



Trabalho de Conclusão de Curso

Desagregação de energia baseada em aprendizagem profunda e transformada de wavelet

Eduardo Gomes dos Santos
egs@laccan.ufal.br

Orientadores:

Prof. Dr. André Luiz Lins de Aquino
Me. Geymerson dos Santos Ramos

Maceió, 22 de Fevereiro de 2022

Eduardo Gomes dos Santos

Desagregação de energia baseada em aprendizagem profunda e transformada de wavelet

Monografia apresentada como requisito parcial para obtenção do grau de Bacharel em Engenharia de Computação do Instituto de Computação da Universidade Federal de Alagoas.

Orientadores:

Prof. Dr. André Luiz Lins de Aquino

Me. Geymerson dos Santos Ramos

Maceió, 22 de Fevereiro de 2022

Monografia apresentada como requisito parcial para obtenção do grau de Bacharel em Engenharia de Computação do Instituto de Computação da Universidade Federal de Alagoas, aprovada pela comissão examinadora que abaixo assina.

Prof. Dr. André Luiz Lins de Aquino - Orientador
Instituto de Computação
Universidade Federal de Alagoas

Me. Geymerson dos Santos Ramos - Coorientador
Instituto de Computação
Universidade Federal de Alagoas

Prof. Dr. Rian Gabriel Santos Pinheiro - Examinador
Instituto de Computação
Universidade Federal de Alagoas

Prof. Dr. Heitor Soares Ramos Filho - Examinador
Instituto de Ciências Exatas
Universidade Federal de Minas Gerais

Catálogo na Fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 - 1767

S237d Santos, Eduardo Gomes dos.

Desagregação de energia baseada em aprendizagem profunda e transformada de wavelet / Eduardo Gomes dos Santos. – 2022.
25 f. : il.

Orientador: André Luiz Lins de Aquino.

Orientador: Geymerson dos Santos Ramos.

Monografia (Trabalho de conclusão de curso em Engenharia de Computação) - Universidade Federal de Alagoas, Instituto de Computação. Maceió, 2022.

Bibliografia: f. 23-25.

1. Wavelets (Matemática). 2. Energia - Desagregação. 3. *Non intrusive load monitoring*. 4. Aprendizagem profunda. I. Título.

CDU: 004.81:159.953.5

Agradecimentos

A minha família, em especial aos meus pais, Edcelma e Edvan, e minha irmã, Layza, por todo o esforço para que eu possa ter uma boa educação, além da dedicação e paciência comigo;

A todos os meus amigos que, apesar das batalhas da vida adulta, estão sempre comigo, me apoiando e me ajudando a crescer, de uma forma ou de outra;

A todos os membros do LaCCAN, por terem me aberto o mundo científico e me ensinarem o quão interessante pode ser a pesquisa.

Resumo

A desagregação de energia é uma área que busca identificar o consumo individual de diferentes aparelhos usando apenas o sinal agregado medido a partir de um único ponto. Este trabalho propõe uma rede neural treinada com dados reduzidos Wavelets para realizar a desagregação de energia. Além da desagregação, que geralmente apenas obtemos uma resposta binária identificando o momento de ativação do aparelho, também estamos interessados em estimar o valor de consumo do aparelho. Consideramos o conjunto de dados UK-DALE para realizar nossos experimentos, o qual contém dados de diferentes aparelhos de cinco casas da Inglaterra. Usando nossa abordagem, em comparação com outro trabalho bem estabelecido, alcançamos melhorias por aparelho de 27,8% (F_1 – score) no processo de desagregação e 11,4% (acurácia estimada) no valor de consumo do aparelho. Nossa principal contribuição foi identificar de forma satisfatória que os coeficientes de aproximação da transformada Wavelet são suficientes para estimar o consumo individual de eletrodomésticos.

Palavras-chave: Desagregação de energia, NILM, Transformada de Wavelet, Aprendizagem Profunda.

Abstract

Energy disaggregation is a field which seeks to identifying individual consumption of different appliances using only the aggregated signal measured from a single point. This work proposes a neural network trained with Wavelets reduced data to perform energy disaggregation. Besides the disaggregation, usually a binary answer by identifying the appliance activation moment, we are interested in estimating the appliance's consumption value. We consider the UK-DALE data set to perform our experiments, containing data from different appliances of five houses from England. Using our strategy, compared with another well-established work, we achieved improvements per appliance of 27.8% (F_1 -score) in the disaggregation process and 11.4% (estimated accuracy) in the appliance's consumption value. Our main contribution was to identify satisfactorily that the coefficients of approximation of the Wavelet transform are enough to estimate the individual consumption of household appliances.

Keywords: Energy disaggregation, NILM, Wavelet transform, Deep learning.

Sumário

Lista de Figuras	vii
Lista de Tabelas	viii
1 Introdução	1
1.1 Trabalhos Correlatos	2
1.2 Contribuições	4
1.3 Estrutura do texto	4
2 Nossa proposta	5
2.0.1 Pré-processamento usando Wavelet	5
2.0.2 Rede neural	9
2.0.3 Reconstrução do sinal desagregado	11
3 Resultados e Discussões	13
4 Considerações Finais	21
Referências bibliográficas	23

Lista de Figuras

2.1	Exemplo da Transformada Wavelet para uma decomposição de 3 níveis. Neste trabalho, usamos apenas 1 nível.	7
2.2	Coeficientes de aproximação e detalhe gerados a partir do sinal agregado. . . .	8
2.3	Funções Escala e Wavelet pré-calculadas a partir de um algoritmo de aproximação. .	9
2.4	Processo de treinamento	10
2.5	Processo de teste	11
3.1	Resultados da desagregação com nossa abordagem no lava-louça da casa 2. . . .	16
3.2	Média do EA com desvio padrão nas casas 5, 2 e 1.	18
3.3	Média do F_1 com desvio padrão nas casas 5, 2 e 1.	19

Lista de Tabelas

3.1	Dados usados na fase de teste	14
3.2	Dados usados na fase de treino	14
3.3	Resultados na casa 1	16
3.4	Resultados na casa 2	16
3.5	Resultados na casa 5	17
3.6	Resultados da nossa proposta nas casas 1, 2 e 5 com outros aparelhos.	17

1

Introdução

O crescimento populacional acelerado gera um aumento natural e direto na demanda mundial de energia. Estima-se que um aumento populacional de aproximadamente 11 bilhões em 2100 (Roser et al., 2013). Conseqüentemente, o uso de aparelhos eletrônicos acompanhará esse crescimento. Assim, é necessário um aumento da produção de energia elétrica, bem como a conscientização do consumidor. Embora a produção de eletricidade dependa de recursos naturais e de aspectos ambientais, a consciência do consumidor pode aumentar com o monitoramento adequado e eficaz do consumo. Esse monitoramento pode nos ajudar a entender quais eletrodomésticos consomem mais ou menos em nossas casas, permitindo-nos planejar estratégias de eficiência energética. Além disso, o monitoramento inteligente tem mostrado grande potencial na compreensão do comportamento baseado no consumo humano e na previsão do consumo.

A desagregação de energia permite identificar o consumo individual de diferentes aparelhos usando apenas o sinal agregado medido a partir de um único ponto. Esta abordagem visa estimar o consumo de cada aparelho no ambiente monitorado. Pode proporcionar, por exemplo, mais transparência nas contas de luz. Além disso, a desagregação em tempo real permite aplicações como detecção de ocupação em uma sala ou identificação de mau funcionamento de máquinas devido a vazamento de corrente elétrica.

Existem diferentes abordagens para realizar a desagregação de energia elétrica, como o Modelo de Markov Oculto (HMM) (Eddy, 2004), Processamento de Sinal de Gráfico (GSP) (Zhao et al., 2015) e Aprendizado Profundo (Kelly and Knottenbelt, 2015b). Propomos uma rede neural treinada com dados reduzidos Wavelets para realizar a desagregação de energia. Além da desagregação, geralmente uma resposta binária identificando o momento de ativação do aparelho, estamos interessados em estimar o valor de consumo do aparelho.

Existem trabalhos anteriores que não usam as principais abordagens discutidas acima. Por exemplo, Dinesh et al. (2015) propôs um método baseado em componentes espectrais não correlacionados de um sinal de consumo de energia. Esta abordagem divide o sinal de potência

em todos os componentes de subespaços usando uma expansão *Karhunen Loève* (KL). Em seguida, eles realizam vários estágios de eliminação iterativa para encontrar a estimativa posterior máxima de uma combinação de aparelhos ativados para uma janela de tempo alvo. No entanto, como os métodos baseados em HMM, aparelhos desconhecidos ou não direcionados levarão a uma piora no desempenho.

