

Data Characterization for Electrical Load Disaggregation using Supervised Learning

Wesley A. Souza¹, Tiago A. Almeida¹

¹Departamento de Computação – Universidade Federal de São Carlos (UFSCar)
Rodovia João Leme dos Santos, SP-264, Km 110 – 18052-780 – Sorocaba – SP – Brazil

wesleyangelino@ieee.org, talmeida@ufscar.br

Abstract. *Electricity is an essential resource and its optimized consumption becomes a growing demand in modern power grids. The recent evolution of micro-controllers allowed the immersion of artificial intelligence into a new metering generation called cognitive meters. Their main feature is the capability to perform energy disaggregation of loads. Consequently, good data representation is required to identify the appliance accurately, so that electrical measurement readings may be used to identify the appliances. In this context, this paper presents a data characterization approach for load disaggregation through 140 features. To validate our proposal, a real data set from 28 appliances were created and used for training three traditional classification algorithms. The results indicate that it is possible to perform an adequate load disaggregation in real-time scenarios, creating a significant number of possibilities to explore additional functionalities, as well as computational intelligence techniques for the cognitive meters.*

Resumo. *A energia elétrica é um serviço essencial e o seu uso otimizado é uma demanda cada vez maior nas redes elétricas modernas. A recente evolução dos microcontroladores permitiu a imersão de inteligência artificial em uma nova geração de medidores, chamados de medidores cognitivos. A principal característica do medidor cognitivo é a possibilidade de realizar a desagregação do consumo de energia elétrica por cargas. Para que isso seja possível, é necessário utilizar uma boa representação para os dados, de tal forma que leituras provenientes de medições elétricas possam ser usadas para identificar cada equipamento elétrico. Neste contexto, este artigo apresenta uma proposta de caracterização de dados para a desagregação do consumo por meio de 140 atributos. Para validar essa proposta, foi construído um conjunto de dados reais coletados a partir de 28 equipamentos elétricos, os quais foram usados para treinar três métodos tradicionais de classificação. Com base nos resultados obtidos, conclui-se que é possível realizar a efetiva desagregação de cargas em cenários reais e em tempo quase real, abrindo um grande conjunto de possibilidades para explorações de novos recursos e empregos de técnicas de inteligência computacional em medidores cognitivos.*

1. Introdução

A energia elétrica é um pilar essencial para o desenvolvimento socioeconômico [Makonin et al. 2013, Barai et al. 2015, Platt et al. 2015] e recentemente tem-se buscado meios otimizados e sustentáveis para usá-la de forma eficiente [Platt et al. 2015]. Neste sentido, intensifica-se a necessidade de detalhar o consumo da energia para oferecer alternativas que aumentem a eficiência energética e que aprimore a gestão da

rede elétrica [Makonin et al. 2013]. Neste contexto, surgem os medidores de energia elétrica como elementos importantes no modelo das modernas redes inteligentes [Barai et al. 2015, Souza et al. 2015]. Os medidores são utilizados em qualquer ponto do sistema, tanto na geração como na transmissão, distribuição ou no uso final de energia elétrica. Como estes pontos de medição necessitam informar dados confiáveis e precisos para a gestão de energia elétrica, os equipamentos de medição devem ser capazes de: armazenar dados e transmiti-los a centrais de controle e monitoramento; integrar-se com os dispositivos elétricos existentes no sistema; realizar procedimentos remotamente; e permitir a implantação de recursos com “inteligência” embarcada [Opris and Caracasian 2013a, Barai et al. 2015].

Diversas propostas vêm sendo feitas para apresentar evoluções que de fato atribuam inteligência ao medidor, sendo que um problema que vem recebendo considerável destaque é a desagregação do consumo de energia por equipamento ligado na rede elétrica [Dong et al. 2012, Opris and Caracasian 2013a, Makonin et al. 2013, Basu et al. 2015, Garcia et al. 2017]. Esse desafio corresponde ao processo de identificar as cargas através da aquisição de grandezas elétricas em um certo tempo [Rathmair and Haase 2013]. Assim, podem ser estabelecidos mecanismos para identificar quais cargas (equipamentos elétricos) foram ligadas ou desligadas no ponto de consumo de energia [Abubakar et al. 2017].

A desagregação efetiva do consumo de energia por equipamento elétrico permitirá: auxiliar o consumidor a analisar e compreender o consumo de energia elétrica por equipamento; obter estratégias e sugestões de economia de energia; produzir relatórios que permitam o gerenciamento e favoreçam o uso consciente e sustentável da energia elétrica; prever e detectar equipamentos com falhas; entre muitos outros recursos [Opris and Caracasian 2013b, Basu et al. 2015, Garcia et al. 2017].

Basicamente, existem três abordagens para a desagregação do consumo: a técnica de *monitoramento não-intrusivo* ou concentrado; a técnica de *monitoramento intrusivo* ou distribuído; e a técnica de *monitoramento semi-intrusivo* ou parcialmente distribuído.

