for escolhido, este modelo é responsável por prever o padrão de teste; caso contrário, as previsões dos n modelos ($n > 1$) precisam ser combinadas usando alguma função para gerar a saída final.

Neste trabalho é proposto um sistema de seleção dinâmica que supõe que as janelas antecedentes mais próximas a um padrão de teste são mais promissoras para a criação da RoC do que as mais distantes [11]. O sistema proposto emprega um *pool* heterogêneo com duas técnicas estatísticas (modelos *Autoregressive Integrated Moving Average* (ARIMA) e Theta) e quatro modelos de aprendizagem de máquina (*Support Vector Regression* (SVR), *Multilayer Perceptron* (MLP), *Extreme Learning Machine* (ELM) e *Long Short-Term Memory* (LSTM)). Por fim, o MPS proposto define quantos, quais e como os modelos serão combinados para cada conjunto de dados.

Uma análise experimental utilizando sete séries temporais do mundo real é conduzida no cenário de um passo à frente. O desempenho do sistema de seleção dinâmica proposto é comparado a modelos individuais e abordagens de seleção dinâmica da literatura em termos de três métricas bem conhecidas: *Mean Squared Error* (MSE), *Mean Absolute Percentage Error* (MAPE), e *Average Relative Prediction Error Variation* (ARV).

1.2 Objetivos

O objetivo geral desse trabalho é propor e avaliar um sistema de seleção dinâmica para previsão de séries temporais capaz de realizar uma busca baseada em dados para maximizar sua acurácia a partir de um comitê heterogêneo de modelos.

Este trabalho tem como objetivos específicos:

- Especificar o funcionamento do sistema, incluindo a definição da região de competência, os parâmetros da seleção dinâmica e os modelos que compõem o comitê;
- Implementar o sistema proposto numa linguagem de programação;
- Conduzir um experimento utilizando conjuntos de dados de séries temporais do mundo real;
- Analisar os resultados obtidos comparando o sistema a outras abordagens da literatura.

1.3 Estrutura do Trabalho

O restante deste trabalho está organizado da seguinte forma: Capítulo 2 introduz conceitos importantes acerca de séries temporais para o entendimento desse trabalho. O Capítulo 3 apresenta trabalhos relacionados em seleção dinâmica para previsão de séries temporais. O Capítulo 4 descreve o sistema de seleção dinâmica proposto. O Capítulo 5 detalha os experimentos realizados para avaliar e comparar o método proposto e as abordagens anteriores. O Capítulo 6

apresenta e discute os resultados do experimento em termos de desempenho e custo computacional. O Capítulo 7 traz considerações finais e delinea trabalhos futuros.

Fundamentação

Séries temporais são sequências de observações sobre uma variável coletadas ao longo do tempo em intervalos geralmente regulares (horas, dias, etc) [18], podendo ser geradas por processos naturais, como variações de temperatura e observações astronômicas, ou criadas por humanos, como índices de mercado e demanda de recursos elétricos. Sua formação pode depender não só de valores passados da própria série mas também de variáveis externas.

Muitas séries temporais apresentam padrões bem estudados na literatura [19], apresentados a seguir:

- **Tendência:** são aumentos ou diminuições a longo prazo na série, representam a direção geral na qual ela se move. Pode ser visto na Figura 2.1(a).
- **Sazonalidade:** são variações periódicas que acontecem em frequências bem definidas, como em certos dias da semana ou meses do ano. Pode ser visto na Figura 2.1(b).
- **Ciclo:** são flutuações a longo prazo que, diferente da sazonalidade, não apresentam comportamento bem definido, como variações de frequência e amplitude.

Quando presentes, esses padrões podem ser interpretados como componentes de uma série temporal, e removidos através de decomposição. Nesse caso, a série pode ser representada por um componente de sazonalidade, um componente de tendência-ciclo, e outro com o resíduo da série, o qual é formado por movimentos aleatórios e ruídos [20]. A decomposição pode ser aditiva ou multiplicativa, como mostram as Equações 2.1 e 2.2, respectivamente, nas quais S_t representa o componente de sazonalidade, T_t o componente de tendência-ciclo, e R_t o componente residual. Esse processo é útil para analisar a série temporal e também para realizar

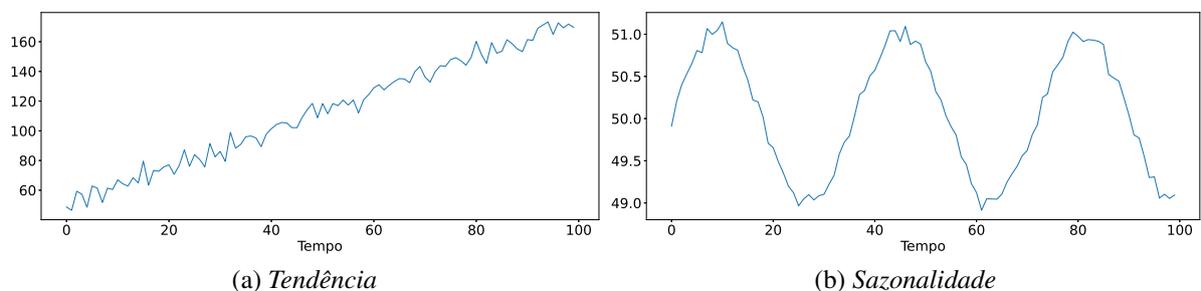


Figura 2.1 Padrões encontrados em séries temporais.

sua previsão, uma vez que há diferentes abordagens para modelar cada componente e prever valores futuros através da combinação das previsões de cada parte.

$$x_t = S_t + T_t + R_t. \quad (2.1)$$

$$x_t = S_t \times T_t \times R_t. \quad (2.2)$$

Uma forma de investigar a relação entre cada observação de uma série temporal e seus valores passados é através da função de autocorrelação (ACF) [19]. Ela mede a correlação linear entre as observações da série e seus atrasos. A Figura 2.2 mostra um exemplo de série temporal e seu *plot* ACF, no qual cada ponto é o valor do coeficiente de autocorrelação para um valor de atraso diferente. Através dele, é possível identificar diferentes padrões na série. No caso da tendência, os coeficientes tendem a ser grandes e diminuir lentamente. Já para sazonalidade, eles são maiores para múltiplos da frequência. Na série temporal do exemplo é possível ver esses dois efeitos em conjunto.

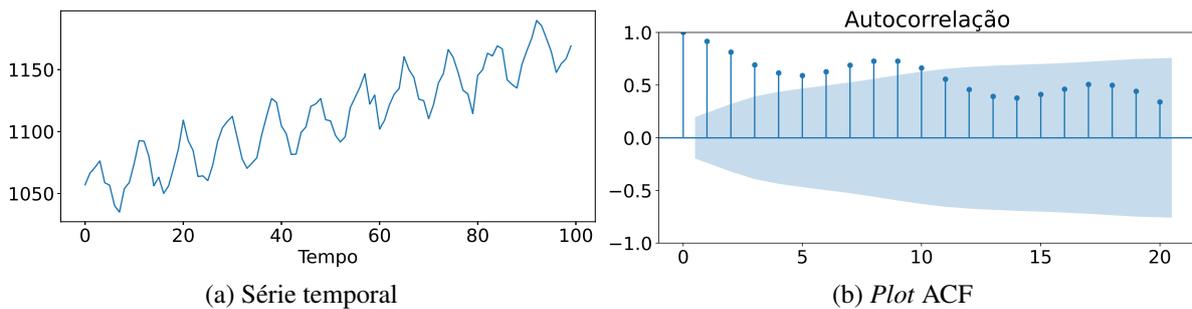


Figura 2.2 Série temporal e respectivo *plot* ACF.

Estacionariedade é um conceito importante para previsão de séries temporais. Uma série é dita estacionária quando suas propriedades estatísticas são constantes ao longo do tempo [21]. Dessa forma, seus valores são gerados a partir de movimentos aleatórios em torno de uma média com variância constante, sem possuir padrões predizíveis a longo prazo. O exemplo mais simples de uma série estacionária é o ruído branco. Algumas abordagens, como modelos auto-regressivos (AR) e de média móvel (MA), se restringem a séries estacionárias. Comportamentos como tendência e sazonalidade são não-estacionários e violam essa restrição.

O processo de diferenciação é útil para transformar uma série não-estacionária em estacionária [22]. Nele, a série original é substituída pela diferença entre cada observação e seu valor anterior, como mostra a Equação 2.3.

$$x'_t = x_t - x_{t-1}. \quad (2.3)$$

Essa transformação pode ser aplicada múltiplas vezes para diminuir ou eliminar tendências e sazonalidades até alcançar uma série estacionária. A Figura 2.3 mostra a série temporal da Figura 2.2 (a) diferenciada uma vez e duas vezes, respectivamente. É possível ver que a série resultante se aproxima de um ruído branco à medida que a ordem aumenta.

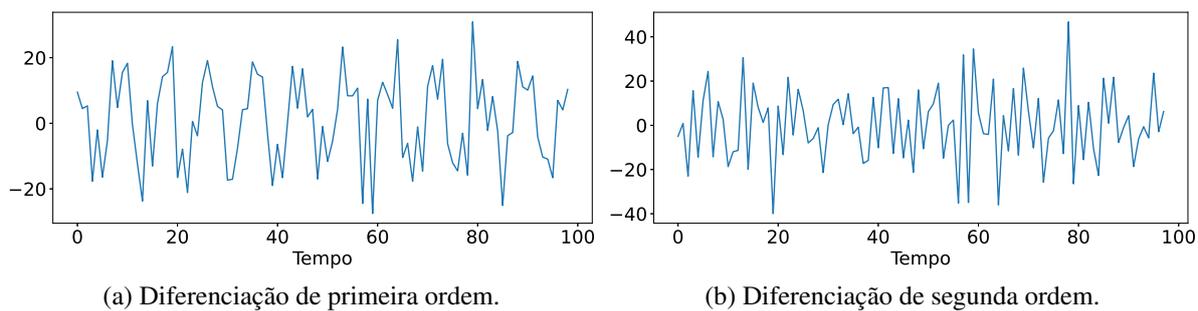


Figura 2.3 Séries temporais resultantes de uma transformação por diferenciação.

Um exemplo de modelo simples utilizado para representação de séries temporais estacionárias é o *random walk* [19], descrito na Equação 2.4. Ele modela a diferença da série como ruído branco, representado por ε_t na equação. É comumente usado como modelo de referência para comparação com outros modelos mais complexos.

$$x_t = x_{t-1} + \varepsilon_t. \quad (2.4)$$

Trabalhos Relacionados

Nesse capítulo são apresentados alguns dos estudos que foram feitos sobre seleção dinâmica de comitês e combinação de previsões. Em [15], é proposta uma abordagem de seleção dinâmica de comitês heterogêneos para prever séries temporais de irradiância solar. Seu comitê é composto por sete diferentes modelos da literatura: ARIMA, SVR, MLP, ELM, *Deep Belief Network* (DBN), *Random Forest* (RF) e *Gradient Boosting* (GB). Aqui a RoC é definida usando uma abordagem de acurácia local, onde os padrões mais próximos de acordo com a distância euclidiana são escolhidos para medir a competência dos modelos no *pool*. [16] enfrenta o problema de adaptação ao desvio de conceito aplicando um mecanismo de detecção de desvio. Essa técnica usa uma medida baseada na correlação de Pearson entre as previsões dos modelos base e a série temporal alvo para remover dinamicamente os modelos de baixo desempenho do *pool*.