[Basu et al. \(2014\)](#) propõe um novo conjunto de meta-recursos para lidar com o NILM como um problema de classificação multi-rótulo. Outros modelos de aprendizado de máquina são usados, como o aluno da árvore de decisão, a máquina de vetores de suporte e o vizinho mais próximo multi-fardo. [Tabatabaei et al. \(2016\)](#) continua com a ideia de classificação multi-rótulo. Eles estendem os trabalhos anteriores empregando conjuntos de recursos no domínio do tempo e no domínio da wavelet para realizar a classificação. Esta abordagem mostra adaptabilidade, mas falha em aparelhos com padrões de carga complexos, como máquinas de lavar e de lavar louça.

O número crescente de dados disponíveis, junto com o alto poder computacional, possibilitou novas técnicas de aprendizado profundo de última geração para resolver problemas de desagregação. Nossa rede neural estima a relação não linear entre dois sinais comparados e realiza a desagregação. Podemos usar as saídas resultantes para diferenciar o consumo de energia de um aparelho da entrada agregada. Ao contrário das outras abordagens de desagregação (GPS ou HMM), que precisam de informações de ativação do dispositivo anterior, nossa rede neural pode aprender essas assinaturas, não exigindo nenhuma alimentação manual de informações para o modelo.

1.1 Trabalhos Correlatos

Vários trabalhos usam o modelo oculto de Markov (HMM) ([Parson et al., 2012](#); [Makonin et al., 2015](#); [Ji et al., 2019](#)) para realizar a desagregação de energia. Especificamente, o proposto por [Parson et al. \(2012\)](#) usam um processo de treinamento que consiste em modelar cada aparelho a partir de um modelo genérico, identificando assinaturas claras no consumo agregado por meio do algoritmo de maximização de expectativa e uma extensão do algoritmo de Viterbi, que filtra o sinal agregado e ignora a interferência de outros aplicativos. Eles subtraíram o consumo de um único aparelho do consumo agregado até desagregar todos os sinais.

Nossa proposta utiliza uma rede neural, que atualmente é o estado da arte em desagregação de energia, superando todos os trabalhos anteriores da literatura que utilizaram o HMM e suas variantes. Usando apenas redes neurais, encontramos vários trabalhos semelhantes aos propostos por [Shin et al. \(2019\)](#). Eles usaram duas redes neurais com camadas convolucionais e densas. O primeiro possui como saída uma janela do sinal desagregado. A segunda saída da rede neural tem a mesma dimensão que a primeira, mas cada neurônio fornece a probabilidade de um dispositivo estar ativo ou não. A combinação dessas duas saídas por meio da

multiplicação por elemento dá o resultado final. Com essa abordagem, eles obtiveram melhores resultados quando comparados a trabalhos anteriores da literatura.

Outro trabalho, proposto por [Zhang et al. \(2018\)](#), apresenta as abordagens: sequência a ponto (*seq2point*) e sequência a sequência (*seq2seq*), em que ambas usam camadas convolucionais e densas, mas o *seq2point* usa todos os dados de uma janela de sinal agregado para inferir um único ponto do sinal desagregado, ou seja, o ponto médio. Por outro lado, o método *seq2seq* prevê todo o sinal da janela. Seu trabalho também mostra que as redes neurais foram capazes de aprender diferentes tipos de assinaturas de aparelhos.

Temos uma contribuição anterior ([Santos et al., 2019](#)) que propõe uma arquitetura de rede neural simples sem um processo de redução de Wavelet para minimizar problemas de escalabilidade. Obtivemos resultados equivalentes aos da literatura durante a execução de nossas redes neurais em um dispositivo embarcado em um tempo muito mais curto. Como desvantagem, identificamos que o tempo geral de treinamento da rede neural em dispositivos embarcados é muito mais longo do que em computadores tradicionais. Além disso, verificamos o desempenho dos processos de treinamento nas novas casas. Identificamos que as redes neurais tendem a aprender assinaturas de dispositivos genéricos (apenas treinando com o mesmo dispositivo em casas diferentes) que poderiam ser usadas em casas nunca antes vistas.

Alguns trabalhos combinam Wavelets com redes neurais para realizar a desagregação de energia. O proposto por [Chang et al. \(2013\)](#) apresenta o teorema de Parseval para calcular um espectro de potência a partir dos coeficientes de transformada Wavelet, visando extrair características do espectro de potência que permitem a uma rede neural identificar eventos de transição de aparelhos. Em nosso método, não usamos nenhuma técnica para extrair características dos coeficientes de transformada Wavelet. Em vez disso, usamos esses coeficientes brutos como entrada para a rede neural. Além disso, a rede neural pode extrair mais características implícitas dos coeficientes.

[Gillis and Morsi \(2016\)](#) discutem uma nova abordagem baseada em aprendizado de máquina semissupervisionado e design Wavelet para prever os rótulos de classe para os dados não rotulados no aplicativo Non Intrusive Load Monitoring (NILM). Eles extraem os recursos do sinal combinando um conjunto de novas Wavelets às classes de carga, usando o projeto Wavelet e a análise de Procrustes.

Observe que a abordagem acima tem muitas semelhanças com a nossa proposta. Porém, utilizam outros algoritmos (menos sensíveis ao sinal do que Daubechies) com a transformada Wavelet para extrair características do sinal que auxiliam na identificação individual das cargas. Ao contrário disso, nossa rede neural extrai recursos e identifica aparelhos diretamente dos coeficientes Wavelet. Desta forma, as redes neurais podem extrair características que não identificaríamos manualmente, obtendo melhores resultados. Nosso trabalho complementa as soluções de redes neurais da literatura atual, pois fornece um método de pré-processamento para redução de dados melhorando os resultados.

1.2 Contribuições

Nossa principal contribuição foi identificar de forma satisfatória que os coeficientes de aproximação da transformada Wavelet são suficientes para estimar o consumo individual de eletrodomésticos. Além disso, quando usamos a transformada inversa nos coeficientes de aproximação inferidos, observamos que tais coeficientes representam o sinal desagregado em sua dimensão original, de forma equivalente e até melhor em alguns casos em comparação com outros trabalhos da literatura.

1.3 Estrutura do texto

Organizamos o restante deste trabalho da seguinte forma: A seção 1 apresentou o contexto e o impacto desse trabalho, os trabalhos relacionados e as contribuições que podemos fazer na literatura. A seção 2 discute os aspectos gerais de nossa abordagem, introduzindo a transformada de Wavelet discreta, definindo formalmente o problema e o uso de aprendizado profundo para desagregação de energia; A seção 3 mostra os resultados; e A seção 4 conclui nosso trabalho.

2

Nossa proposta

Para desagregação, consideramos o consumo agregado (C) de cada casa como a soma de n aparelhos na casa ($A = \{a_1, a_2, \dots, a_n\}$)

$$C = a_1 + a_2 + \dots + a_n + \varepsilon. \quad (2.1)$$

Como há interferências na leitura do consumo dos aparelhos, nós adicionamos um ruído Gaussiano ε (Santos et al. (2019)). Note que C é uma série temporal que representa o consumo agregado dos aparelhos na casa.

Nossa estratégia de desagregação consiste em: i) Pré-processamento dos dados baseado em wavelet; ii) uso da rede neural para extrair os sinais; e iii) reconstrução do sinal original através da transformada Wavelet inversa. As subseções a seguir descrevem essas etapas.

2.0.1 Pré-processamento usando Wavelet

No nosso modelo, cada dispositivo requer uma rede neural F . A combinação de todos os aparelhos $a_i \in A$ requer n redes neurais. A desagregação instantânea em um momento t para um ambiente com n aparelhos requer os seguintes cálculos

$$\begin{array}{c} F_1(C) \\ F_2(C) \\ \vdots \\ F_n(C) \end{array} \quad (2.2)$$

Para permitir a execução de várias redes neurais, consideramos uma fase de pré-processamento para reduzir a quantidade de dados por meio de uma transformada wavelet discreta (DWT) (Nikookar, 2013). Inicialmente, considere uma série temporal genérica x (em nosso caso representando o consumo agregado C ou o consumo individual de um aparelho

$a_i \in A$). Calculamos os coeficientes de aproximação $Q_\phi[j_0, k]$ e os coeficientes de detalhe $Q_\psi[j, k]$ definidos, respectivamente, por

$$Q_\phi[j_0, k] = \frac{1}{\sqrt{M}} \sum_t x[t] 2^{j_0/2} \phi[2^{j_0}t - k], \quad (2.3)$$

e

$$Q_\psi[j, k] = \frac{1}{\sqrt{M}} \sum_t x[t] 2^{j_0/2} \psi[2^j t - k], \text{ for } j \geq j_0, \quad (2.4)$$

onde $M = 2^J$ é o fator de normalização, limitado pela resolução desejada (J), onde $j = 1, 2, \dots, J$, $k = 1, 2, \dots, 2^j$, e j_0 é a resolução inicial.

As funções Escala (ϕ) e Wavelet (ψ) são respectivamente definidas por

$$\phi[2^j t - k] = \sum_{m=2k+n} h[m - 2k] \sqrt{2} \phi[2^{j+1}t - m], \quad (2.5)$$

e

$$\psi[2^j t - k] = \sum_{m=2k+n} g[m - 2k] \sqrt{2} \psi[2^{j+1}t - m], \quad (2.6)$$

$h[n]$ é o conjunto de coeficientes tal que $\phi[2^j t - k]$ pode ser expressada como uma soma ponderada de suas versões deslocadas $\phi[2^{j+1}t - k]$. Da mesma forma temos $g[n]$, mas usando $\psi[2^j t - k]$.

