

Desagregação de energia baseada em Deep Learning e Transformação de Wavelet

Eduardo G. Santos¹, André L. L. Aquino²

¹Universidade Federal de Alagoas (UFAL)
Maceió – AL – Brazil

{egs, alla}@laccan.ufal.br

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. 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.*

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. 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.*

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]. Consequentemente, 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. 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. Isso pode proporcionar, por exemplo, mais transparência nas contas de luz.

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. A rede neural artificial 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 (Processamento de Sinal de Gráfico (GSP) [Zhao et al. 2015] ou Modelo de Markov Oculto (HMM) [Eddy 2004]), que precisam de informações de ativação do dispositivo anterior, a rede neural pode aprender essas assinaturas, não exigindo nenhuma alimentação manual de informações para o modelo.

Propomos uma rede neural treinada com dados reduzidos através da transformada de Wavelets para realizar a desagregação de energia. 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. Além da *Excelência Acadêmica no Encontro de IC* da UFAL (2020), este trabalho gerou as seguintes publicações:

1. Artigo no periódico IEEE Transactions on Industrial Informatics (**Qualis-2020 A1, Impact Factor 10.215**) [Santos et al. 2021].
2. Artigo no evento internacional SBESC (**Qualis-2019 B2**) [Santos et al. 2019b].
3. Artigo no evento regional ERBASE 2019 (**best paper**) [Santos et al. 2019a].

2. Trabalhos Correlatos

Considerando o uso de redes neurais para o problema em tela, encontramos vários trabalhos semelhantes aos propostos por [Shin et al. 2019]. Eles usaram duas redes neurais com camadas convolucionais e densas. A primeira 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.

Outro trabalho, proposto por [Zhang et al. 2018], apresenta as abordagens: sequência para ponto (*seq2point*) e sequência para sequência (*seq2seq*), em que ambas usam camadas convolucionais e densas, mas o *seq2point* usa todos os dados de uma janela de tamanho fixo do sinal agregado para inferir um único ponto do sinal desagregado, o qual corresponde ao ponto médio da janela do sinal individual de um determinado aparelho. 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.

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. [Gillis and Morsi 2016] discutem

uma nova abordagem baseada em aprendizado de máquina semi-supervisionado e design Wavelet para prever os rótulos de classe para os dados não rotulados no aplicativo NILMTK [Batra et al. 2014]. 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 com a transformada Wavelet para extrair características do sinal que auxiliam na identificação individual das cargas. Ao contrário disso, nossa rede neural extrai informações e identifica aparelhos diretamente dos coeficientes Wavelet. Desta forma, as redes neurais podem extrair características que não identificaríamos manualmente, obtendo melhores resultados. Nosso trabalho avança no estado da arte em relação às soluções de redes neurais encontradas na literatura atual, pois fornece um método de pré-processamento para redução de dados melhorando significativamente os resultados.

3. Metodologia

Para desagregação, consideramos o consumo agregado $C = a_1 + a_2 + \dots + a_n + \epsilon$ de cada casa como a soma de n aparelhos na casa $A = \{a_1, a_2, \dots, a_n\}$, onde a_i representa uma série temporal do consumo individual de um determinado aparelho i . Como há interferências na leitura do consumo dos aparelhos, nós adicionamos um ruído Gaussiano ϵ nessa representação. Note que C é uma série temporal que representa o consumo agregado dos aparelhos na casa.

Uma visão geral da nossa técnica é representada na Figura 1. No passo 1) temos o uso da transformada de wavelet na serie temporal do consumo agregado. No passo 2) obtemos os coeficientes de aproximação e de detalhe. No passo 3) usamos os coeficientes de aproximação para treino/teste da rede neural. No passo 4) a rede neural infere os coeficientes de aproximação de um determinado aparelho. Nos passos 5) e 6), a transformada de wavelet inversa é usada nos coeficientes inferidos afim de obter a série temporal desagregada em sua dimensão original. Em resumo, os coeficientes de aproximação do sinal agregado são obtidos através da transformada de wavelet discreta e usados como entrada da rede neural. Com isso, a rede neural irá inferir os coeficientes de aproximação do consumo individual de um determinado aparelho, logo após usamos a transformada de wavelet inversa para transformar os coeficientes inferidos no sinal desagregado do aparelho alvo. Note que cada aparelho deve ter uma rede neural associada para fazer a sua desagregação.

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 da base de dados, nós a re-amostramos usando diferentes funções Wavelet (Coiflets, Daubechies, Meyer, Haar, Reverse Biorthogonal, Symlet e suas variações) e usamos o DTW para medir a similaridade entre os coeficientes de aproximação e a série agregada original C . Calculamos a média dos custos das casas e a que obteve maior similaridade foi usando a Daubechies-18, portanto a função escolhida para usarmos.