Algumas outras abordagens concentram-se em encontrar os melhores pesos de combinação para os modelos disponíveis no *pool*. [17] utiliza uma função não linear baseada no erro gaussiano complementar para pesar e combinar dinamicamente previsões de séries temporais feitas por um comitê heterogêneo. Em [23], um comitê heterogêneo composto por dez modelos diferentes é proposto, onde as configurações de combinação são ajustadas dinamicamente usando o algoritmo de otimização bayesiana (BOA). [24] apresenta um método para combinar previsões por meio de um meta-modelo, que é pré-treinado usando uma coleção de características de séries temporais. Logo, para uma nova série temporal, o método pode extrair suas características e atribuir pesos aos modelos base do conjunto. [25] usa aprendizado por reforço para ajustar os pesos dos modelos base dinamicamente. O problema é modelado como um Processo de Decisão de Markov (MDP), onde cada estado descreve um passo de tempo da série temporal, e cada ação representa uma possível combinação de peso a ser atribuída.

A proposta se destaca dos demais trabalhos ao usar realizar uma otimização baseada no conjunto de dados capaz de decidir os melhores parâmetros de seleção dinâmica, como o número de modelos a serem selecionados a cada ponto de teste, o tamanho da região de competência e a função de combinação a ser utilizada. Além disso, ela também utiliza como critério para formação da RoC as janelas mais recentes ao padrão de teste, abordagem que foi apenas implementada anteriormente utilizando um comitê homogêneo [11], enquanto a proposta utiliza um comitê heterogêneo composto por seis diferentes modelos.

Sistema Proposto

4.1 Visão Geral

O processo geral é ilustrado na Figura 4.1, e pode ser dividido em duas fases: (I) Treinamento e (II) Teste. A fase (I) recebe os conjuntos de dados de treinamento e validação como entrada e gera os modelos treinados e os parâmetros da seleção dinâmica como saída. Os parâmetros de seleção dinâmica consistem no tamanho da Região de Competência (RoC) (k), o número de modelos (n) para gerar a previsão final e o tipo de combinação (média ou mediana) se mais de um modelo de previsão é selecionado.

A primeira etapa da Fase (I) consiste em criar um *pool* de modelos para prever a série temporal. Para aumentar a diversidade, o *pool* é composto por diferentes modelos que são treinados e validados para a escolha de seus melhores hiperparâmetros. Essa parte é essencial para garantir que o *pool* terá a capacidade de prever diferentes padrões.

Na segunda etapa da Fase (I), os modelos treinados do *pool* são utilizados para buscar os melhores parâmetros das etapas dinâmicas de seleção e combinação. Esta etapa é realizada usando o conjunto de validação. Assim, define-se o tamanho da RoC (k), a quantidade de modelos (n) utilizados na etapa de combinação e o tipo de combinação (média e mediana). Após a conclusão da fase de treinamento (Fase (I)), o sistema proposto pode ser empregado para prever padrões de teste não vistos.

Para cada padrão de teste (x_t), a Fase (II) do sistema proposto recebe como entrada o *pool* treinado, e os parâmetros para as etapas de seleção dinâmica e combinação. Um diagrama passo a passo dessa fase é mostrado na Figura 4.2.

A etapa de seleção dinâmica é onde os melhores modelos do *pool* para prever cada padrão de teste são selecionados. Isso envolve criar a RoC com as k janelas passadas ($W_{t-1}, W_{t-2}, \dots, W_{t-k}$) em relação ao padrão de teste atual (x_t) e avaliar o desempenho médio de cada modelo do *pool* nessas janelas. Cada janela de tempo W_t contém l observações ($x_{t-k}, x_{t-(k+1)}, x_{t-(k+2)}, \dots, x_{t-(k+l-1)}$). Então, cada modelo m_j do *pool* gera uma previsão $\hat{x}_{(t-k)+1}$ para cada janela W_t que compõe a região de competência. O desempenho de um modelo m_j do *pool* é medido usando o erro médio entre sua previsão e o valor real para cada W_t que compõe a RoC.

Após a seleção dos n modelos com os menores erros, a previsão final pode ser gerada. Se $n = 1$, a etapa de combinação é desnecessária; caso contrário, as previsões dos $n > 1$ modelos são geradas usando a combinação por média ou mediana.

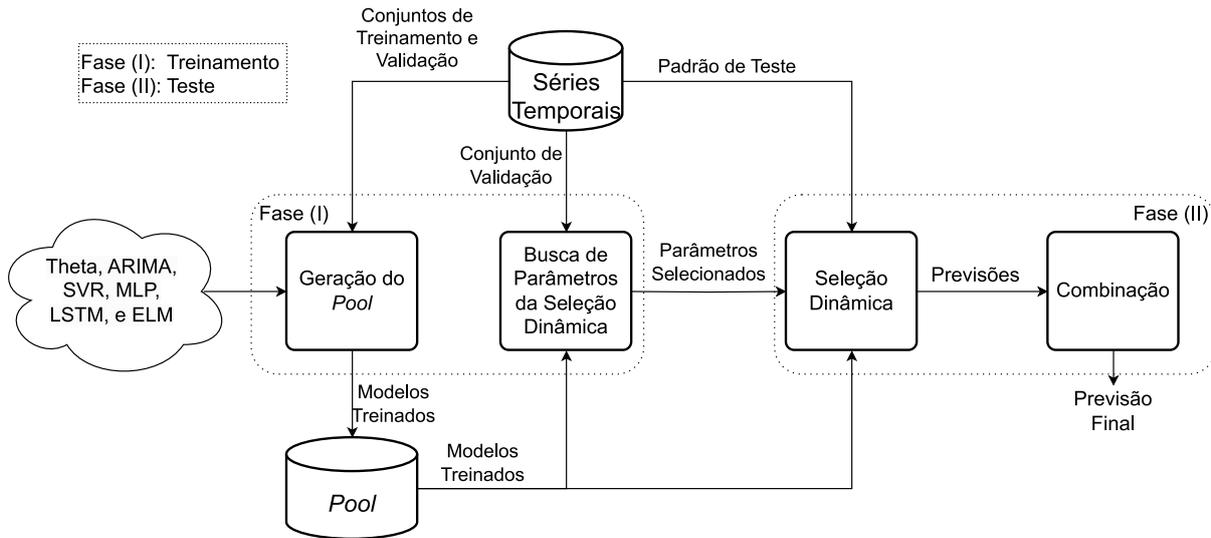


Figura 4.1 Visão geral do sistema proposto.

4.2 Modelos do Comitê

O sistema proposto emprega um comitê heterogêneo composto por seis modelos: ARIMA, Modelo Theta, SVR, MLP, ELM e LSTM. Esses modelos de aprendizado de máquina e estatísticos foram escolhidos porque apresentam características distintas e são amplamente empregados na literatura de previsão de séries temporais [2, 26, 27, 28, 29, 30, 31, 32, 33]. Essa seção descreve o funcionamento dos modelos do comitê.

4.2.1 Autoregressive Integrated Moving Average (ARIMA)

ARIMA é uma generalização do modelo *Autoregressive Moving Average* (ARMA). Ele decompõe uma série temporal numa parte auto-regressiva (AR) e numa parte de média móvel (MA) dos erros de previsão passados, podendo usar também uma parte integrada (I), a qual substitui a série temporal por sua diferenciação [34].

Um modelo auto-regressivo prevê valores futuros de uma série com base na combinação linear de p valores passados da própria variável, chamados de atrasos. Já um modelo de média móvel prevê a variável como uma média ponderada de q erros passados. O processo de diferenciação é aplicado a uma série para torná-la estacionária, dessa forma removendo componentes de tendência e sazonalidade e estabilizando sua média ao longo do tempo. Esse processo pode ser aplicado múltiplas vezes, sendo esse número de vezes a ordem de diferenciação d . As três partes em conjunto compõem o ARIMA(p,d,q), cuja formulação está presente na Equação 4.1

$$\hat{x}_t = \phi_1 \hat{x}_{t-1} + \dots + \phi_p \hat{x}_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad (4.1)$$

em que \hat{x}_t é a série temporal diferenciada de ordem d , ε_t são os erros de previsão, e ϕ e θ são os coeficientes de auto-regressão e de média móvel, respectivamente. O modelo é estimado usando a metodologia de Box–Jenkins [35], a qual se baseia numa estratégia iterativa que con-

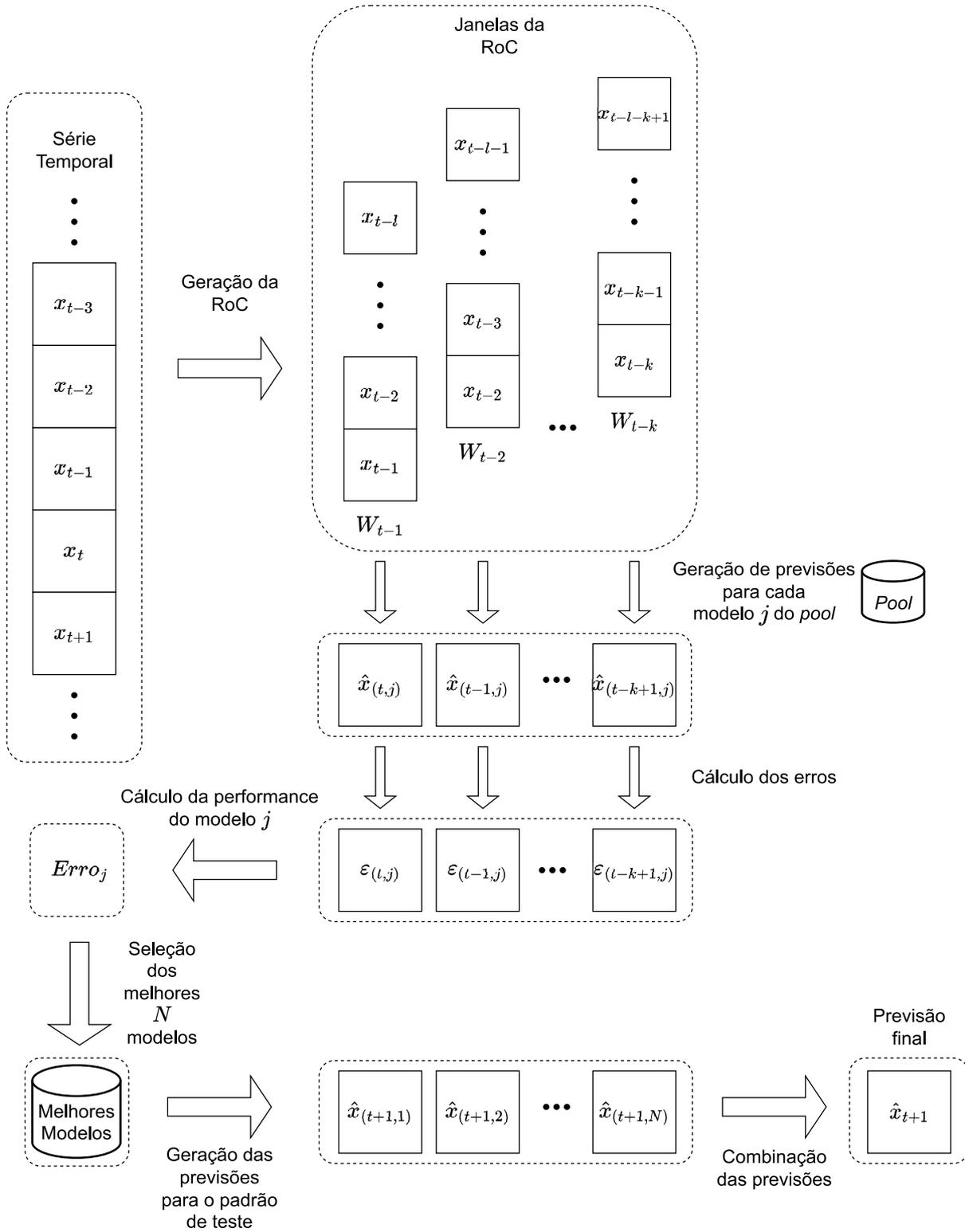


Figura 4.2 Etapas de Seleção Dinâmica e Combinação.

siste em propor um modelo inicial, calcular seus coeficientes e analisar os ruídos para verificar a sua adequação.