Nós definimos as funções ϕ e ψ baseada em alguma família de wavelet, e com isso nós podemos calcular os valores dos filtros $h[n]$ e $g[n]$ tal que as equações 2.5 e 2.6 sejam respeitadas. Finalmente, a Transformada de Wavelet Discreta Inversa (IDWT) é dada por

$$x[t] = \frac{1}{\sqrt{M}} \sum_k Q_\phi[j_0, k] \phi[2^{j_0}t - k] + \frac{1}{\sqrt{M}} \sum_{j=j_0} \sum_k Q_\psi[j, k] \psi[2^j t - k]. \quad (2.7)$$

Mallat (1989) mostrou que é possível realizar a decomposição e reconstrução do DWT usando bancos de filtros de 2 canais por meio de um algoritmo hierárquico conhecido como algoritmo piramidal. A figura 2.1 ilustra o processo de decomposição de um sinal x em 3 níveis ($J = 3$). Na Fig. 2.1, ${}^l Q_\phi$ e ${}^l Q_\psi$ representam, respectivamente, os coeficientes de aproximação e detalhe para o nível l . O processo começa com uma série temporal x com tamanho N . Aplicamos as Eqs. 2.3 e 2.4 gerando os coeficientes ${}^1 Q_\phi$ e ${}^1 Q_\psi$ com tamanho $N/2$. Repetimos toda a operação, onde ${}^l Q_\phi$ gera ${}^{2l} Q_\phi$ e ${}^{2l} Q_\psi$. Repetimos o processo resultando em ${}^3 Q_\phi$ e ${}^3 Q_\psi$ com o tamanho $N/8$. Depois que o sinal passa pelo filtro, fazemos uma redução em seu espectro de frequência pela metade, o que significa que metade da amostra pode ser descartada, de acordo

com a regra de Nyquist (Landau, 1967). Como há uma redução da série temporal pela metade com cada decomposição, Q_ϕ^3 tem 1/8 do tamanho de x .

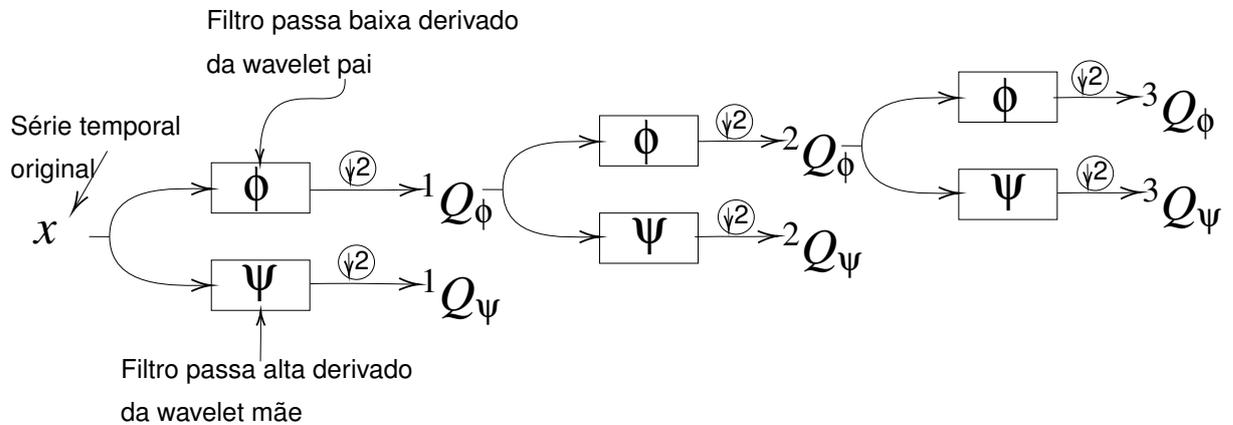


Figura 2.1: Exemplo da Transformada Wavelet para uma decomposição de 3 níveis. Neste trabalho, usamos apenas 1 nível.

A Fig. 2.2 ilustra um exemplo dos coeficientes de aproximação e detalhe gerados através da transformação Wavelet em um sinal agregado. Na primeira imagem (Fig. 2.2), temos 1000 amostras de um sinal agregado. Logo abaixo estão os 500 coeficientes de aproximação que calculamos a partir da série agregada. Finalmente, também temos os coeficientes de detalhe. É importante notar que mesmo com 500 amostras, os coeficientes de aproximação representam muito bem a série agregada. Os coeficientes de aproximação fornecem frequências mais baixas, nenhum ruído e representam melhor a série de tempo original.

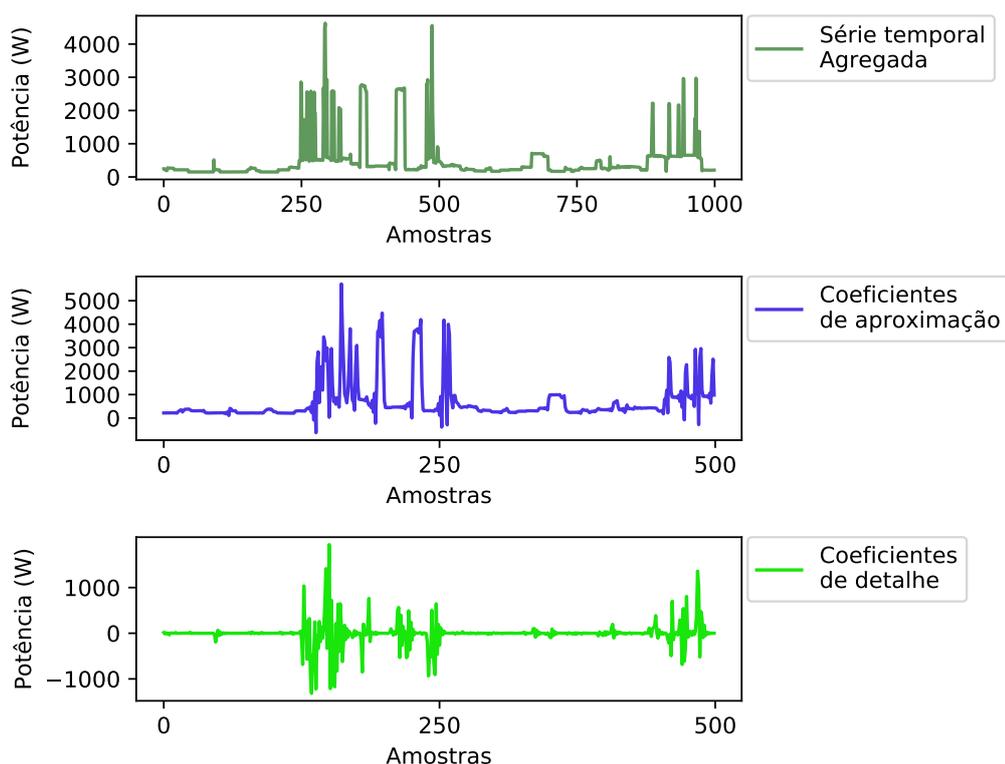


Figura 2.2: Coeficientes de aproximação e detalhe gerados a partir do sinal agregado.

Para escolher uma função Wavelet, usamos o algoritmo de sincronização dinâmica de tempo (DTW)(Müller, 2007), que mede a similaridade entre diferentes séries temporais. Para cada série agregada C das cinco casas do conjunto de dados UK-DALE, nós a re-amostramos usando diferentes funções Wavelet (Coiflets, Daubechies, Meyer, Haar, Reverse Biorthogonal, Symlet e suas variações) e usamos DTW para medir a similaridade entre os coeficientes de aproximação e a série agregada original C . Calculamos a média dos custos das cinco casas. O primeiro (custo mais baixo) foi Daubechies-18, custando 225,062; o segundo foi Daubechies-19, com 225.341; este último foi o Biorthogonal Reverso-1.5 e Daubechies-01, com 252,445 e 252,490, respectivamente.

Com base na função Wavelet de Daubechies-18, podemos transformar a série agregada C e todas as séries individuais $a_i \in A$ em seu coeficiente de aproximação Q_ϕ .

A família Daubechies deve obedecer a uma série de equações e propriedades fora do escopo deste trabalho. Para obter mais informações, consulte Rowe and Abbott (1995). Usamos um algoritmo de aproximação para calcular os coeficientes do filtro (Rowe and Abbott, 1995). Em particular, as funções Escala e Wavelet de Daubechies-18 estão na Fig. 2.3.