Nossa rede neural convolucional 1D consiste em oito camadas: seis convolucionais, uma densa e uma de saída. A entrada de nossa rede é uma janela deslizante de tamanho fixo (400 amostras) dos coeficientes de aproximação obtidos através da Transformada

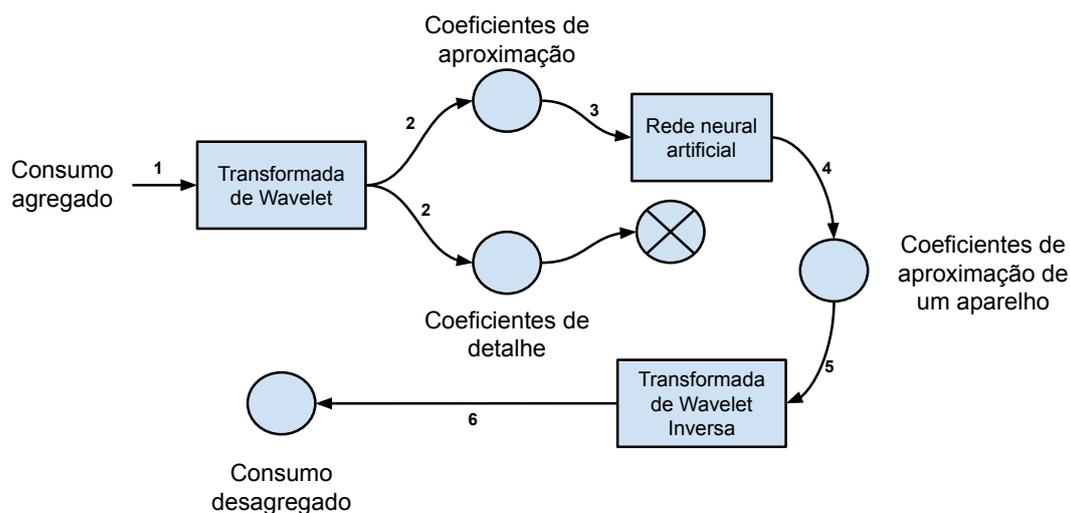


Figura 1. Visão geral de nossa técnica de desagregação de energia.

de Wavelet no sinal agregado. As três primeiras camadas convolucionais são idênticas. Eles têm 20 filtros, cada um com tamanho 12. As três camadas convolucionais a seguir apresentam 15, 10 e 7 filtros, 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, a qual se assemelha a modelagem *seq2point* ([Zhang et al. 2018]). Em resumo, o modelo infere um ponto médio da janela correspondente no sinal individual (semelhante a *seq2point*), e além disso infere os pontos adjacentes, que são o ponto imediatamente anterior e posterior. Na Seção 4 apresentaremos os demais hiper-parâmetros da rede neural.

4. Resultados e discussão

Para avaliar o desempenho do nosso método, usamos um conjunto de dados público chamado UK-DALE [Kelly and Knottenbelt 2015b], 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. Selecionamos 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. Selecionamos 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 ocorreram durante a coleta de dados na construção desse conjunto de dados, dessa forma ocasionando leituras dos sinais individuais não existentes no sinal agregado e perda de sincronicidade, os quais pode impactar negativamente o processo de aprendizado da rede neural. Portanto, 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.

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], e usamos o erro quadrático médio (MSE) como a função de perda da rede neural. 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]. Os segmentos de teste usados foram:

Casa 1 - Início 2014/04/25 23:22:00 e fim 2014/07/21 07:38:30;

Casa 2 - Início 2013/09/20 00:43:10 e fim 2013/10/10 05:15:20; e

Casa 5 - Início 2014/09/07 23:13:10 e fim 2014/09/24 15:40:50.

Já para o treino utilizamos os seguintes segmentos:

Casa 1 - Início 2012/11/09 00:00:00 e fim 2014/04/25 23:21:00;

Casa 2 - Início 2013/05/22 00:00:00 e fim 2013/09/20 00:43:00; e

Casa 5 - Início 2014/06/29 00:00:00 e fim 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] usam 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 2015a, 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. para 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

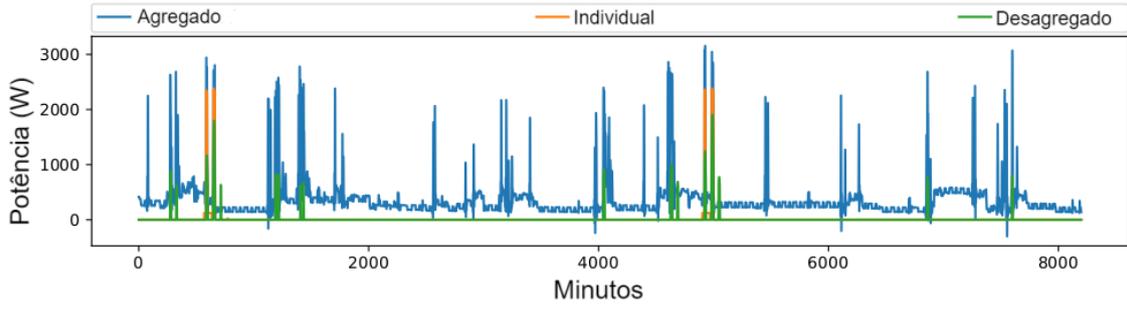


Figura 2. Desagregação do lava-louças da CASA 2.