4.2.2 Modelo Theta

O modelo Theta [36, 37] se baseia na decomposição de séries temporais a partir de dois componentes, também chamados de linhas theta. A previsão ocorre calculando os valores futuros de cada linha e depois combinando para obter a previsão final. Cada linha foca em comportamentos diferentes da série temporal, e são definidas pelo parâmetro θ , o qual modifica a curvatura da série temporal ao ser aplicado em sua segunda diferença, como mostra a Equação 4.2. Quando θ está entre 0 e 1, flutuações a curto prazo são menores, dando mais ênfase ao comportamento de longo prazo. Um valor de 0 é equivalente a uma regressão linear. Quando θ é maior que 1, as flutuações são intensificadas, dando mais ênfase aos efeitos de curto prazo.

$$x''_{novo} = x'' \cdot \theta. \quad (4.2)$$

Normalmente duas linhas theta são utilizadas. Uma possui $\theta = 0$ para focar no comportamento a longo prazo e é definida a partir de uma regressão linear. A outra possui $\theta = 2$ para focar nos efeitos a curto prazo e é extrapolada a partir de *Simple Exponential Smoothing* (SES). Na presença de sazonalidade, ela é removida por decomposição no início do processo, e adicionada ao final para compor a previsão do modelo.

4.2.3 Support Vector Regression (SVR)

SVR é uma extensão do modelo *Support Vector Machine* (SVM) para regressão [38]. SVM é um classificador que atua criando um hiperplano ou conjunto de hiperplanos capazes de separar amostras de classes diferentes em regiões distintas, maximizando a distância entre o hiperplano e os pontos mais próximos de cada lado. Em regressão linear, SVR tenta encontrar a linha que melhor se ajusta aos pontos. Essa linha é o hiperplano que melhor captura as observações dentro de uma margem, como mostra a Equação 4.3

$$f(x) = w \cdot x + b, \quad (4.3)$$

em que x é o vetor de entrada e w é o vetor de pesos determinado pelo problema de minimização de risco através de uma função de custo. Essa função é caracterizada pela discrepância entre o valor predito e o valor observado. O modelo busca gerar previsões a uma distância máxima de ε do valor esperado. Essa distância funciona como uma margem, dentro da qual os pontos do conjunto de dados de treinamento podem estar sem contabilizar uma penalidade na função de custo. Para casos em que um modelo linear seja insuficiente para se ajustar aos pontos, é possível usar uma função *kernel* que mapeia a entrada em um espaço de características de dimensão maior. Nesse espaço é possível encontrar uma função linear que se ajuste aos pontos e ao mesmo tempo seja não-linear no espaço original.

4.2.4 *Multilayer Perceptron (MLP)*

MLP é uma classe de redes neurais do tipo *feedforward*, ou seja, as conexões entre seus nós não formam ciclos e a informação de entrada flui apenas em uma direção [39]. Esse algoritmo possui uma camada de entrada, uma ou mais camadas escondidas e uma camada de saída. Os nós da primeira camada representam os valores de cada característica de entrada. As seguintes camadas são compostas por neurônios que recebem uma combinação linear dos valores da camada anterior, e em seguida usam uma função de ativação não-linear para gerar suas saídas. Durante o treinamento, o modelo calcula o erro da saída e ajusta os pesos de suas conexões, partindo dos nós da camada de saída até os nós da camada de entrada, seguindo um algoritmo conhecido como *backpropagation* [40]. MLP é capaz de aproximar funções não-lineares e pode ser usado tanto para classificação quanto para regressão. Nesse trabalho foi implementado com uma única camada oculta e utilizou a função logística como função de ativação.

4.2.5 *Extreme Learning Machine (ELM)*

ELM [41] é outra classe de redes neurais do tipo *feedforward* que geralmente possui uma camada escondida e se diferencia do MLP pelo seu processo de treinamento. Enquanto o MLP usa o algoritmo de *backpropagation* para ajustar os pesos por toda a rede, ELM utiliza pesos aleatórios para suas camadas ocultas e ajusta os pesos da camada de saída durante a aprendizagem.

Nesse trabalho foi usada uma rede neural com uma camada oculta. Os pesos das conexões entre a camada de entrada e a camada oculta (W_1) são determinados aleatoriamente. Já os pesos das conexões entre a camada oculta e a camada de saída (W_2) são calculados a partir da pseudo-inversa de Moore-Penrose, a qual procura a solução de melhor ajuste pelo método dos quadrados mínimos e de menor norma. O cálculo é descrito na Equação 4.4

$$W_2 = \sigma(W_1 X)^+ Y, \quad (4.4)$$

em que σ é a função de ativação, X a matriz de variáveis independentes, Y a matriz de variáveis resposta e \cdot^+ a pseudo-inversa.

4.2.6 *Long Short-Term Memory (LSTM)*

LSTM pertence ao tipo de rede neurais recorrentes (RNN). Ao contrário das redes *feedforward*, as conexões entre os nós de uma RNN podem formar ciclos, fazendo com que a saída de um nó funcione como entrada para o mesmo. Isso permite guardar informação contextual por tempo indeterminado, podendo ser útil para realizar previsões de passos de tempo futuros [42].

Um problema das RNN tradicionais é que a informação contextual que elas podem armazenar é limitada. Isso ocorre porque a influência da entrada na camada oculta, e consequentemente na saída da rede neural, pode decair exponencialmente à medida que ela circula as conexões da rede [43]. Esse problema é conhecido na literatura como *vanishing gradient*. Para contorná-lo, as redes neurais LSTM empregam uma arquitetura composta por unidades especiais que se comportam como máquinas de estados. Cada uma delas possui uma célula de memória, uma porta de entrada, uma porta de saída, e uma porta de esquecimento. As portas

de saída e entrada controlam o fluxo de informação pela célula, enquanto a porta de esquecimento determina quais informações dos estados serão descartadas. Dessa forma, é possível guardar informações relevantes de estados passados por um longo prazo para realizar previsões em passos de tempo futuros.

Protocolo Experimental

Um conjunto de sete séries temporais foi usado para avaliar o sistema proposto: *Electricity*, *Goldman Sachs*, *Pollution*, *Star*, *Sunspot*, *Vehicle* e *Wine*. Esses conjuntos de dados podem ser acessados em¹. A Figura 5.1 mostra o *plot* das séries.

Essa base de dados é composta por séries temporais com diferentes características, como tendência, sazonalidade, ciclos, entre outras. É possível ver que as séries *Electricity*, *Pollution* e *Vehicle* possuem uma tendência positiva. Já *Goldman* e *Wine* possuem tendência negativa, enquanto *Sunspot* e *Star* não aparentam possuir tendência. Algumas das séries do experimento possuem sazonalidade, e para estimar suas frequências foi usada a função de autocorrelação (ACF), cujo *plot* é mostrado na Figura 5.2. Nesses casos, a autocorrelação é mais forte nos atrasos que são múltiplos da frequência, permitindo fazer a identificação. A Tabela 5.1 mostra as frequências aproximadas de sazonalidade estimadas para cada série, as quais serviram como informação de entrada para o modelo Theta.

Cada série temporal foi normalizada para o intervalo $[0,1]$ e dividida em três conjuntos seguindo a ordem temporal: treinamento, validação e teste, que contêm 50%, 25% e 25% das observações, respectivamente. O primeiro conjunto foi utilizado para treinar os modelos propostos e da literatura, enquanto o conjunto de validação foi empregado para selecionar os melhores hiperparâmetros dos modelos e os parâmetros de seleção dinâmica. Por fim, o conjunto de teste foi utilizado para avaliar o desempenho do modelo proposto, juntamente com outros métodos da literatura e modelos individuais que compõem o comitê. Todos os modelos foram avaliados no cenário de previsão de um passo à frente.

¹https://github.com/EraylsonGaldino/dataset_time_series

Tabela 5.1 Frequências de sazonalidade estimadas para cada série temporal, medidas em passos de tempo.

| Série Temporal | Frequência |
|--------------------|------------|
| <i>Electricity</i> | 12 |
| <i>Goldman</i> | - |
| <i>Pollution</i> | 3 |
| <i>Star</i> | 28 |
| <i>Sunspot</i> | 11 |
| <i>Vehicle</i> | 12 |
| <i>Wine</i> | 12 |

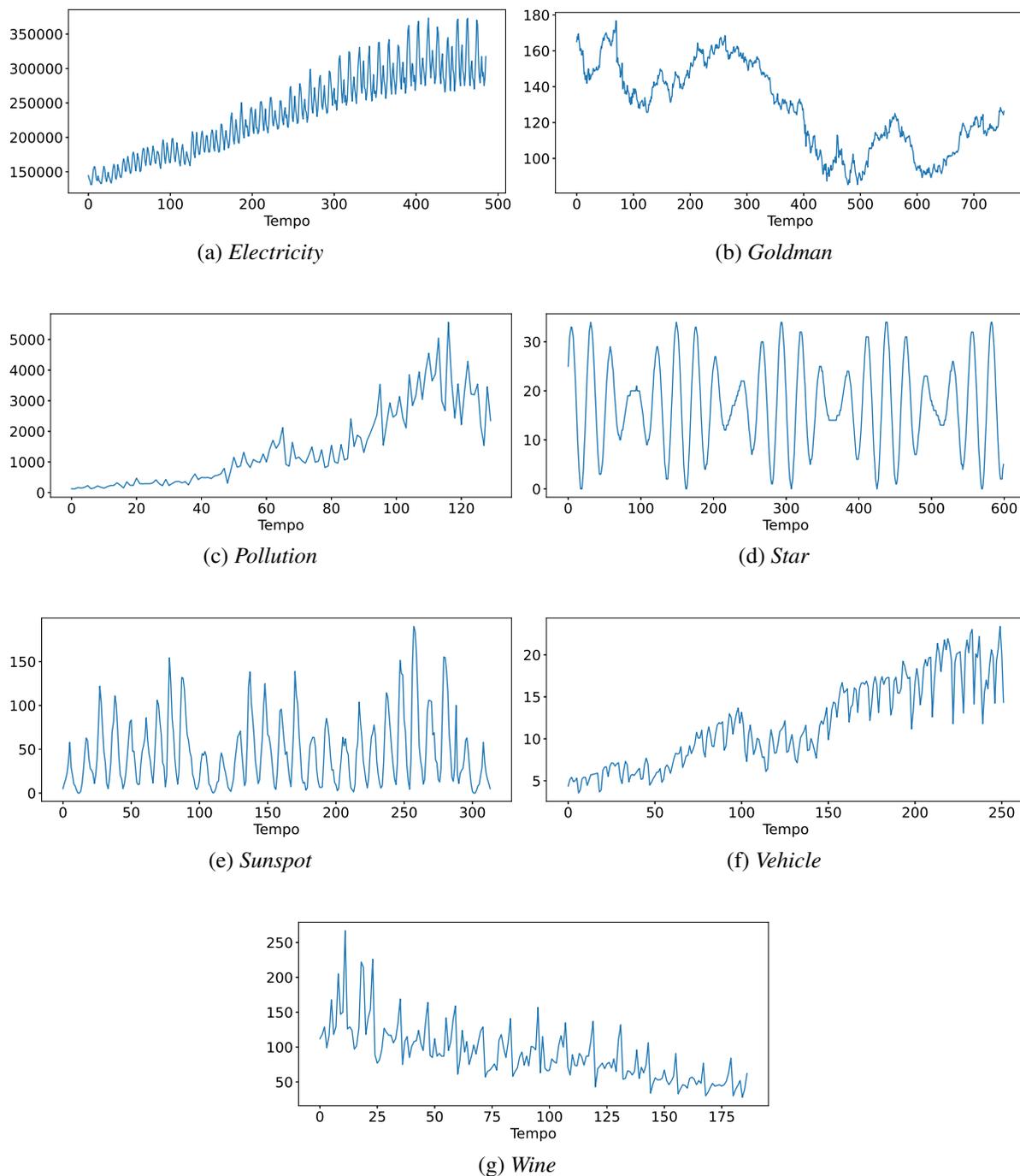


Figura 5.1 Plot das séries temporais utilizadas no experimento.