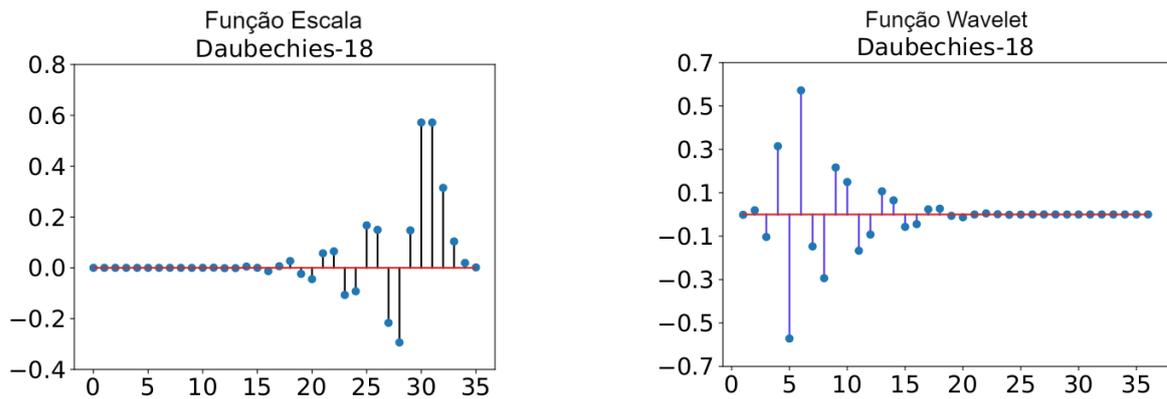


Figura 2.3: Funções Escala e Wavelet pré-calculadas a partir de um algoritmo de aproximação.

2.0.2 Rede neural

Nossa rede neural convolucional 1D é a mesma usada em Santos et al. (2019), consiste em oito camadas: seis convolucionais, uma densa e uma de saída. A entrada de nossa rede é a janela deslizante dos coeficientes aproximados das Wavelets (pré-processamento) do sinal agregado com tamanho 400 ($W = 400$). As três primeiras camadas convolucionais são idênticas. Eles têm 20 filtros, cada um com tamanho 12. As três camadas convolucionais a seguir apresentam um número reduzido de filtros (15, 10 e 7, respectivamente, de tamanho 9) em cada camada. A camada densa possui 512 neurônios, e a última apresenta três neurônios. A saída de nossa rede tem dimensão três que corresponde à estimativa de três pontos do coeficiente de sinal desagregado ($\hat{Q}_\phi^{a_i}$) do aparelho i para cada janela deslizante dos dados de entrada. Em resumo, o modelo infere um ponto médio $\hat{Q}_\phi^{a_i}[t + W/2]$ (a informação instantânea atual), um ponto a direita $\hat{Q}_\phi^{a_i}[t + 1 + W/2]$ (futuro), e o ponto esquerdo $\hat{Q}_\phi^{a_i}[t - 1 + W/2]$ (passado). A seção 2.0.3 apresenta a definição dos hiper-parâmetros da rede.

Processo de treinamento

O processo de treinamento de cada aparelho i , ilustrado na Figura 2.4, segue a sequência:

1. Cálculo dos coeficientes de aproximação Q_ϕ^C (sinal agregado) e $Q_\phi^{a_i}$ (sinal do aparelho individual) baseado em 2.3. Na Figura 2.4, esses coeficientes são os pontos dentro dos retângulos em baixo e no topo, respectivamente.
2. Uso das entradas na rede neural iF para cálculo do erro a ser usado no *backpropagation*. As entradas são a janela $Q_\phi^C[t : t + W]$ e os três pontos de $Q_\phi^{a_i}(t + W/2 - 1, t + W/2, t + W/2 + 1)$.
3. Execução do deslizamento da janela passo a passo ($t = t + 1$) de $t = 0$ até $t = N - W$, onde N é o tamanho de Q_ϕ^C e $Q_\phi^{a_i}$. Perceba que o Q_ϕ^C e $Q_\phi^{a_i}$ foram obtidos usando a Equação 2.3.

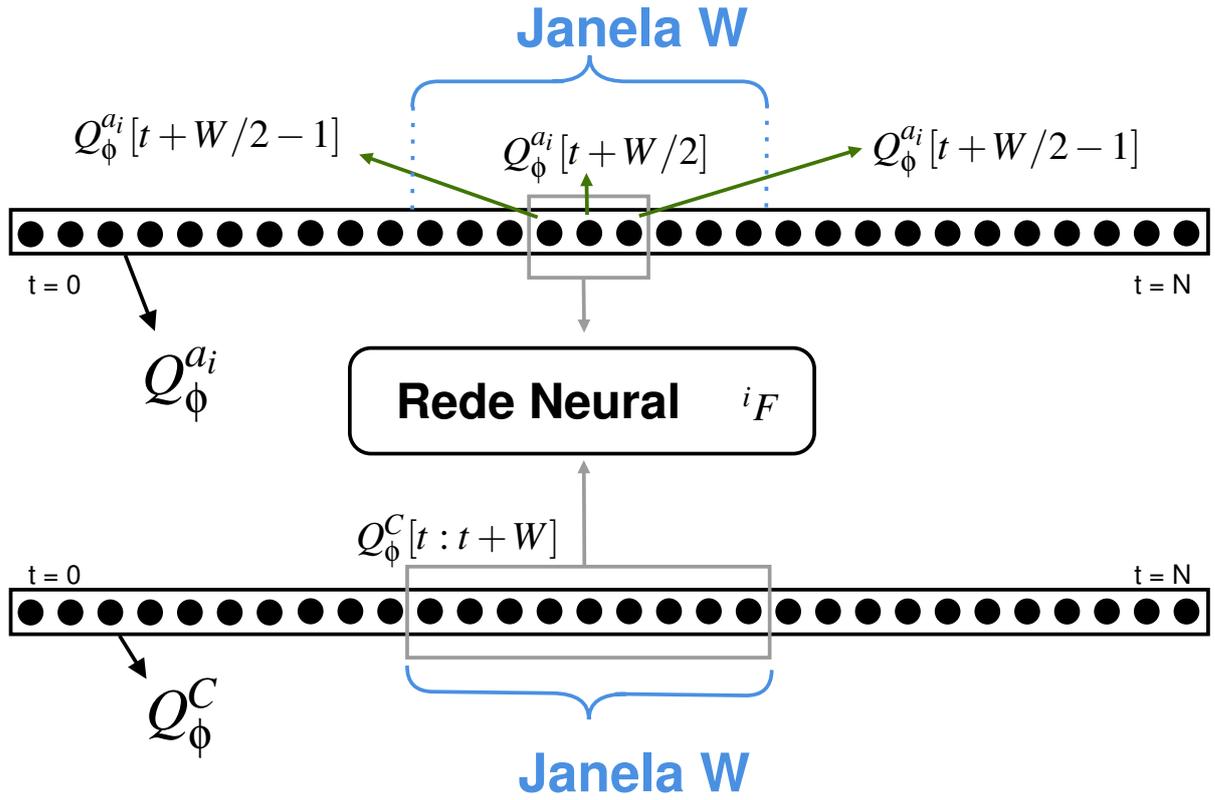


Figura 2.4: Processo de treinamento

Após o processo acima, para cada aparelho i , temos nossa rede neural treinada, e podemos realizar a desagregação para os dispositivos treinados.

Processo de teste

Usamos como entrada da rede neural apenas o sinal agregado Q_ϕ^C . Assim, executamos $i_F(Q_\phi^C)$ em uma janela deslizante de tamanho W . Em nosso processo, enumeramos cada passo W como $w_1, w_2, \dots, w_{t'}$. Dessa forma, em cada passo, a rede neural recebe w_j e gera 3 pontos. Portanto, ao final temos $w_{t'}$ amostras geradas onde cada uma possui 3 pontos:

$$i_F^{w_j}(Q_\phi^C[t:t+W]) \rightarrow k^{w_j}[1:3]. \tag{2.8}$$

onde $k^{w_j}[1:3]$ representa os 3 pontos previstos pela rede neural na etapa w_j . No final, temos que combinar todas as saídas w_j afim de reconstruir \hat{Q}_ϕ^{ai} . Fazemos isso usando a seguinte equação:

$$\hat{Q}_\phi^{ai}[j] = \begin{cases} k^{w_j}[1], & \text{if } j = 1 \\ (k^{w_{j-1}}[2] + k^{w_j}[1])/2, & \text{if } j = 2 \\ (k^{w_{j-2}}[3] + k^{w_{j-1}}[2])/2, & \text{if } j = t' + 1 \\ k^{w_{j-2}}[3], & \text{if } j = t' + 2 \\ (k^{w_{j-2}}[3] + k^{w_{j-1}}[2] + k^{w_j}[1])/3, & \text{outro} \end{cases} \tag{2.9}$$

onde t' é o número total de janelas e $\hat{Q}_\phi^{a_i}$ são os coeficientes inferidos. Observe que $\hat{Q}_\phi^{a_i}$ tem $t' + 2$ amostras. A relação de t' com N , é dada por $t' = N - W$. Portanto, $\hat{Q}_\phi^{a_i}$ tem o tamanho $N - W + 2$.