A técnica de monitoramento não-intrusivo ou concentrado, do inglês *Non-intrusive Load Monitoring* (NILM), consiste em usar um único medidor alocado no ponto consumidor e, portanto, reduz consideravelmente os custos de implantação do sistema [Azzini et al. 2018]. Nesta abordagem, realizar a desagregação do consumo por equipamento elétrico através da análise de sinais elétricos é um problema ainda em aberto [Rathmair and Haase 2013, Sadeghianpourhamami et al. 2017]. Para isso, pode ser observada a assinatura de potência dos equipamentos elétricos e os indicadores matemáticos ou estatísticos que os caracterizem de forma adequada [Abubakar et al. 2017].

A desagregação do consumo por cargas pode ser vista como um problema tradicional de classificação multiclasse. Assim, a primeira etapa no *pipeline* deste processo de classificação é a *caracterização das amostras*. Neste contexto, a extração de características provenientes dos equipamentos elétricos é o principal desafio nos sistemas NILM, pois tal extração é a chave para a correta identificação dos equipamentos. Diversos trabalhos vêm sendo publicados na literatura especializada, apresentando diferentes estratégias para desagregação de consumo [Figueiredo et al. 2011, Dong et al. 2012, Fernandes et al. 2013, Xiao and Cheng 2019], sendo que a maioria foca em atributos extraídos de comportamentos elétricos coletados em baixa e alta frequência (em relação à taxa de amostragem de sinais). Trabalhos recentes consideram o uso de algoritmos de aprendizagem de máquina [Figueiredo et al. 2011, Kelly and Knottenbelt 2015, Gillis et al. 2016, Zhang et al. 2019] e sugerem o emprego de atributos provenien-

tes do tempo e de parâmetros elétricos [Figueiredo et al. 2011, Makonin et al. 2013, Kelly and Knottenbelt 2015, Sadeghianpourhamami et al. 2017, Souza et al. 2018]. Esses trabalhos apresentam métodos com capacidade preditiva, em termos de acurácia, que oscilam entre 70% a 98%. Tais trabalhos empregam, geralmente, bases de dados compostas por poucas amostras e poucos atributos, variando assim a acurácia reportada.

Com o intuito de aumentar a capacidade dos modelos preditivos para a efetiva caracterização de cargas e desagregação do consumo de energia elétrica em técnicas NILM, este artigo oferece e analisa uma proposta detalhada de caracterização de dados. A hipótese desta pesquisa é que a caracterização apropriada dos dados é capaz de oferecer informações relevantes para os algoritmos tradicionais de aprendizado de máquina e, assim, obter alta capacidade preditiva aliada a um baixo custo computacional, de forma que a caracterização de cargas possa ser empregada na prática e em cenários reais.

Este artigo está organizado da seguinte maneira: a Seção 2 apresenta a revisão dos estudos existentes na literatura, a proposta de caracterização dos dados e a análise dos atributos; a Seção 3 apresenta a coleta de dados reais para avaliação da proposta; a análise de desempenho dos algoritmos de aprendizado de máquina para a validação da proposta de caracterização de dados é oferecida na Seção 4; por fim, a Seção 5 apresenta conclusões e direcionamentos para trabalhos futuros.

2. Caracterização dos dados

A caracterização de dados é a primeira e uma das mais importantes etapas em um processo de aprendizado de máquina. A representação das amostras e a qualidade dos dados são essenciais para se obter uma boa solução para o problema, independentemente do método de aprendizado selecionado [Blum and Langley 1997].

Com a intenção de propor atributos relevantes para representar os dados das cargas elétricas de cada equipamento, primeiramente foi realizado um estudo minucioso da literatura existente. Em resumo, os principais atributos empregados são apresentados na Tabela 1. Destes, tem-se atributos que podem ser extraídos em baixa ou alta frequência de amostragem, conforme descritos a seguir.

2.1. Atributos de baixa frequência

A amostragem de sinal em baixa frequência permite que sejam computados valores que não necessitam de rigorosos detalhes das formas de onda de tensão e corrente, tampouco de hardware com poder de aquisição em alta frequência. Dentro do contexto de desagregação do consumo por cargas, esta técnica é chamada de análise macroscópica [Souza et al. 2018]. Outra característica dos atributos de baixa frequência é a simplicidade de cálculo, não havendo necessidade de processadores com muito poder computacional. Portanto, os atributos como a potência ativa, potência reativa, potência aparente, tensão e corrente eficazes são exemplos dos atributos típicos que podem ser computados em sistemas de baixa frequência. Neste contexto, os principais atributos presentes em trabalhos relevantes da literatura são:

- **Potência aparente (A):** equivalente ao produto da tensão e corrente instantânea. Corresponde à combinação das potências que fluem entre a carga e a fonte;
- **Potência ativa (P):** parcela de potência que é dissipada e responsável pelo trabalho útil. Em outras palavras, é a potência útil que flui entre a fonte e a carga;
- **Potência reativa (Q):** parcela de potência que circula entre a carga e a fonte e que não gera trabalho útil. Geralmente, aparece quando existem componentes indutivos ou capacitivos no circuito;

Tabela 1. Principais atributos empregados em trabalhos relevantes que usam técnicas NILM.