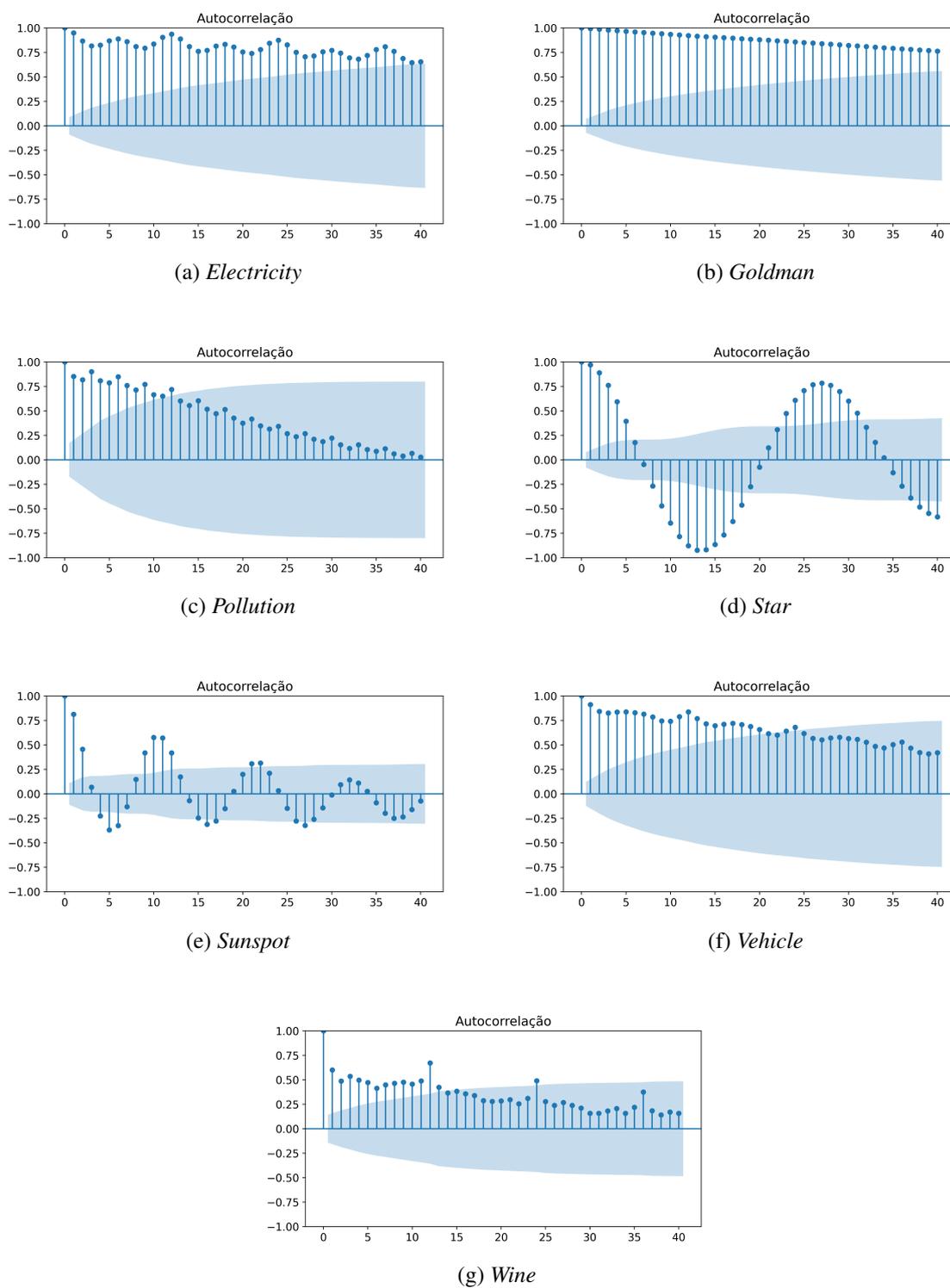


Figura 5.2 Plot ACF das séries temporais do experimento.

Tabela 5.2 Espaço de hiperparâmetros dos modelos que compõem o comitê do sistema proposto.

| Modelo | Parâmetros | Valores |
|--------|--------------------------|-------------------------------------|
| ARIMA | p, d, q | Método de Hyndman [44] |
| Theta | α | <i>Simple Exponential Smoothing</i> |
| | b_0 | <i>Ordinary Least Squares</i> |
| | θ_1 | 0 |
| | θ_2 | 2 |
| MLP | Tamanho da camada oculta | 20, 50, 100 |
| | Tolerância | 1E-3, 1E-4, 1E-5 |
| | Função de ativação | Função Logística |
| ELM | Algoritmo | Pseudo-inversa de Moore-Penrose |
| | Função de Ativação | Tangente Hiperbólica |
| | Tamanho da camada oculta | 20, 50, 100, 200, 500 |
| SVR | Kernel | RBF |
| | C | 1, 10, 100, 1000 |
| | γ | 1, 1E-1, 1E-2, 1E-3 |
| | ϵ | 1E-1, 1E-2, 1E-3, 1E-4, 1E-5 |
| LSTM | Unidades | 15, 25 |
| | Tamanho do lote | 1, 10, 100 |
| | Épocas | 50, 100 |
| | Função de ativação | ReLU, Tangente Hiperbólica |

Para os modelos de aprendizagem de máquina (SVR, MLP, ELM e LSTM), cada série foi reestruturada como janelas deslizantes compostas por 20 observações (atrasos). Desta forma, a previsão é formulada como uma tarefa de aprendizagem supervisionada onde os atrasos pertencentes a uma janela são as características de uma linha de entrada, a qual é usada para prever a próxima observação na série. Para os modelos estatísticos (ARIMA e Theta), os parâmetros de modelo foram atualizados após cada novo padrão de teste avaliado, e em seguida usados para prever o próximo ponto.

As melhores configurações para os modelos de aprendizagem de máquina que compõem o comitê heterogêneo do sistema proposto foram estabelecidas usando uma otimização de hiperparâmetros a partir da técnica *grid search*. Essa seleção foi realizada com base no desempenho do conjunto de validação. A Tabela 5.2 mostra o espaço de busca usado para cada hiperparâmetro. Para o modelo ARIMA, foi empregado o algoritmo *stepwise* descrito por Hyndman e Khandakar (2008) [44]. O modelo Theta ajusta um modelo de *Simple Exponential Smoothing* (SES) para estimar α e uma regressão de *Ordinary Least Squares* (OLS) para estimar b_0 [45].

Os parâmetros de seleção dinâmica do sistema proposto também foram estabelecidos a partir do desempenho do conjunto de validação. Este processo de busca foi realizado para

Tabela 5.3 Parâmetros de Seleção Dinâmica para o sistema proposto definidos baseado na performance no conjunto de validação.

| Série Temporal | k | n | comb |
|--------------------|----|---|---------|
| <i>Electricity</i> | 3 | 5 | Mediana |
| <i>Goldman</i> | 6 | 5 | Mediana |
| <i>Pollution</i> | 12 | 2 | Média |
| <i>Star</i> | 12 | 1 | - |
| <i>Sunspot</i> | 4 | 3 | Mediana |
| <i>Vehicle</i> | 14 | 1 | - |
| <i>Wine</i> | 19 | 1 | - |

determinar o tamanho da RoC (k), o número de modelos selecionados para gerar a previsão final (n) e o operador de combinação ($comb$). Os valores de k usados na busca pertencem ao intervalo $[1, 20]$, enquanto os valores de n pertencem ao intervalo $[1, 6]$. Se $n = 1$, seleciona-se a previsão do melhor modelo de acordo com seu desempenho na região de competência. Caso contrário, quando $n > 1$, um operador é utilizado para combinar as previsões dos melhores n modelos. Este operador é escolhido entre média e mediana. Para cada série temporal, uma configuração ($k, n, comb$) foi escolhida usando a abordagem de *grid search* com base no desempenho no conjunto de validação. A Tabela 5.3 mostra os parâmetros selecionados para o sistema proposto.

O sistema proposto foi comparado com os modelos individuais que compõem o comitê e dois sistemas de seleção dinâmica: DSLA [15] e DSNAW [11]. Estes dois últimos sistemas empregam um comitê homogêneo composto por 100 SVRs. Os SVRs foram treinados usando diferentes amostras criadas por meio da reamostragem (com substituição) das janelas da série temporal original. Os sistemas DSLA e DSNAW também empregaram uma fase de validação para a escolha dos parâmetros k, n e $comb$. Neste caso, n pertence ao intervalo $[1, 20]$.

Após a fase de teste, os resultados foram avaliados usando três métricas diferentes: *Mean Squared Error* (MSE), *Mean Absolute Percentage Error* (MAPE), e *Average Relative Prediction Error Variation* (ARV). MSE mede a diferença média quadrática entre o valor observado e o valor real, como mostra a Equação 5.1. Já a métrica MAPE, descrita na Equação 5.2, mede a acurácia em unidades percentuais, calculada como a razão média entre o erro absoluto de cada previsão e o valor real. Na Equação 5.3 está descrita a métrica ARV, a qual compara o desempenho de um determinado método de previsão com a previsão utilizando a média. Um valor de ARV menor que 1 indica que o modelo é melhor do que usar a média, enquanto um valor de ARV maior que 1 indica o contrário. Um valor igual a 1 indica que ambas as abordagens são equivalentes.

$$MSE = \frac{1}{N} \sum_{t=1}^N (real_t - previsto_t)^2. \quad (5.1)$$

$$MAPE = \frac{100}{N} \sum_{t=1}^N \left| \frac{real_t - previsto_t}{real_t} \right|. \quad (5.2)$$

$$ARV = \frac{\sum_{t=1}^N (real_t - previsto_t)^2}{\sum_{t=1}^N (real - previsto_t)^2}. \quad (5.3)$$

Durante a execução do sistema a métrica MSE foi usada para fins de comparação entre os modelos. Ela serviu para definir as melhores configurações nas escolhas de parâmetros realizadas pela técnica *grid search* durante a fase de treinamento e também para ranquear os modelos de melhor performance durante a seleção dinâmica durante a fase de teste.

A métrica razão, definida na Equação 5.4, foi empregada para comparar o desempenho do sistema proposto com outros modelos da literatura em termos de diferença percentual.

$$razão = \frac{(\varepsilon_a - \varepsilon_b)}{\varepsilon_a} \times 100, \quad (5.4)$$

em que ε_a é o desempenho do modelo da literatura e ε_b é o desempenho do sistema proposto. MSE foi usado como medida de erro.

O Oráculo é um modelo teórico de seleção dinâmica que sempre escolhe o melhor modelo do *pool* para cada padrão de teste. Este modelo é útil para determinar o limite superior de um determinado *pool*. Equação 5.5 quantifica o quão perto uma abordagem de seleção dinâmica está de seu oráculo.

$$razão_o = 100 - \frac{(MAPE_b - MAPE_o)}{MAPE_b} \times 100, \quad (5.5)$$

em que $MAPE_b$ e $MAPE_o$ são os desempenhos de uma dada abordagem de seleção dinâmica, e do respectivo oráculo para o mesmo *pool*. Quanto mais próximo de 100%, melhor é a seleção dinâmica realizada por uma determinada abordagem.

A participação de cada modelo do comitê no resultado final foi medida com o objetivo de investigar a consistência das escolhas da seleção dinâmica do sistema proposto com relação ao desempenho dos modelos. A participação foi medida utilizando a Equação 5.6

$$participação = \frac{|X_s|}{|X_t|} \times 100, \quad (5.6)$$

em que $|X_t|$ representa o número total de padrões do conjunto de teste, e $|X_s|$ representa o número desses padrões nos quais determinado modelo foi selecionado para participar da previsão final. Um valor de 100% indica que o modelo participou de todas as previsões do conjunto de teste.