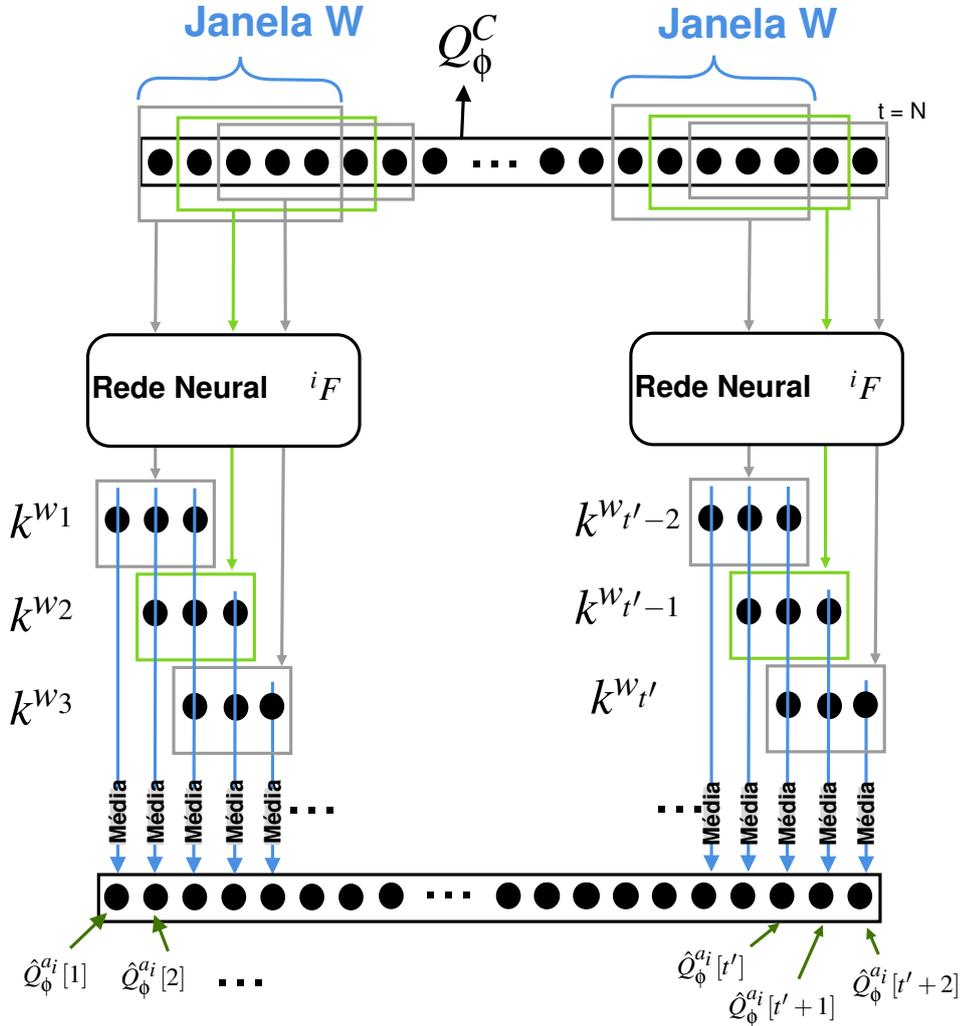


Figura 2.5: Processo de teste

A Fig. 2.5 representa o processo de teste e uso da rede neural i_F . Para cada janela w_j de Q_ϕ^C , o modelo inferirá 3 pontos de $\hat{Q}_\phi^{a_i}$, ou seja, $k^{w_j}[1 : 3]$. Depois disso, usamos a Eq. 2.9 para compor os coeficientes inferidos $\hat{Q}_\phi^{a_i}$ do aparelho a_i .

2.0.3 Reconstrução do sinal desagregado

Nosso processo de desagregação considera a entrada da rede neural apenas os coeficientes de aproximação do sinal agregado, que infere os coeficientes de aproximação da série desagregada. A última etapa é a reconstrução da série temporal em sua dimensão original (\hat{a}_i) a qual é realizada por meio da transformação Wavelet inversa conforme ilustrado na Eq. 2.7.

Em resumo, a execução completa do processo consiste em: i) Cálculo dos coeficientes de aproximação do sinal agregado usando transformada de Wavelet; ii) uso dos coeficientes

de aproximação na rede neural para a mesma inferir os coeficientes de aproximação do sinal desagregado; iii) o modelo infere os coeficientes de aproximação do sinal desagregado; e iv) o cálculo da transformada Wavelet inversa para estimar o sinal desagregado em sua dimensão original.

3

Resultados e Discussões

Para verificar o desempenho do nosso método, usamos o conjunto de dados UK-DALE ([Kelly and Knottenbelt, 2015a](#)), que possui dados de cinco casas da Inglaterra. O conjunto de dados possui o consumo de energia agregado registrado a cada segundo, e os dados no nível do aparelho estão com resolução de seis segundos. Seleccionamos as casas 1, 2 e 5 para avaliar nossa proposta. A casa 1 tem quatro moradores com 54 eletrodomésticos, e eles coletam os dados ao longo de 786 dias. As casas 2 e 5 possuem dois moradores cada, com 20 e 26 eletrodomésticos, respectivamente, e os dados coletados ao longo de 234 e 137 dias, respectivamente. Seleccionamos a máquina de lavar (WM), a lavadora-secadora (WD) e a lava-louças (DW) devido à complexidade de seu acionamento e porque o trabalho utilizado para comparar com nossa solução também os utiliza.

Perdas e atrasos de leitura podem ocorrer nas amostras durante a coleta de dados, ocasionando leituras dos sinais individuais não existentes no sinal agregado e perda de sincronicidade, impactando o processo de aprendizado da rede neural. Reduzimos a frequência de amostragem do sinal principal e individual para $1/60Hz$, criando duas séries temporais do mesmo tamanho para garantir que cada medição existe em ambas as séries. Treinamos nossas redes neurais dez vezes em cada casa. Este número de execuções já foi satisfatório, pois o desvio padrão foi mínimo, indicando que nossa rede convergiu.

O processo de treinamento considera as séries temporais de potência aparente para o sinal principal e a potência ativa para o individual, uma vez que o sinal principal de algumas casas não possui medidas de potência ativa. É essencial normalizar (faixa de 0 a 1) os dados para manter sua escala e evitar a saturação das funções de ativação.

Para aliviar o problema do gradiente de fuga ([Hochreiter, 1998](#)) e melhorar a aprendizagem da rede neural, usamos a função de ativação Unidade Linear Exponencial (ELU) ([Clevert et al.](#),

2015), definida como

$$u(n) = \begin{cases} n, & \text{if } n \geq 0 \\ \alpha(e^n - 1), & \text{if } n < 0. \end{cases} \quad (3.1)$$

A ELU é uma função linear se n for maior que 0, e se n for menor que 0, é exponencial de acordo com um valor de *alpha*. Usamos como padrão $\alpha = 1$.

Escolhemos o erro quadrático médio (MSE) como a função de perda da rede neural

$$L(\hat{Q}_\phi^{a_i}, Q_\phi^{a_i}) = \frac{1}{T} \sum_{t=1}^T (\hat{Q}_\phi^{a_i}[t] - Q_\phi^{a_i}[t])^2, \quad (3.2)$$

onde \hat{y} é a saída prevista no instante t e y é a saída real. Além disso, usamos o otimizador Adam Kingma and Ba (2014) para acelerar a convergência da rede e o tempo de treinamento. As taxas usadas para o momento de primeira e segunda ordem foram 0,9 e 0,999, respectivamente.

Treinamos as redes em uma NVIDIA GTX 1080 Ti, considerando 200 épocas com uma taxa de aprendizado de 1×10^{-4} , 100 épocas com uma taxa de 1×10^{-5} e 50 épocas com uma taxa de 1×10^{-6} . Usamos o algoritmo *backpropagation* para treinar o modelo Hecht-Nielsen (1992). A tabela 3.1 apresenta os segmentos de teste usados e a tabela 3.2 apresenta os segmentos de treino usados.

Tabela 3.1: Dados usados na fase de teste

Casa	Início	Fim
1	2014/04/25 23:22:00	2014/07/21 07:38:30
2	2013/09/20 00:43:10	2013/10/10 05:15:20
5	2014/09/07 23:13:10	2014/09/24 15:40:50

Tabela 3.2: Dados usados na fase de treino

Casa	Início	Fim
1	2012/11/09 00:00:00	2014/04/25 23:21:00
2	2013/05/22 00:00:00	2013/09/20 00:43:00
5	2014/06/29 00:00:00	2014/09/07 23:13:00

Em relação ao estado-da-arte, revisamos minuciosamente os trabalhos na literatura e identificamos alguns deles que relatam o desempenho de aparelhos com base no mesmo conjunto de dados. Kong et al. (2019) usa a classificação de amostra de $1/60Hz$, a mesma usada em nosso trabalho. Eles estendem a pesquisa e comparam com outros trabalhos com diferentes avaliações de amostras. Eles compararam com métodos de *benchmark* (Kelly and Knottenbelt,

2015b; Zhang et al., 2018), com o mesmo conjunto de dados e utilizando as mesmas métricas. Os resultados alcançados foram melhores que os dos *benchmarks*. Nosso objetivo é mostrar a eficácia do uso da Wavelet neste processo. Assim, decidimos usar o trabalho de Kong et al. como nossa linha de base de comparação. Em trabalhos futuros, gostaríamos de avaliar esta abordagem em diferentes classificações de amostras com diferentes conjuntos de dados e compará-la com outros trabalhos da literatura.