F_1 – score. O EA informa a exatidão das estimativas de consumo de energia:

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

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 a partir dos coeficientes inferidos pela rede neural usando a transformação de Wavelet inversa neles.

O F_1 – score enfatiza a precisão da identificação da ativação dos aparelhos-alvo, desconsiderando o quão precisas são as estimativas de consumo. Ele é dado por

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

com $precisao = \frac{TP}{TP+FP}$ e $recall = \frac{TP}{TP+FN}$. 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, alguns aparelhos possuem ativações e longos períodos de inatividade. Portanto, é razoável usar o F_1 – score.

Inicialmente, a Figura 2 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.

A tabela 1 apresenta os resultados sobre as CASAS 1, 2 e 5, e os eletrodomésticos máquina de lavar (WM), a lava-louças (DW) e lavadora-secadora (WD), 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, Identificamos que Kong et al. obtiveram uma precisão estimada de 96% (EA) em comparação com nosso resultado de 85% para máquina de lavar louça. Esse comportamento ocorre porque a CASA 1 possui muitos dados de alta qualidade, o que pode ter habilitado os modelos de

Tabela 1. Resultados CASAS 1, 2 e 5

	DW					WM / WD para a CASA 5						
	TP	TN	FP	FN	EA	F1	TP	TN	FP	FN	EA	F1
CASA 1												
Nosso 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
CASA 2												
Nosso 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
CASA 5												
Nosso 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

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.

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 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 (CASA 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 padrão do EA e F_1 de todas as casas. Para o DW, obtivemos um desvio padrão de 0,022 e 0,028 (EA e f_1). No WM obtivemos 0,034 e 0,082 (EA e f_1). Por fim, no WD temos 0,012 e 0,045 (EA e f_1). Analisando esses valores, 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 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 2 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.

Tabela 2. Resultados para as 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

Na 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.

5. Conclusão

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 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.

O projeto para reprodução desse experimento está em repositório aberto disponível para download em <https://gitlab.com/eduardoGo/nilm-wavelet>. 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.

Referências

- Batra, N., Kelly, J., Parson, O., Dutta, H., Knottenbelt, W., Rogers, A., Singh, A., and Srivastava, M. (2014). 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.
- Chang, H.-H., Lian, K.-L., Su, Y.-C., and Lee, W.-J. (2013). Power-spectrum-based wavelet transform for nonintrusive demand monitoring and load identification. *IEEE Transactions on Industry Applications*, 50(3):2081–2089.
- Clevert, D., Unterthiner, T., and Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). In Bengio, Y. and LeCun, Y., editors, *4th Int. Conf. on Learn. Represent., ICLR 2016*.
- Eddy, S. R. (2004). What is a hidden Markov model? *Nature Biotechnology*, 22(10):1315–1316.

- Gillis, J. M. and Morsi, W. G. (2016). Non-intrusive load monitoring using semi-supervised machine learning and wavelet design. *IEEE Transactions on Smart Grid*, 8(6):2648–2655.
- Hecht-Nielsen, R. (1992). Theory of the backpropagation neural network. In *Neural networks for perception*, pages 65–93. Elsevier.
- Hochreiter, S. (1998). 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.
- Kelly, J. and Knottenbelt, W. (2015a). 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.
- Kelly, J. and Knottenbelt, W. (2015b). The uk-dale dataset, domestic appliance-level electricity demand and whole-house demand from five uk homes. *Scientific data*, 2:150007.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kong, W., Dong, Z. Y., Wang, B., Zhao, J., and Huang, J. (2019). 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.
- Müller, M. (2007). Dynamic time warping. *Information retrieval for music and motion*, pages 69–84.
- Roser, M., Ritchie, H., and Ortiz-Ospina, E. (2013). World population growth. Most recent substantial revision in May 2019.
- Santos, E., Freitas, C., and Aquino, A. (2019a). Uma arquitetura de aprendizagem profunda para desagregação de energia. In *Anais da XIX Escola Regional de Computação Bahia, Alagoas e Sergipe*, pages 314–322, Porto Alegre, RS, Brasil. SBC.
- Santos, E., Ramos, G., and Aquino, A. (2021). An energy disaggregation approach based on deep neural network and wavelet transform. *IEEE Transactions on Industrial Informatics*, pages 1–1.
- Santos, E. G., Freitas, C. G. S., and Aquino, A. L. L. (2019b). A deep learning approach for energy disaggregation considering embedded devices. In *2019 IX Brazilian Symposium on Computing Systems Engineering (SBESC)*, pages 1–8.
- Shin, C., Joo, S., Yim, J., Lee, H., Moon, T., and Rhee, W. (2019). Subtask gated networks for non-intrusive load monitoring. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1150–1157.
- Zhang, C., Zhong, M., Wang, Z., Goddard, N., and Sutton, C. (2018). Sequence-to-point learning with neural networks for non-intrusive load monitoring. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.
- Zhao, B., Stankovic, L., and Stankovic, V. (2015). 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.