O projeto foi implementado na linguagem de programação *Python* e os experimentos foram executados num computador com CPU *Ryzen 5 3600* e 16GB de memória RAM.

CAPÍTULO 6

Resultados

A Tabela 6.1 mostra os valores de desempenho em termos de MSE, MAPE e ARV do sistema proposto e de modelos da literatura. O sistema proposto obteve o melhor desempenho na maioria das métricas em dois conjuntos de dados, *Pollution* e Vinho. DSLA obteve o melhor resultado em três séries temporais, *Electricity*, *Star* e *Vehicle*. Os modelos individuais, ARIMA, Theta e MLP, alcançaram o melhor desempenho nas séries *Goldman*, *Wine* e *Sunspot*, respectivamente. É importante destacar que para a série *Wine*, a proposta foi capaz de identificar e selecionar o melhor modelo para a maioria das amostras do conjunto de teste, alcançando a acurácia do melhor modelo individual. Embora o sistema proposto tenha obtido o melhor desempenho em apenas dois conjuntos de dados, ele foi capaz de se adaptar à maioria deles, alcançando o melhor resultado médio. Seguindo esse pensamento, foi realizada uma análise usando a classificação dos modelos em cada conjunto de dados.

A Tabela 6.2 mostra a classificação dos modelos em cada conjunto de dados de acordo com as métricas MSE, MAPE e ARV. Assim, são criados três rankings para cada série. O sistema proposto ficou entre os três melhores modelos em 16 dos 21 casos, Theta alcançou as três melhores classificações em 10 casos e os métodos DSLA e DSNAW ficaram entre os três modelos de melhor acurácia em 9 casos. O sistema proposto alcançou as melhores classificações de média e mediana analisando o desempenho geral. Esse resultado mostra que, embora a proposta não tenha obtido o melhor desempenho na maioria dos casos, ela é capaz de se adaptar a diferentes cenários.

Table 6.3 mostra a métrica *razão* calculada de acordo com a Equação 5.4. A razão corresponde à diferença percentual em relação ao MSE entre o sistema proposto e os modelos individuais e comitês da literatura. O sistema proposto alcançou uma vantagem consistente em todos os conjuntos de dados sobre a maioria dos modelos da literatura. Para as séries *Pollution* e *Wine*, a proposta superou todos os modelos da literatura. Em média, o sistema proposto superou todos os outros modelos da literatura nessa métrica.

A Tabela 6.4 mostra a proporção de observações de teste nas quais a previsão de um modelo do comitê heterogêneo foi usada para compor a previsão do sistema proposto. A soma de cada linha é igual a $n \times 100\%$, já que quando $n > 1$ múltiplas previsões são usadas para compor a previsão final do mesmo padrão de teste. Na série temporal *Electricity*, a maioria dos modelos atingiu uma participação próxima ou igual a 100%. Como o n escolhido foi 5 para esse conjunto de dados, a cada padrão de teste apenas um modelo ficava de fora. É possível ver que para a maioria dos pontos esse modelo foi o ARIMA, o que está consistente com sua performance quando analisado o MSE na Tabela 6.1. Na série temporal *Goldman* há um cenário parecido. Nesse caso, os três modelos com melhor performance, especificamente SVR, Theta e ARIMA, possuem uma participação próxima ou igual a 100%. Os outros três modelos com resultado

Tabela 6.1 Resultados de performance para os métodos avaliados. Valores em negrito representam os melhores resultados para cada par de série temporal e métrica.

| Série Temporal | Métrica | SVR | Theta | ARIMA | MLP | ELM | LSTM | DSLA | DSNAW | Proposto |
|--------------------|---------|---------|----------------|----------------|----------------|---------|---------|----------------|-------------|----------------|
| <i>Electricity</i> | MSE | 1.41E-3 | 2.71E-3 | 1.55E-2 | 1.91E-3 | 2.15E-3 | 6.35E-3 | 1.03E-3 | 1.25E-3 | 1.39E-3 |
| | MAPE | 4.02 | 5.87 | 13.9 | 4.48 | 4.67 | 9.10 | 3.43 | 4.05 | 4.01 |
| | ARV | 1.11E-1 | 1.95E-1 | 7.14E-1 | 1.87E-1 | 2.37E-1 | 2.75E-1 | 7.21E-2 | 1.01E-1 | 1.27E-1 |
| <i>Goldman</i> | MSE | 4.21E-4 | 3.98E-4 | 3.99E-4 | 6.29E-4 | 4.98E-4 | 4.92E-4 | 4.97E-4 | 4.26E-4 | 4.13E-4 |
| | MAPE | 9.39 | 8.89 | 8.86 | 12.2 | 10.1 | 10.3 | 10.4 | 9.32 | 9.23 |
| | ARV | 2.93E-2 | 2.73E-2 | 2.73E-2 | 3.72E-2 | 3.21E-2 | 3.70E-2 | 3.53E-2 | 2.91E-2 | 2.86E-2 |
| <i>Pollution</i> | MSE | 1.28E-1 | 1.70E-2 | 1.77E-2 | 4.52E-2 | 4.26E-1 | 1.85E-1 | 2.40E-2 | 2.48E-2 | 1.68E-2 |
| | MAPE | 57.1 | 18.6 | 21.4 | 34.7 | 108 | 68.0 | 22.9 | 24.5 | 21.5 |
| | ARV | 1.04 | 1.62 | 1.19 | 8.83E-1 | 9.55E-1 | 1.11 | 9.05E-1 | 8.74E-1 | 8.46E-1 |
| <i>Star</i> | MSE | 4.16E-5 | 4.72E-3 | 1.87E-4 | 1.17E-4 | 5.58E-5 | 1.69E-4 | 2.95E-5 | 5.30E-5 | 3.34E-5 |
| | MAPE | 1.73 | 20.7 | 4.14 | 3.12 | 3.62 | 3.95 | 1.52 | 2.05 | 1.81 |
| | ARV | 5.92E-4 | 7.24E-2 | 2.64E-3 | 1.66E-3 | 7.88E-4 | 2.41E-3 | 4.18E-4 | 7.50E-4 | 4.74E-4 |
| <i>Sunspot</i> | MSE | 2.14E-2 | 2.14E-2 | 1.59E-2 | 1.37E-2 | 2.61E-2 | 1.78E-2 | 1.77E-2 | 1.43E-2 | 1.50E-2 |
| | MAPE | 58.1 | 82.7 | 61.6 | 52.8 | 117 | 68.3 | 53.6 | 52.4 | 54.3 |
| | ARV | 5.81E-1 | 3.11E-1 | 2.87E-1 | 2.51E-1 | 3.66E-1 | 3.09E-1 | 4.29E-1 | 2.90E-1 | 2.82E-1 |
| <i>Vehicle</i> | MSE | 1.46E-2 | 1.33E-2 | 2.10E-2 | 1.62E-2 | 4.39E-2 | 3.09E-2 | 1.51E-2 | 1.43E-2 | 1.49E-2 |
| | MAPE | 13.5 | 13.8 | 17.0 | 14.6 | 23.3 | 18.2 | 13.5 | 14.2 | 14.7 |
| | ARV | 8.01E-1 | 1.01 | 1.79 | 9.26E-1 | 1.00 | 8.41E-1 | 6.50E-1 | 8.35E-1 | 1.01 |
| <i>Wine</i> | MSE | 5.28E-3 | 1.04E-3 | 5.43E-3 | 3.18E-3 | 2.09E-3 | 5.37E-3 | 6.98E-3 | 1.71E-3 | 1.04E-3 |
| | MAPE | 124 | 47.3 | 123 | 102 | 58.9 | 132 | 92.5 | 62.3 | 47.3 |
| | ARV | 7.51E-1 | 3.77E-1 | 1.58 | 7.87E-1 | 4.91E-1 | 7.66E-1 | 6.08E-1 | 4.69E-1 | 3.77E-1 |

Tabela 6.2 Ranking das abordagens avaliadas baseado nas métricas de erro.

| Série Temporal | Métrica | SVR | Theta | ARIMA | MLP | ELM | LSTM | DSLA | DSNAW | Proposto |
|--------------------|---------|------|----------|----------|----------|------|------|----------|----------|-------------|
| <i>Electricity</i> | MSE | 4 | 7 | 9 | 5 | 6 | 8 | 1 | 2 | 3 |
| | MAPE | 3 | 7 | 9 | 5 | 6 | 8 | 1 | 4 | 2 |
| | ARV | 3 | 6 | 9 | 5 | 7 | 8 | 1 | 2 | 4 |
| <i>Goldman</i> | MSE | 4 | 1 | 2 | 9 | 8 | 6 | 7 | 5 | 3 |
| | MAPE | 5 | 2 | 1 | 9 | 6 | 7 | 8 | 4 | 3 |
| | ARV | 5 | 2 | 1 | 9 | 6 | 8 | 7 | 4 | 3 |
| <i>Pollution</i> | MSE | 7 | 2 | 3 | 6 | 9 | 8 | 4 | 5 | 1 |
| | MAPE | 7 | 1 | 2 | 6 | 9 | 8 | 4 | 5 | 3 |
| | ARV | 6 | 9 | 8 | 3 | 5 | 7 | 4 | 2 | 1 |
| <i>Star</i> | MSE | 3 | 9 | 8 | 6 | 5 | 7 | 1 | 4 | 2 |
| | MAPE | 2 | 9 | 8 | 5 | 6 | 7 | 1 | 4 | 3 |
| | ARV | 3 | 9 | 8 | 6 | 5 | 7 | 1 | 4 | 2 |
| <i>Sunspot</i> | MSE | 7 | 8 | 4 | 1 | 9 | 6 | 5 | 2 | 3 |
| | MAPE | 5 | 8 | 6 | 2 | 9 | 7 | 3 | 1 | 4 |
| | ARV | 9 | 6 | 3 | 1 | 7 | 5 | 8 | 4 | 2 |
| <i>Vehicle</i> | MSE | 3 | 1 | 7 | 6 | 9 | 8 | 5 | 2 | 4 |
| | MAPE | 2 | 3 | 7 | 5 | 9 | 8 | 1 | 4 | 6 |
| | ARV | 2 | 7 | 9 | 5 | 6 | 4 | 1 | 3 | 8 |
| <i>Wine</i> | MSE | 6 | 1 | 8 | 5 | 4 | 7 | 9 | 3 | 1 |
| | MAPE | 8 | 1 | 7 | 6 | 3 | 9 | 5 | 4 | 1 |
| | ARV | 6 | 1 | 9 | 8 | 4 | 7 | 5 | 3 | 1 |
| Média | | 4.76 | 4.76 | 6.10 | 5.38 | 6.57 | 7.14 | 3.90 | 3.38 | 2.86 |
| Mediana | | 5 | 6 | 7 | 5 | 6 | 7 | 4 | 4 | 3 |

Tabela 6.3 Diferença percentual (% MSE) de acordo com a Equação 5.4.