Escolhemos a solução de Kong et al. porque eles possuem a melhor qualidade de solução, resultados numéricos e relevância. Em suma, eles inferem toda a janela deslizante de consumo dos eletrodomésticos. Eles usam duas redes neurais para cada dispositivo. A primeira desagrega o sinal, semelhante a nós. A segunda realiza o pós-processamento para eliminar o ruído do sinal desagregado inferido pela primeira rede neural. Em comparação das técnicas, em vez de usar uma segunda rede neural (que requer outra etapa de treinamento), usamos a transformada Wavelet para reduzir o ruído na série de entrada. Também inferimos o consumo individual de cada dispositivo a cada momento (pois temos o sinal na resolução de 1 minuto, então nosso modelo infere a cada minuto), enquanto o trabalho de Kong et al. infere apenas o consumo médio da janela atual. Então, por exemplo, se temos uma janela de tamanho 300 e nosso sinal tem resolução de 1 minuto, então o modelo proposto por Kong et al. irá inferir o consumo individual dos aparelhos a cada 5 horas (300 minutos). Além disso, nosso modelo infere o mesmo consumo instantâneo três vezes. Essa execução dá ao nosso modelo a oportunidade de autocorreção na detecção de ativações de dispositivos.

Para avaliar o desempenho de desagregação, usamos a acurácia estimada EA e o F_1 – *score*. O EA informa a exatidão das estimativas de consumo de energia:

$$EA = 1 - \frac{\sum_{t=1}^T |\hat{y}_t - y_t|}{2 \sum_{t=1}^T y_t}, \quad (3.3)$$

onde \hat{y}_t é o consumo inferido pelo modelo no instante t , y_t é o consumo real e T é o tempo total. Observe que podemos obter \hat{y}_t de $\hat{Q}_\phi^{a_i}$ com a transformação de Wavelet inversa.

O F_1 – *score* enfatiza a precisão da identificação da ativação dos aparelhos-alvo, desconsiderando o quão precisas são as estimativas de consumo:

$$F_1 = \frac{2 \times \textit{precisao} \times \textit{recall}}{\textit{precisao} + \textit{recall}}, \quad (3.4)$$

com $\textit{precisao} = \frac{TP}{TP+FP}$ e $\textit{recall} = \frac{TP}{TP+FN}$. Os verdadeiros positivos (TP) representam casos em que a rede neural previu corretamente que um dispositivo estava LIGADO; os falsos positivos (FP) são as previsões LIGADAS, mas o dispositivo estava DESLIGADO; os falsos negativos (FN) são previsões erradas de um aparelho DESLIGADO; e o verdadeiro negativo (TN) representa as previsões corretas quando o dispositivo estava DESLIGADO. Essa métrica combina precisão e *recall* para trazer um único número que indica a qualidade geral dos modelos e funciona bem mesmo com conjuntos de dados que têm classes desproporcionais. No nosso caso,

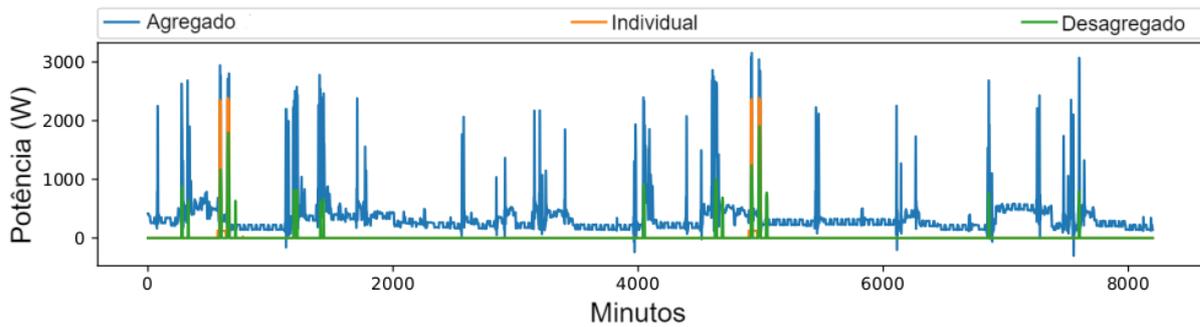


Figura 3.1: Resultados da desagregação com nossa abordagem no lava-louça da casa 2.

alguns aparelhos possuem ativações e longos períodos de inatividade. Portanto, é razoável usar o $F1 - score$. Os autores em Kong et al. (2019) não mencionaram o limite usado para calcular o $F1 - score$. Assim, usamos os valores padrão de NILMTK Batra et al. (2014). Inicialmente, a Fig. 3.1 apresenta um exemplo de resultados de inferência para ativação da máquina de lavar louça. Este exemplo mostra que a ativação da máquina de lavar louça tem dois picos de energia (linha laranja do sinal do aparelho original (a_i)). A rede neural inferiu o primeiro e o segundo pico de energia de forma satisfatória (linha verde do sinal desagregado (\hat{a}_i)). Perceba que, neste exemplo, a série temporal tem cerca de 8.000 pontos. Como o sinal tem resolução de 1 minuto, temos cerca de cinco dias.

Tabela 3.3: Resultados na casa 1

	DW						WM					
	TP	TN	FP	FN	EA	F1	TP	TN	FP	FN	EA	F1
Este trabalho	30	278	1	0	0.85	0.98	58	244	0	7	0.72	0.94
Kong et Al.	–	–	–	–	0.96	0.91	–	–	–	–	0.63	0.64

Tabela 3.4: Resultados na casa 2

	DW						WM					
	TP	TN	FP	FN	EA	F1	TP	TN	FP	FN	EA	F1
Este trabalho	18	50	1	1	0.85	0.94	7	61	0	2	0.65	0.87
Kong et Al.	–	–	–	–	0.92	0.90	–	–	–	–	0.50	0.52

Kong et al. avalia sua abordagem usando a resolução de 6 segundos e 1 minuto. As tabelas 3.3 - 3.5 apresentam os resultados sobre as casas e eletrodomésticos avaliados por Kong et al. com uma resolução de 1 minuto (a resolução exata usada neste artigo) e usada aqui para comparar com nossa proposta. Também incluímos os valores dos valores da matriz de confusão. Como Kong et al. não forneceu os valores para a resolução de 1 minuto, preenchemos a tabela com '-'. A primeira avaliação foi com a casa 1 e os aparelhos DW e WM (ver 3.3). Identificamos que Kong et al. obtiveram uma precisão estimada de 96% (EA) em comparação com nosso

Tabela 3.5: Resultados na casa 5

	DW						WD					
	TP	TN	FP	FN	EA	F1	TP	TN	FP	FN	EA	F1
Este trabalho	5	46	0	0	0.879	1	3	43	0	5	0.53	0.54
Kong et Al.	–	–	–	–	0.70	0.68	–	–	–	–	0.41	0.24

Tabela 3.6: Resultados da nossa proposta nas casas 1, 2 e 5 com outros aparelhos.

CASA 1						CASA 2						CASA 5					
KT		FR		MW		KT		FR		MW		KT		FR		MW	
EA	F_1	EA	F_1	EA	F_1	EA	F_1	EA	F_1	EA	F_1	EA	F_1	EA	F_1	EA	F_1
0.744	0.915	0.949	0.835	0.634	0.564	0.874	0.884	0.926	0.894	0.703	0.648	0.649	0.833	0.694	0.621	-0.489	0.406

resultado de 85% para máquinas de lavar louça. Esse comportamento ocorre porque a casa 1 possui muitos dados de alta qualidade, o que pode ter habilitado os modelos de Kong et al. aprenderem o consumo melhor que o nosso uma vez que reduzimos os dados sem conseguir aumentar ainda mais a sua qualidade. Outra razão é que o modelo proposto por Kong et al. infere apenas o consumo médio para cada janela, enquanto nosso modelo infere a janela inteira. Portanto, é mais desafiador ajustar o consumo do dispositivo em cada ponto da janela. Sobre a pontuação de F_1 , nossa proposta atingiu 98%, 7% maior do que a proposta por Kong et al.. Esse comportamento ocorre porque nosso modelo infere o mesmo consumo três vezes, permitindo que seja mais provável detectar uma ativação, de forma que a pontuação de F_1 de nosso modelo tende a ser maior. Para a máquina de lavar, nossa proposta foi melhor na EA e $F_1 - score$, com uma melhoria de 9% e 30%, respectivamente. Relembrando, pontuações mais altas de F_1 significam melhor detecção de ativação, enquanto um valor mais alto de EA significa uma melhor estimativa de consumo.