| Série Temporal | SVR | Theta | ARIMA | MLP | ELM | LSTM | DSLA | DSNAW |
|--------------------|-------|--------|-------|-------|-------|-------|--------|--------|
| <i>Electricity</i> | 1.88 | 48.78 | 91.06 | 27.35 | 35.61 | 78.15 | -35.07 | -10.97 |
| <i>Goldman</i> | 1.82 | -3.94 | -3.62 | 34.30 | 17.01 | 16.06 | 16.87 | 3.03 |
| <i>Pollution</i> | 86.82 | 1.03 | 5.26 | 62.82 | 96.05 | 90.92 | 29.94 | 32.09 |
| <i>Star</i> | 19.78 | 99.29 | 82.11 | 71.45 | 40.14 | 80.24 | -13.33 | 36.98 |
| <i>Sunspot</i> | 29.73 | 29.79 | 5.83 | -9.74 | 42.39 | 15.84 | 15.03 | -5.37 |
| <i>Vehicle</i> | -2.49 | -12.31 | 28.75 | 7.53 | 65.92 | 51.55 | 1.23 | -4.71 |
| <i>Wine</i> | 80.29 | 0.00 | 80.84 | 67.35 | 50.35 | 80.65 | 85.10 | 39.11 |
| Média | 31.12 | 23.23 | 41.46 | 37.30 | 49.64 | 59.06 | 14.25 | 12.88 |

inferior na métrica MSE desse conjunto de dados obtiveram porcentagens menores. Na série temporal *Pollution* apenas dois modelos eram escolhidos a cada padrão de teste. ARIMA, MLP e Theta alcançaram as maiores porcentagens, uma vez que conseguiram uma vantagem de uma ordem de grandeza na métrica MSE em relação aos outros modelos. O modelo com pior MSE nesse conjunto, especificamente ELM, não participou da previsão de nenhum padrão de teste. Na série temporal *Star* apenas um modelo era escolhido para compor cada previsão final. Os dois modelos que se destacaram em performance, SVR e ELM, foram os únicos a participarem das previsões nesse conjunto de dados. No cenário da série *Sunspot*, mesmo com apenas metade dos modelos sendo usados para compor cada previsão, todos eles conseguiram participação. Ainda assim, MLP, ARIMA e LSTM se destacaram na métrica MSE e alcançaram porcentagens maiores. Na série temporal *Vehicle* há um cenário parecido com a série *Star*. Para esse conjunto $n = 1$ e apenas os dois modelos com melhores resultados, Theta e SVR, revezaram as previsões dos padrões de teste. Por último, a série *Wine* apresentou um resultado curioso, onde novamente o n escolhido foi 1, porém dessa vez apenas um modelo, Theta, participou de todas as previsões. Analisando seu resultado de MSE, o modelo se distanciou dos demais e dessa forma conseguiu compor sozinho todas as previsões para essa série temporal. No geral, Theta obteve a maior média de participação, seguido por SVR e MLP. Ao comparar com a Tabela 6.3, é possível ver que a participação média de cada modelo do comitê segue a mesma ordem que o resultado médio da métrica MSE. Uma vez que essa métrica foi utilizada para ranquear os modelos durante a seleção dinâmica, a relação de concordância entre esses resultados é uma evidência de que as escolhas realizadas pelo sistema foram consistentes com o desempenho dos modelos.

Na Tabela 6.5 é possível ver os resultados de performance para o oráculo do comitê heterogêneo, utilizado no sistema proposto, e o oráculo do comitê homogêneo, utilizado nas abordagens DSNAW e DSLA. Em todas as combinações de conjunto de dados e métricas o comitê homogêneo obteve vantagem, a qual chegou a ser de duas ordens de grandeza nas métricas MSE, MAPE e ARV para a série *Pollution*, e na métrica ARV para as séries *Star* e *Wine*. Quando comparada a performance do comitê homogêneo aos resultados do SVR, modelo base que compõe o comitê, sua vantagem chega a ser de quatro ordens de grandeza na métrica MSE para o conjunto *Pollution*. Mesmo dispondo de um comitê mais poderoso, as abordagens DS-

Tabela 6.4 Porcentagem dos padrões de teste onde um modelo do comitê foi escolhido para participar da previsão final (%), de acordo com a Equação 5.6.

| Séries Temporais | SVR | Theta | ARIMA | MLP | ELM | LSTM |
|--------------------|-------|-------|-------|-------|-------|-------|
| <i>Electricity</i> | 100 | 98.36 | 26.23 | 100 | 97.54 | 77.87 |
| <i>Goldman</i> | 95.24 | 99.47 | 100 | 52.38 | 75.66 | 77.25 |
| <i>Pollution</i> | 12.12 | 42.42 | 75.76 | 60.61 | 0 | 9.09 |
| <i>Star</i> | 60 | 0 | 0 | 0 | 40 | 0 |
| <i>Sunspot</i> | 48.10 | 22.78 | 63.29 | 75.95 | 26.58 | 63.29 |
| <i>Vehicle</i> | 31.75 | 68.25 | 0 | 0 | 0 | 0 |
| <i>Wine</i> | 0 | 100 | 0 | 0 | 0 | 0 |
| Média | 49.60 | 61.61 | 37.90 | 41.28 | 34.25 | 32.50 |

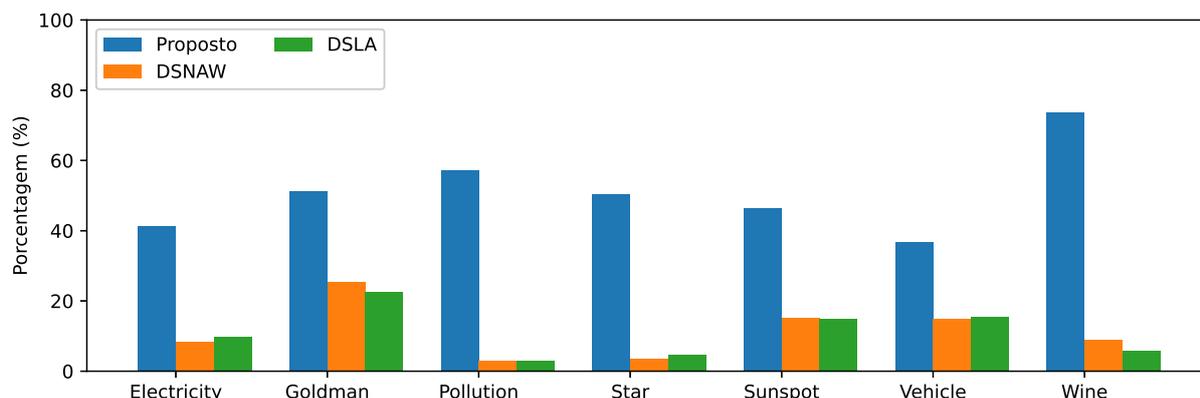


Figura 6.1 Razão de performance entre as abordagens de seleção dinâmica e seus oráculos em termos da Equação 5.5.

NAW e DSLA não conseguiram um resultado geral superior em relação ao sistema proposto. Para analisar esse resultado, a Figura 6.1 mostra a métrica $razão_o$ descrita na Equação 5.5, comparando a seleção dinâmica realizada pelo sistema proposto, DSLA e DSNAW. Quanto mais próximo $razão_o$ estiver de 100% (desempenho do oráculo), melhor será a seleção dinâmica realizada pelo modelo. Pode-se observar que o sistema proposto se aproxima mais de seu oráculo do que o DSLA e o DSNAW. Uma explicação razoável é que escolher o melhor modelo no comitê homogêneo é muito mais desafiador, pois possui 100 preditores. Ao mesmo tempo, a proposta emprega um comitê heterogêneo com apenas seis modelos. Mesmo com um tamanho de *pool* pequeno, a abordagem proposta teve um desempenho geral melhor do que o DSLA e o DSNAW, o que mostra que o emprego de um comitê heterogêneo pode oferecer melhor acurácia com mais eficiência.

A Figura 6.2 mostra a soma dos tempos de execução para todas as séries temporais nas fases de treinamento e teste de cada abordagem. O tempo de treinamento das abordagens de seleção

Tabela 6.5 Resultados de performance para os oráculos criados a partir de cada comitê.

| Série Temporal | Métrica | Oráculo Heterogêneo | Oráculo Homogêneo |
|--------------------|---------|---------------------|-------------------|
| <i>Electricity</i> | MSE | 2.94E-4 | 4.09E-5 |
| | MAPE | 1.66 | 3.40E-1 |
| | ARV | 2.14E-2 | 2.92E-3 |
| <i>Goldman</i> | MSE | 2.15E-4 | 1.10E-4 |
| | MAPE | 4.72 | 2.36 |
| | ARV | 1.45E-2 | 7.69E-3 |
| <i>Pollution</i> | MSE | 6.46E-3 | 2.66E-5 |
| | MAPE | 12.3 | 7.02E-1 |
| | ARV | 3.15E-1 | 1.03E-3 |
| <i>Star</i> | MSE | 9.57E-6 | 4.09E-7 |
| | MAPE | 9.11E-1 | 7.19E-2 |
| | ARV | 1.36E-4 | 5.78E-6 |
| <i>Sunspot</i> | MSE | 6.23E-3 | 3.16E-3 |
| | MAPE | 25.2 | 7.94 |
| | ARV | 1.06E-1 | 5.53E-2 |
| <i>Vehicle</i> | MSE | 3.69E-3 | 1.34E-3 |
| | MAPE | 5.40 | 2.10 |
| | ARV | 1.94E-1 | 6.43E-2 |
| <i>Wine</i> | MSE | 4.70E-4 | 3.00E-5 |
| | MAPE | 34.8 | 5.50 |
| | ARV | 1.26E-1 | 6.97E-3 |

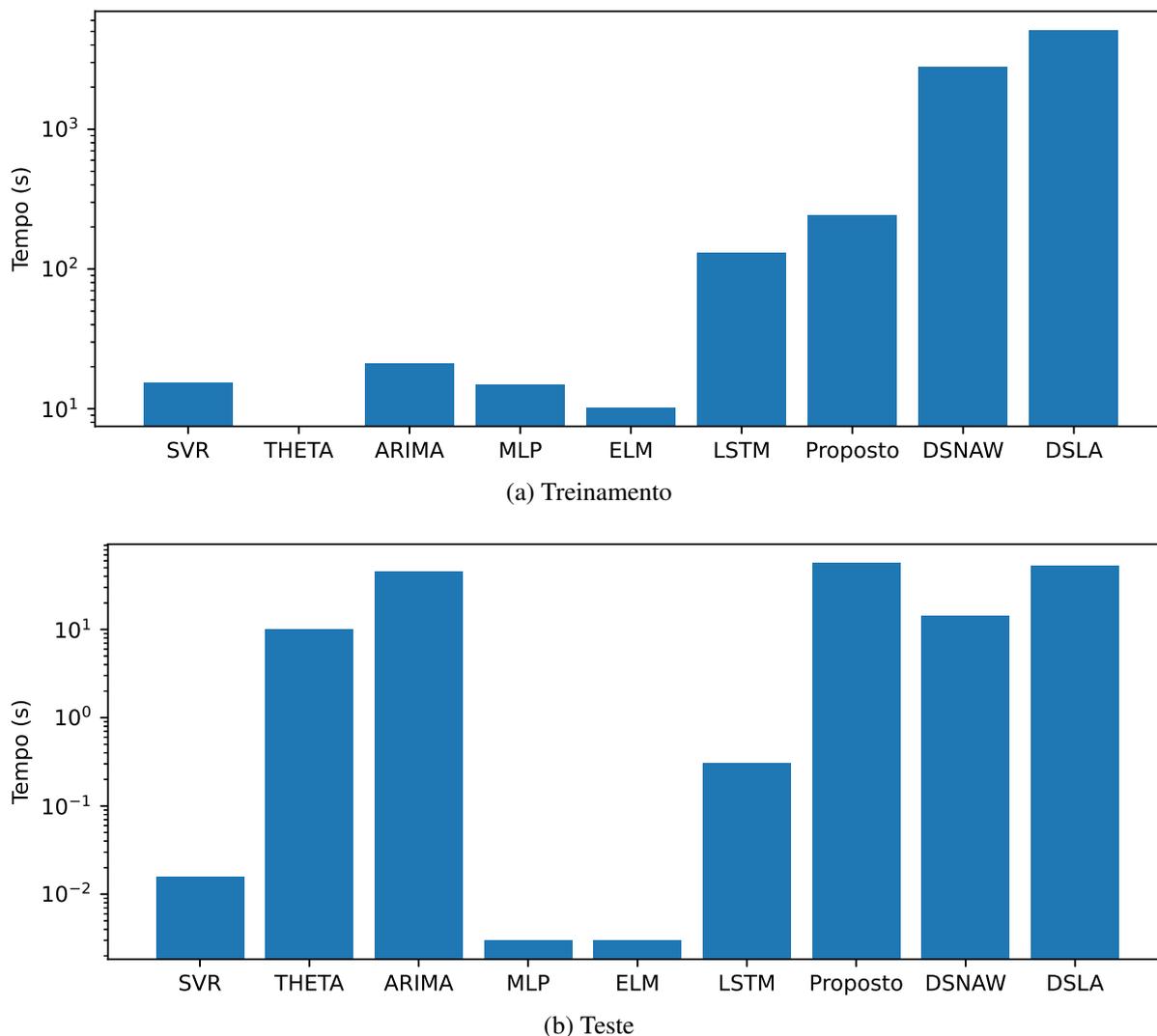


Figura 6.2 Tempos de execução em escala logarítmica.

dinâmica compreende o treinamento dos modelos do comitê e a busca pelos parâmetros de seleção dinâmica. O tempo de teste de uma abordagem de seleção dinâmica consiste no gasto formando a RoC e gerando previsões para cada modelo individual. A proposta funcionou muito mais rápida na fase de treinamento do que as outras abordagens de seleção dinâmica baseadas em comitês homogêneos maiores. Este último não apenas gastou mais tempo treinando cada modelo, mas também foi muito mais lento para encontrar os melhores parâmetros de seleção dinâmica devido ao grande número de preditores a serem avaliados. Na fase de teste, a proposta teve uma execução mais lenta devido aos modelos Theta e ARIMA, pois são ajustados a cada entrada de teste, impactando no tempo final dessa fase.

As Figuras 6.3 a 6.9 mostram, para cada série temporal, o *plot* das previsões do sistema proposto e da melhor solução entre as demais abordagens de modelo individual e de comitês de acordo com a métrica MAPE. Na série *Electricity*, ambas as abordagens DSLA e sistema

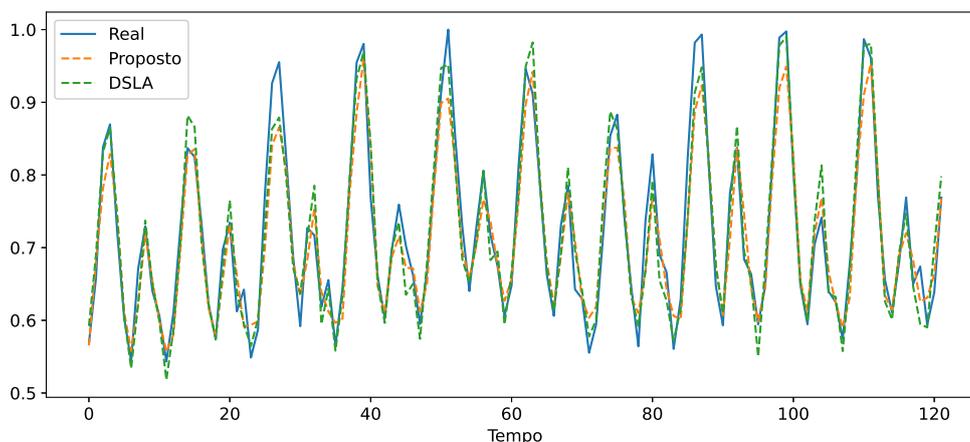


Figura 6.3 Plot de previsões para a série temporal *Electricity*.

proposto conseguiram uma boa aproximação da curva real, com MAPEs de apenas 3.43% e 4.01%, respectivamente. Na série *Goldman*, as soluções da proposta e do modelo ARIMA foram semelhantes, tendo a mesma forma da curva de valores reais, porém defasadas. De fato, é possível ver pela Tabela 6.1 que todas as abordagens conseguiram resultados de erro próximos. Os MAPEs foram de 8.86% para ARIMA e 9.23% para o sistema proposto. Esse conjunto de dados é de uma série financeira, as quais são usualmente representadas por modelos de *random walk* [46] e bastante difíceis de prever devido a movimentos aleatórios. Na série *Pollution*, o modelo Theta apresentou variações menores e conseguiu um MAPE melhor de 18.6%, comparado aos 21.5% do sistema proposto. No cenário da série *Star*, as soluções do DSLA e da proposta foram semelhantes e chegaram muito próximo da curva real, com MAPEs de apenas 1.52% e 1.81%, respectivamente. Na série *Sunspot*, as soluções do método DSNAW e da proposta foram próximos, mas se distanciaram em alguns momentos da curva real, resultando em MAPEs de 52.4% e 54.3%, respectivamente. Na série *Vehicle*, embora os métodos DSLA e proposto tenham apresentado curvas bem diferentes, atingiram MAPEs próximos de 13.5% e 14.7%, respectivamente. A série *Wine* foi o conjunto de dados que obteve os piores resultados na métrica MAPE. O modelo Theta participou sozinho de todas as previsões dos pontos de teste no sistema proposto, resultando em curvas iguais e num MAPE de 47.3%.

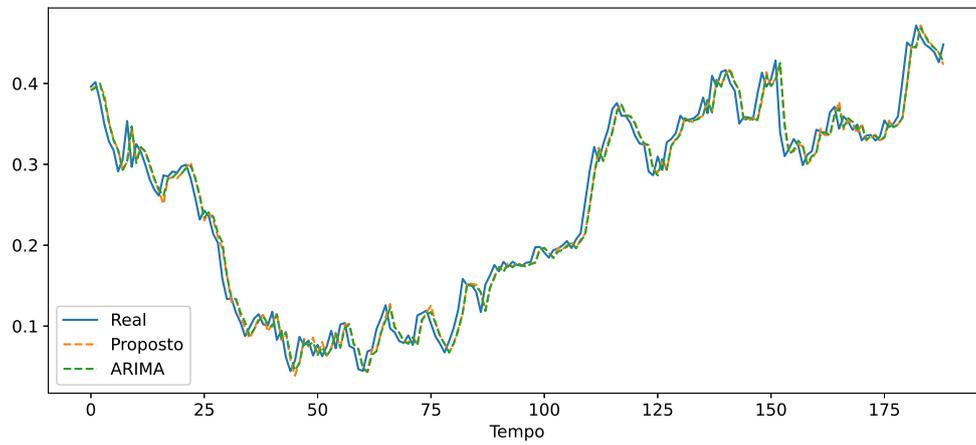


Figura 6.4 Plot de previsões para a série temporal *Goldman*.

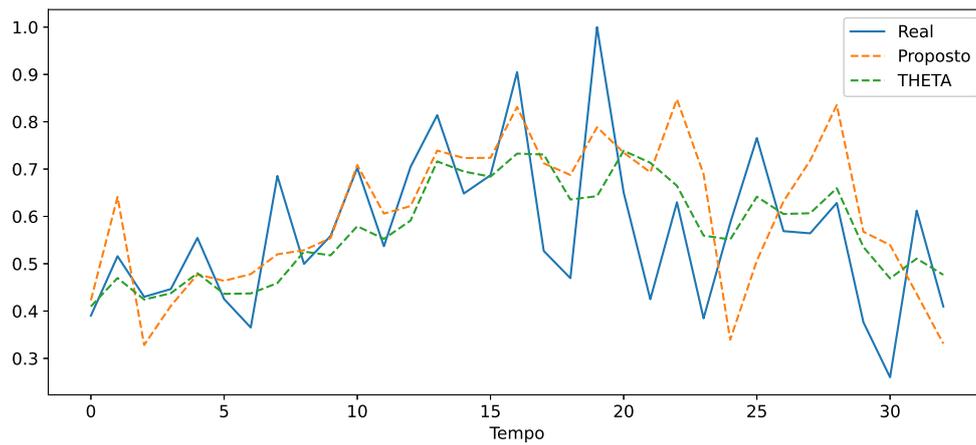


Figura 6.5 Plot de previsões para a série temporal *Pollution*.