A segunda avaliação foi com a casa 2 nos aparelhos DW e WM. Para a máquina de lavar louça, registramos EA inferior em 7% e $F_1 - score$ superior em 4%. Este comportamento ocorre pelos mesmos motivos descritos para DW na casa 1. Considerando a máquina de lavar, também obtivemos melhores resultados para EA (15%) e $F_1 - score$ (35%). O WM na casa 2 se comporta mal por apresentar um comportamento parecido com ruído no meio da ativação do dispositivo. Ele aumentou nosso desvio padrão em WM (veja a figura 3.3).

Até agora, nas casas 1 e 2, nosso modelo apresentou um maior $F_1 - score$ e um EA semelhante, o que significa que nossa proposta é mais adequada para detectar ativação do que inferir o consumo do aparelho. A ativação da detecção é mais importante do que a previsão do consumo, pois, na maioria dos aparelhos, é possível obter o consumo médio utilizando as especificações do aparelho.

A avaliação subsequente foi com a casa 5 e os aparelhos DW e WD. Nossa proposta atingiu EA (17,9%) maior e F_1 (32%) maior para a máquina de lavar louça, e EA (12%) maior e F_1 (30%) maior para a lavadora-secadora. Esses comportamentos ocorrem porque o número de

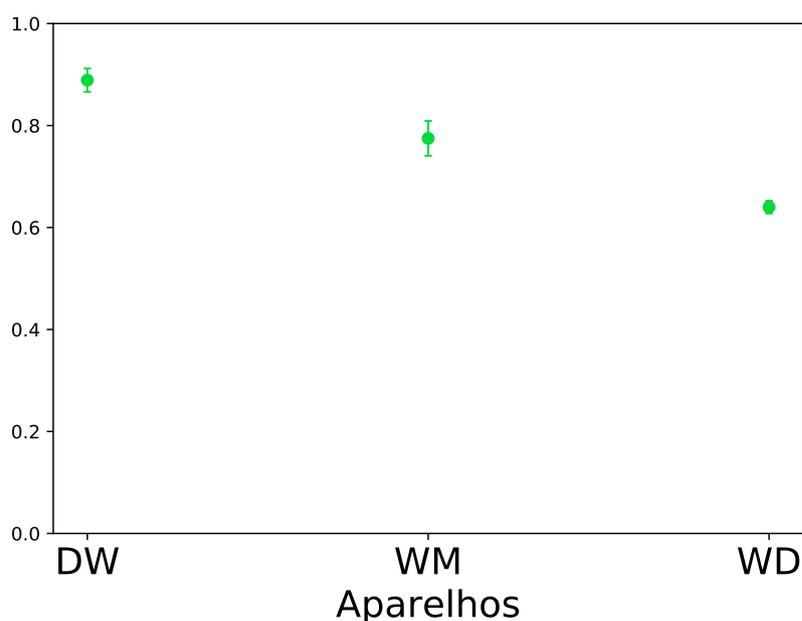


Figura 3.2: Média do EA com desvio padrão nas casas 5, 2 e 1.

amostras da casa 5 é menor que as anteriores e apresenta mais ruídos (51 amostras). Assim, nossa proposta suaviza o sinal através da transformada Wavelet, melhorando os resultados. Além disso, nossa janela de slide tem passo 1, aumentando a quantidade de dados testados.

Por fim, calculamos a média (Casas 1, 2 e 5) dos resultados para cada dispositivo testado. Para EA , os resultados obtidos são melhores para todos os dispositivos, com aumento de pontos percentuais de 2%, 11,5% e 12%, para DW, WM e WD, respectivamente. No F_1 , nossa proposta também foi superior em todos os dispositivos. O aumento em pontos percentuais foi de 14%, 32% e 30% para os dispositivos DW, WM e WD, respectivamente. Nosso modelo é mais adequado para detectar a ativação do aparelho, então o F_1 tende a ser maior, e os resultados confirmam isso. Nosso modelo também infere toda a janela de consumo individual. Assim, é mais provável que nosso modelo faça uma autocorreção ao inferir o consumo de uma janela. Nesse sentido, o modelo proposto por Kong et al. infere o consumo de uma janela apenas uma vez, levando a rede neural a cometer erros. Portanto, se o consumo na janela não tiver muitos detalhes complexos, nosso modelo será superior. Também consideramos a média geral para as três casas e os três dispositivos. Em média, nossa proposta melhora os pontos percentuais de EA em 7,9%. Corresponde a um aumento no desempenho de 11,4% (EA de 0,61 para 0,689). O aumento no desempenho de F_1 foi 31,2% (F_1 de 0,55 para 0,80). Nosso modelo foi muito melhor na casa 5. Assim, a média aumentou significativamente.

Para aumentar a confiabilidade de nossos experimentos, realizamos o treino e o teste 10 vezes para cada dispositivo em cada casa. Assim, obtemos a média e o desvio da pontuação EA e F_1 de todas as casas. Os resultados estão nas Figs 3.2 e 3.3.

Nas Figs. 3.2 e 3.3, podemos ver que os modelos de rede neural convergiram uma vez que o desvio padrão é pequeno, exceto para o F_1 na WM. Nesse caso, treinamos a rede neural com

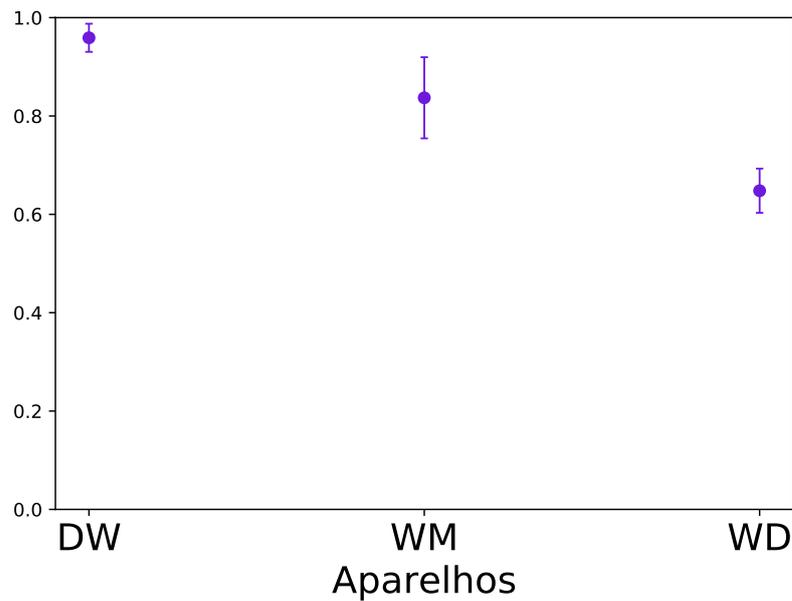


Figura 3.3: Média do F_1 com desvio padrão nas casas 5, 2 e 1.

muitas amostras, sem melhora devido ao conjunto de dados.

Com base no comportamento médio, nossa proposta é melhor do que a de Kong et al. Para melhor verificar o desempenho de nossa proposta, consideramos outros dispositivos (não usados por Kong et al.) Nas casas acima. Existem muitos dispositivos, mas a maioria deles tem apenas dois estados. Executamos nosso modelo apenas com chaleira (KT), geladeira (FR) e micro-ondas (MW). A tabela 3.6 apresenta os resultados. A casa 1 é a que tem mais dados para treinar redes neurais. Portanto, os modelos podem aprender a detectar os padrões de consumo de um dispositivo de forma mais genérica. Este comportamento justifica o bom desempenho da nossa proposta. Da mesma forma, a casa 2 também apresenta mais dados e os padrões de consumo dos aparelhos apresentam bom comportamento. Assim, a rede neural pode aprender as características de um aparelho de forma adequada. Observe que o MW (nas casas 1 e 2) é o aparelho onde o modelo obteve os menores valores das métricas. Ele possui padrões de consumo que não possuem detalhes exclusivos. Além disso, a duração da ativação varia (depende do comportamento humano), portanto, as métricas têm valores mais baixos para este dispositivo. Por outro lado, os modelos obtiveram bom desempenho para os demais aparelhos, uma vez que sua ativação é mais complexa e possui características únicas.

Casa 5, o conjunto de dados é pequeno. Desse modo, a rede neural tem dificuldade em aprender assinaturas de dispositivos. Obtemos um EA e F_1 de 65% e 83% para o KT. Nosso modelo foi capaz de detectar com sucesso as ativações do KT, uma vez que seu F_1 foi de alto valor. Porém, o EA foi menor, portanto nosso modelo não conseguiu inferir o consumo do aparelho nas ativações de forma muito satisfatória. Porém, podemos resolver este problema obtendo o consumo médio do KT de ativações, uma vez que este dispositivo é simples e possui apenas dois estados de ativação (ligado e desligado). Com o FR, os valores de F_1 e EA são

semelhantes. Para as ativações detectadas, o modelo inferiu o consumo de forma razoável. Vimos que tínhamos 23 ativações de RF nos dados de teste. Nosso modelo inferiu todas as ativações. No entanto, ele detectou algumas ativações falsas, que acabaram reduzindo o F_1 para 62%. O FR é um aparelho cíclico mas com baixo consumo quando comparado a outros aparelhos. Não é fácil detectar suas ativações quando outros aparelhos são ativados. No MW, nosso modelo teve o pior desempenho (impraticável). Como dissemos acima, este dispositivo tem uma assinatura sem características exclusivas e tem durações de ativação muito diferentes. Além disso, como na casa 5, o conjunto de dados é pequeno, exige melhor treinamento da rede neural com mais amostras.



Considerações Finais

Este trabalho propôs um processo de redução de dados por meio da transformada Wavelet para treinar redes neurais em problemas de monitoramento de carga não intrusiva. Nossa proposta prevê o sinal desagregado de máquina de lavar, lava-louças e máquina de lavar e secar roupa, superando trabalhos anteriores da literatura. Esses resultados confirmam que o pré-processamento da transformada Wavelet, com a função Wavelet Daubechies 18 para reduzir e reconstruir a série temporal de consumo, aumentou a eficiência de predição da rede neural. Podemos atribuir isso a uma diminuição significativa do ruído, que é os dados de alta frequência que não foram restaurados no processo de reconstrução.

Pesquisas futuras guiarão nosso trabalho no sentido de explorar e avaliar níveis mais altos de redução de dados usando a transformada Wavelet, o impacto no tempo de treinamento da rede neural e novas técnicas de redução de dados. Também planejamos investigar a transferência de conhecimento em redes neurais, permitindo a mesma rede neural em casas diferentes. Além disso, esperamos comparar diferentes abordagens com outros conjuntos de dados, identificar fraudes nas contas de eletricidade e explorar a detecção de falhas em diferentes ambientes com diversos aparelhos. Além disso, esperamos explorar os coeficientes de detalhe. Por fim, as informações de alta frequência também podem ser úteis para certos aparelhos com padrões temporais mais intrincados.

Alguns desdobramento deste projeto foram publicados nas seguintes conferências/periódicos:

- SANTOS, Eduardo; FREITAS, Christopher; AQUINO, André. Uma arquitetura de aprendizagem profunda para desagregação de energia. In: ESCOLA REGIONAL DE COMPUTAÇÃO BAHIA, ALAGOAS E SERGIPE (ERBASE) , 2019, Ilhéus. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2019 . p. 314-322.
- SANTOS, Eduardo G.; FREITAS, Christopher GS; AQUINO, Andre LL. A Deep Learning approach for Energy Disaggregation considering Embedded Devices. In: 2019 IX Brazilian

Symposium on Computing Systems Engineering (SBESC). IEEE, 2019. p. 1-8.

- SANTOS, Eduardo; RAMOS, Geymerson; AQUINO, Andre. An Energy Disaggregation Approach Based on Deep Neural Network and Wavelet Transform. IEEE Transactions on Industrial Informatics, 2021. doi: 10.1109/TII.2021.3130416.

Por se tratar de um Trabalho de Conclusão de Curso, é importante ressaltar o aprendizado obtido, através do estudo dos assuntos expostos e suas aplicações, ajudando assim na construção e avanço da fronteira do conhecimento.

Este período foi muito importante para o meu crescimento intelectual, profissional e pessoal. No âmbito profissional foi fundamental no desenvolvimento da minha capacidade técnica de superar desafios, me levando até a conseguir um estágio de 6 meses na Google. Outro benefício é que trabalhei com grandes pesquisadores pertencentes ao LaCCAN, em especial com meu orientador Prof Dr. André Luis Lins de Aquino, onde já tive a oportunidade de acompanhar a publicação de três artigos.

Referências bibliográficas

- Kaustav Basu, Vincent Debusschere, Seddik Bacha, Ujjwal Maulik, and Sanghamitra Bondyopadhyay. Nonintrusive load monitoring: A temporal multilabel classification approach. *IEEE Transactions on industrial informatics*, 11(1):262–270, 2014.
- Nipun Batra, Jack Kelly, Oliver Parson, Haimonti Dutta, William Knottenbelt, Alex Rogers, Amarjeet Singh, and Mani Srivastava. Nilmtk: an open source toolkit for non-intrusive load monitoring. In *Proceedings of the 5th international conference on Future energy systems*, pages 265–276. ACM, 2014.
- Hsueh-Hsien Chang, Kuo-Lung Lian, Yi-Ching Su, and Wei-Jen Lee. Power-spectrum-based wavelet transform for nonintrusive demand monitoring and load identification. *IEEE Transactions on Industry Applications*, 50(3):2081–2089, 2013.
- Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). In Yoshua Bengio and Yann LeCun, editors, *4th Int. Conf. on Learn. Represent., ICLR 2016*, 2015.
- Chinthaka Dinesh, Buddhika W Nettasinghe, Roshan Indika Godaliyadda, Mervyn Parakrama B Ekanayake, Janaka Ekanayake, and Janaka V Wijayakulasooriya. Residential appliance identification based on spectral information of low frequency smart meter measurements. *IEEE Transactions on smart grid*, 7(6):2781–2792, 2015.
- Sean R Eddy. What is a hidden Markov model? *Nature Biotechnology*, 22(10):1315–1316, oct 2004.
- Jessie M Gillis and Walid G Morsi. Non-intrusive load monitoring using semi-supervised machine learning and wavelet design. *IEEE Transactions on Smart Grid*, 8(6):2648–2655, 2016.
- Robert Hecht-Nielsen. Theory of the backpropagation neural network. In *Neural networks for perception*, pages 65–93. Elsevier, 1992.

- Sepp Hochreiter. The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(2):107–116, oct 1998.
- TY Ji, L Liu, TS Wang, WB Lin, MS Li, and QH Wu. Non-intrusive load monitoring using additive factorial approximate maximum a posteriori based on iterative fuzzy c-means. *IEEE Transactions on Smart Grid*, 2019.
- Jack Kelly and William Knottenbelt. The uk-dale dataset, domestic appliance-level electricity demand and whole-house demand from five uk homes. *Scientific data*, 2:150007, 2015a.
- Jack Kelly and William Knottenbelt. Neural nilm: Deep neural networks applied to energy disaggregation. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 55–64. ACM, 2015b.
- Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- Weicong Kong, Zhao Yang Dong, Bo Wang, Junhua Zhao, and Jie Huang. A practical solution for non-intrusive type ii load monitoring based on deep learning and post-processing. *IEEE Transactions on Smart Grid*, 11(1):148–160, 2019.
- H. J. Landau. Sampling, data transmission, and the nyquist rate. *Proceedings of the IEEE*, 55(10):1701–1706, 1967.
- Stephen Makonin, Fred Popowich, Ivan V Bajić, Bob Gill, and Lyn Bartram. Exploiting hmm sparsity to perform online real-time nonintrusive load monitoring. *IEEE Transactions on smart grid*, 7(6):2575–2585, 2015.
- Stephane G Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7): 674–693, 1989.
- Meinard Müller. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84, 2007.
- Homayoun Nikookar. *Wavelet radio: adaptive and reconfigurable wireless systems based on wavelets*. Cambridge University Press, 2013.
- Oliver Parson, Siddhartha Ghosh, Mark Weal, and Alex Rogers. Non-intrusive load monitoring using prior models of general appliance types. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.
- Max Roser, Hannah Ritchie, and Esteban Ortiz-Ospina. World population growth, 2013. Most recent substantial revision in May 2019.

Alistair CH Rowe and Paul C Abbott. Daubechies wavelets and mathematica. *Computers in Physics*, 9(6):635–648, 1995.

Eduardo G. Santos, Cristopher G. S. Freitas, and Andre L. L. Aquino. A deep learning approach for energy disaggregation considering embedded devices. In *2019 IX Brazilian Symposium on Computing Systems Engineering (SBESC)*, pages 1–8, 2019.

DOI [10.1109/SBESC49506.2019.9046095](https://doi.org/10.1109/SBESC49506.2019.9046095).

Changho Shin, Sunghwan Joo, Jaeryun Yim, Hyoseop Lee, Taesup Moon, and Wonjong Rhee. Subtask gated networks for non-intrusive load monitoring. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1150–1157, 2019.

Seyed Mostafa Tabatabaei, Scott Dick, and Wilsun Xu. Toward non-intrusive load monitoring via multi-label classification. *IEEE Transactions on Smart Grid*, 8(1):26–40, 2016.

Chaoyun Zhang, Mingjun Zhong, Zongzuo Wang, Nigel Goddard, and Charles Sutton. Sequence-to-point learning with neural networks for non-intrusive load monitoring. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Bochao Zhao, Lina Stankovic, and Vladimir Stankovic. Blind non-intrusive appliance load monitoring using graph-based signal processing. In *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 68–72. IEEE, 2015.