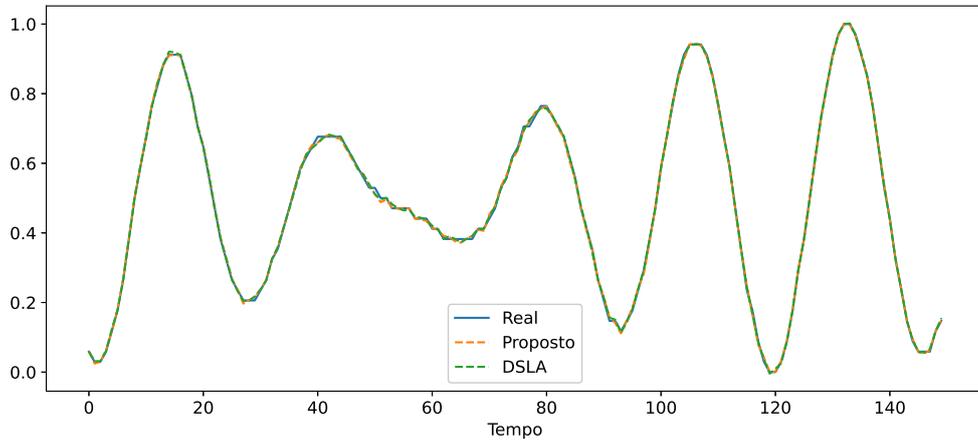


Figura 6.6 *Plot de previsões para a série temporal Star.*

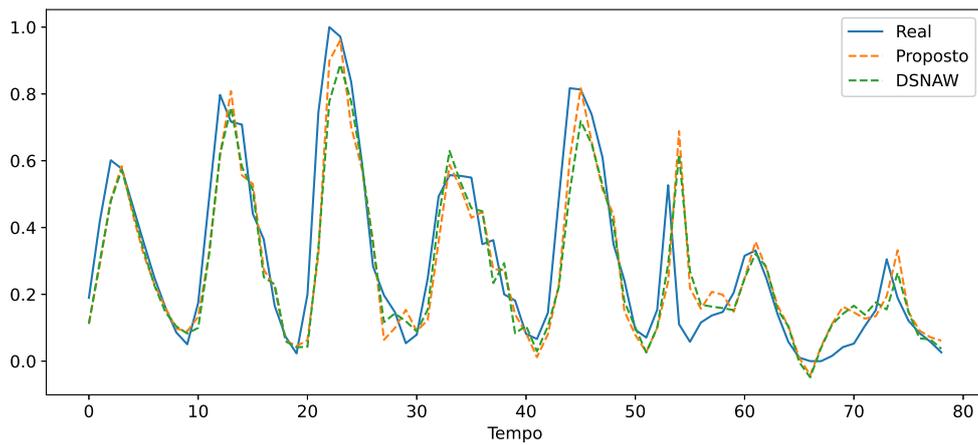


Figura 6.7 *Plot de previsões para a série temporal Sunspot.*

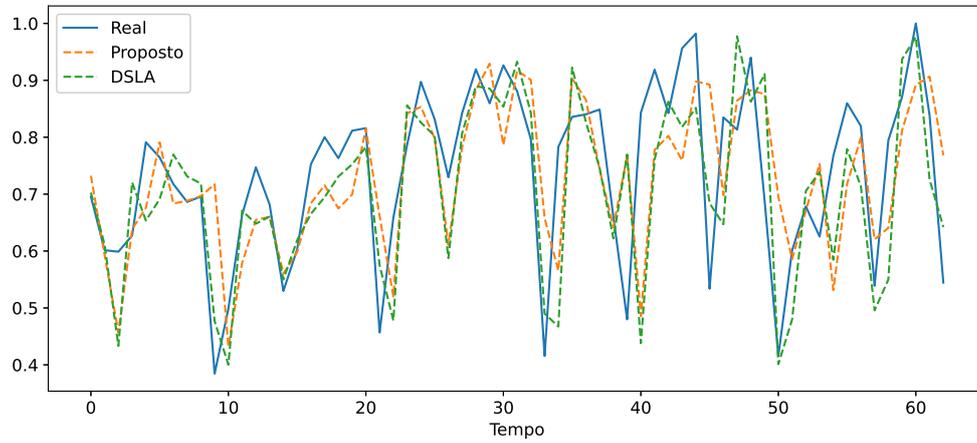


Figura 6.8 *Plot de previsões para a série temporal Vehicle.*

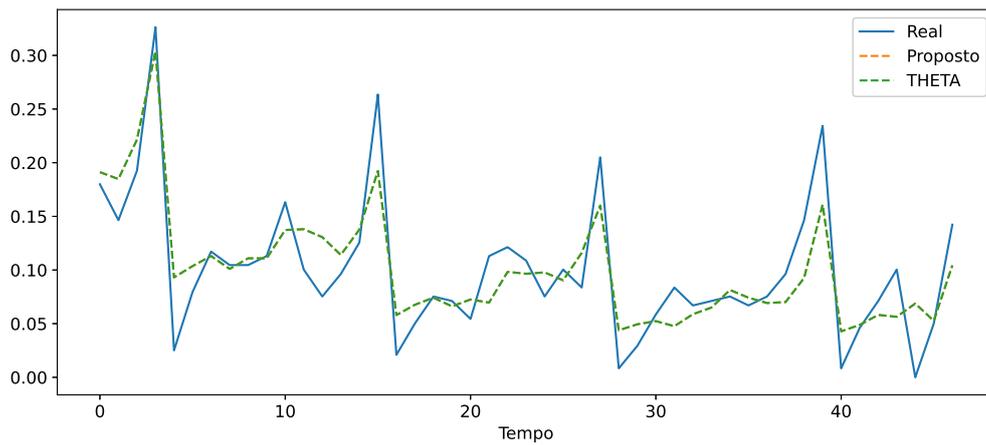


Figura 6.9 *Plot de previsões para a série temporal Wine.*

Conclusão

Este trabalho propôs um sistema inteligente de seleção dinâmica que busca o tamanho da RoC mais adequado, o conjunto de modelos mais competente e a melhor função de combinação para maximizar sua acurácia. A RoC compreende as janelas de tempo mais próximas de um padrão de teste e é usada para acessar a competência de um *pool* heterogêneo de modelos. O comitê é composto por seis modelos diferentes: ARIMA, Theta, SVR, MLP, ELM e LSTM.

A avaliação experimental foi realizada com sete séries temporais bem conhecidas usando três métricas (MSE, MAPE e ARV). O sistema proposto alcançou os melhores resultados gerais em comparação com as abordagens de modelos individuais e seleção dinâmica, classificando-se entre os três primeiros em 16 das 21 combinações considerando as métricas de desempenho e séries temporais. Este resultado mostra que a proposta pode fornecer estabilidade e robustez usando um comitê mais diversificado. O sistema proposto também foi mais eficiente em termos de tempo durante a fase de treinamento do que as outras abordagens de seleção dinâmica.

Com relação a trabalhos futuros, o uso de combinações treináveis é uma alternativa promissora para aumentar a acurácia da proposta [12]. A busca pelos melhores parâmetros dinamicamente é outra diretriz relevante, pois o número de modelos com bom desempenho, o melhor operador de combinação e o tamanho da RoC mais adequado podem mudar ao longo do tempo. Além disso, diferentes combinações de modelos podem levar a uma melhor capacidade de generalização e robustez.

Referências Bibliográficas

- [1] Omer Berat Sezer, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing*, 90:106181, 2020.
- [2] Shruti Kaushik, Abhinav Choudhury, Pankaj Kumar Sheron, Nataraj Dasgupta, Sayee Natarajan, Larry A. Pickett, and Varun Dutt. AI in Healthcare: Time-Series Forecasting Using Statistical, Neural, and Ensemble Architectures. *Frontiers in Big Data*, 3, 2020.
- [3] Zahra Karevan and Johan A.K. Suykens. Transductive LSTM for time-series prediction: An application to weather forecasting. *Neural Networks*, 125:1–9, 2020.
- [4] Ioannis Pierros and Ioannis Vlahavas. Architecture-Agnostic Time-Step Boosting: A Case Study in Short-Term Load Forecasting. In *Artificial Neural Networks and Machine Learning – ICANN 2022*, pages 556–568, Cham, 2022. Springer Nature Switzerland.
- [5] Elorrieta, Felipe, Eyheramendy, Susana, and Palma, Wilfredo. Discrete-time autoregressive model for unequally spaced time-series observations. *A&A*, 627:A120, 2019.
- [6] Zahra Hajirahimi and Mehdi Khashei. Hybrid structures in time series modeling and forecasting: A review. *Engineering Applications of Artificial Intelligence*, 86:83–106, 2019.
- [7] Ahmed Tealab. Time series forecasting using artificial neural networks methodologies: A systematic review. *Future Computing and Informatics Journal*, 3(2):334–340, 2018.
- [8] Bryan Lim and Stefan Zohren. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194):20200209, 2021.
- [9] Changqing Cheng, Akkarapol Sa-Ngasoongsong, Omer Beyca, Trung Le, Hui Yang, Zhenyu Kong, and Satish Bukkapatnam. Time Series Forecasting for Nonlinear and Nonstationary Processes: A Review and Comparative Study. *IIE Transactions*, 01 2015.
- [10] Spyros Makridakis, Evangelos Spiliotis, and Vassilis Assimakopoulos. The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36, 07 2019.

- [11] Eraylson G. Silva, Paulo S. G. De Mattos Neto, and George D. C. Cavalcanti. A Dynamic Predictor Selection Method Based on Recent Temporal Windows for Time Series Forecasting. *IEEE Access*, 9:108466–108479, 2021.
- [12] Paulo SG De Mattos Neto, Paulo Renato A Firmino, Hugo Siqueira, Yara De Souza Tadano, Thiago Antonini Alves, Joao Fausto L De Oliveira, Manoel Henrique Da Nóbrega Marinho, and Francisco Madeiro. Neural-based ensembles for particulate matter forecasting. *IEEE Access*, 9:14470–14490, 2021.
- [13] Min Qi and Guoqiang Peter Zhang. An investigation of model selection criteria for neural network time series forecasting. *European Journal of Operational Research*, 132(3):666–680, 2001.
- [14] Luís Torgo and Mariana Oliveira. Ensembles for Time Series Forecasting. 01 2014.
- [15] Domingos Santos, Paulo De Mattos Neto, Joao Oliveira, Hugo Siqueira, Tathiana Mikamura Barchi, Aranildo Lima, Francisco Madeiro, Douglas Dantas, Attilio Converti, Alex Pereira, José Bione Filho, and Manoel Marinho. Solar irradiance forecasting using dynamic ensemble selection. *Applied Sciences*, 12, 03 2022.
- [16] Amal Saadallah, Florian Priebe, and Katharina Morik. A Drift-Based Dynamic Ensemble Members Selection Using Clustering for Time Series Forecasting. In *Machine Learning and Knowledge Discovery in Databases*, pages 678–694, Cham, 2020. Springer International Publishing.
- [17] Vitor Cerqueira, Luis Torgo, Mariana Oliveira, and Bernhard Pfahringer. Dynamic and Heterogeneous Ensembles for Time Series Forecasting. In *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 242–251, 2017.
- [18] Jiawei Han, Micheline Kamber, and Jian Pei. Data mining concepts and techniques, third edition, 2012.
- [19] Hyndman, R.J., Athanasopoulos, G. Forecasting: principles and practice, 2nd edition, 2018. [Online; accessed 14-April-2023].
- [20] Eleazar C Nwogu, Iheanyi S Iwueze, Kelechukwu CN Dozie, and Hope I Mbachu. Choice between mixed and multiplicative models in time series decomposition. *Int J Stat Appl Math*, 9(5):153–159, 2019.
- [21] Guy P Nason. Stationary and non-stationary time series. *Statistics in volcanology*, 60, 2006.
- [22] David A Dickey and Sastry G Pantula. Determining the order of differencing in autoregressive processes. *Journal of Business & Economic Statistics*, 5(4):455–461, 1987.
- [23] Liang Du, Ruobin Gao, Ponnuthurai Suganthan, and David Wang. Bayesian Optimization Based Dynamic Ensemble for Time Series Forecasting. *Information Sciences*, 591, 01 2022.

- [24] Pablo Montero-Manso, George Athanasopoulos, Rob J. Hyndman, and Thiyanga S. Tagalaga. FFORMA: Feature-based forecast model averaging. *International Journal of Forecasting*, 36(1):86–92, 2020. M4 Competition.
- [25] Yuwei Fu, Di Wu, and Benoit Boulet. Reinforcement Learning Based Dynamic Model Combination for Time Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(6):6639–6647, Jun. 2022.
- [26] Sima Siami-Namini, Neda Tavakoli, and Akbar Siami Namin. A Comparison of ARIMA and LSTM in Forecasting Time Series. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1394–1401, 2018.
- [27] Domingos S. de O. Santos Júnior, João F.L. de Oliveira, and Paulo S.G. de Mattos Neto. An intelligent hybridization of ARIMA with machine learning models for time series forecasting. *Knowledge-Based Systems*, 175:72–86, 2019.
- [28] Ji-Long Chen and Guo-Sheng Li. Evaluation of support vector machine for estimation of solar radiation from measured meteorological variables. *Theoretical and Applied Climatology*, 115, 05 2013.
- [29] José Manuel Valente and Sebastián Maldonado. SVR-FFS: A novel forward feature selection approach for high-frequency time series forecasting using support vector regression. *Expert Systems with Applications*, 160:113729, 2020.
- [30] Pedro Henrique Borghi, Oleksandr Zakordonets, and João Paulo Teixeira. A COVID-19 time series forecasting model based on MLP ANN. *Procedia Computer Science*, 181:940–947, 2021.
- [31] Gang Song and Qun Dai. A novel double deep ELMs ensemble system for time series forecasting. *Knowledge-Based Systems*, 134:31–49, 2017.
- [32] Dimitrios D. Thomakos and Konstantinos Nikolopoulos. Forecasting Multivariate Time Series with the Theta Method. *Journal of Forecasting*, 34(3):220–229, 2015.
- [33] Steven Elsworth and Stefan Güttel. Time Series Forecasting Using LSTM Networks: A Symbolic Approach, 2020.
- [34] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [35] Paul Newbold. The principles of the box-jenkins approach. *Operational Research Quarterly (1970-1977)*, 26(2):397–412, 1975.
- [36] Vassilis Assimakopoulos and K. Nikolopoulos. The Theta model: A decomposition approach to forecasting. *International Journal of Forecasting*, 16:521–530, 10 2000.
- [37] José Fiorucci, Tiago Pellegrini, Francisco Louzada, and Fotios Petropoulos. The Optimised Theta Method. 03 2015.

- [38] Vladimir N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., 1995.
- [39] Hind Taud and JF Mas. Multilayer perceptron (mlp). *Geomatic approaches for modeling land change scenarios*, pages 451–455, 2018.
- [40] Barry J Wythoff. Backpropagation neural networks: a tutorial. *Chemometrics and Intelligent Laboratory Systems*, 18(2):115–155, 1993.
- [41] Jian Wang, Siyuan Lu, Shuihua Wang, and Yudong Zhang. A review on extreme learning machine. *Multimedia Tools and Applications*, 81, 05 2021.
- [42] Hasim Sak, Andrew W. Senior, and Françoise Beaufays. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. *CoRR*, abs/1402.1128, 2014.
- [43] Alex Graves, Marcus Liwicki, Santiago Fernández, Roman Bertolami, Horst Bunke, and Jürgen Schmidhuber. A novel connectionist system for unconstrained handwriting recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5):855–868, 2009.
- [44] Rob Hyndman and Yeasmin Khandakar. Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, 26, 07 2008.
- [45] Statsmodels contributors. The Theta Model, 2023. [Online; accessed 24-March-2023].
- [46] Ratnadip Adhikari and R. Agrawal. A combination of artificial neural network and random walk models for financial time series forecasting. *Neural Computing and Applications*, Accepted, 03 2013.
- [47] G.Peter Zhang. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50:159–175, 2003.