Atributos	Descrição	Referências
A	Potência aparente	[Makonin et al. 2013] [Figueiredo et al. 2011]
P	Potência ativa	[Hart 1992] [Drenker and Kader 1999] [Xiao and Cheng 2019] [Marceau and Zmeureanu 2000] [Baranski and Voss 2004]
Q	Potência reativa	[Marceau and Zmeureanu 2000] [Baranski and Voss 2004]
V	Potência residual	[Souza et al. 2018]
U_{RMS}	Valor eficaz (RMS) de tensão	[Ruzzelli et al. 2010]
I_{RMS}	Valor eficaz (RMS) de corrente	[Ruzzelli et al. 2010] [Kelly and Knottenbelt 2015]
$PF(\lambda)$	Fator de potência	[Souza et al. 2018] [Teshome et al. 2016] [Figueiredo et al. 2011]
$QF(\lambda_Q)$	Fator de reatividade	[Souza et al. 2018]
$VF(\lambda_D)$	Fator de distorção	[Souza et al. 2018]
$U_{1,3,5,\dots,19}$	Componentes harmônicas de tensão	[Sultanem 1991]
$I_{1,3,5,\dots,19}$	Componentes harmônicas de corrente	[Sultanem 1991] [Shaw et al. 2008] [Bouhouras et al. 2019]
DHT_U	Distorção harmônica total de tensão	[Lin et al. 2016]
DHT_I	Distorção harmônica de corrente	[Dong et al. 2012] [Chang 2012] [Lin et al. 2016]
I_f	Corrente inativa	[Huang et al. 2015]

- **Tensão eficaz e corrente eficaz** (U_{RMS} e I_{RMS}): valor quadrático médio de um sinal – tensão ou corrente – que equivale ao valor real absorvido pela carga;
- **Fator de potência** (λ): razão entre a potência ativa com as demais potências que indica a eficiência da energia no circuito;
- **Fator de reatividade** (λ_Q): possui relação direta com a potência reativa e corresponde proporcionalmente ao desfasamento entre tensão e corrente.

2.2. Atributos de alta frequência

A amostragem de sinal em alta frequência permite que sejam computados valores provenientes do detalhamento das formas de onda de tensão e corrente. Dentro da área de desagregação do consumo de cargas, os atributos de alta frequência também são chamados de características microscópicas [Souza et al. 2018, Schirmer and Mporas 2019]. Neste contexto, os principais atributos presentes em trabalhos relevantes da literatura são:

- **Potência residual** (V): parcela de potência referente à característica de não-linearidade na carga;
- **Fator de distorção ou assimetria** (λ_D): razão entre a potência residual com as outras parcelas. Representa a presença de distorção harmônica em função da carga;
- **Componentes harmônicas** ($U_{1,3,\dots,19}$ e $I_{1,3,\dots,19}$): pode-se descrever um sinal periódico (no caso tensão e corrente) como um somatório de sinais senoidais

múltiplos de frequência da rede elétrica através da extração de componentes harmônicas pela transformada rápida de Fourier (FFT, do inglês *Fast Fourier Transform*). O mais comum é extrair os valores de amplitude das componentes ímpares (da 1ª até a 19ª) da tensão (U) e da corrente (I);

- **Distorção harmônica total** (DHT_U e DHT_I): corresponde à quantificação de harmônicas de um sinal, como da tensão (U) e corrente (I);
- **Corrente inativa** (I_f): corresponde ao valor eficaz da quantificação de corrente que não tem função útil de um circuito elétrico.

3. Base de dados

Os atributos apresentados na Seção 2 é a base para a caracterização dos dados proposta neste trabalho. Para avaliá-la, foi realizada a coleta e rotulação de dados reais empregando técnicas tradicionais de aprendizado de máquina. A coleta foi feita através de medições de tensão e corrente de uma instalação real, considerando circuitos monofásicos com tensão nominal senoidal de 127 Volts a 60 Hz. A corrente foi adquirida através de um sensor SCT-03 com taxa de amostragem de sinal equivalente a 15,36 kHz (256 sinais coletados em cada ciclo de 60 Hz). Os sinais dos sensores foram posteriormente condicionados e enviados para o dispositivo embarcado, no caso um *BeagleBone Black* (BBB) via Conversor Analógico para Digital (ADC).

A Figura 1 apresenta o diagrama do hardware construído para a coleta dos dados. Neste hardware, a fonte CA/CC realiza a alimentação da placa de condicionamento de sinais e do BBB. Ela oferece saídas de tensão contínua de +5 V, +/-15 V, sendo as três tensões para a placa de condicionamento e +5 V para o BBB. Na placa de condicionamento, os sensores recebem os sinais de tensão e corrente e é realizado o condicionamento do sinal para a saída compatível com o ADC do BBB.

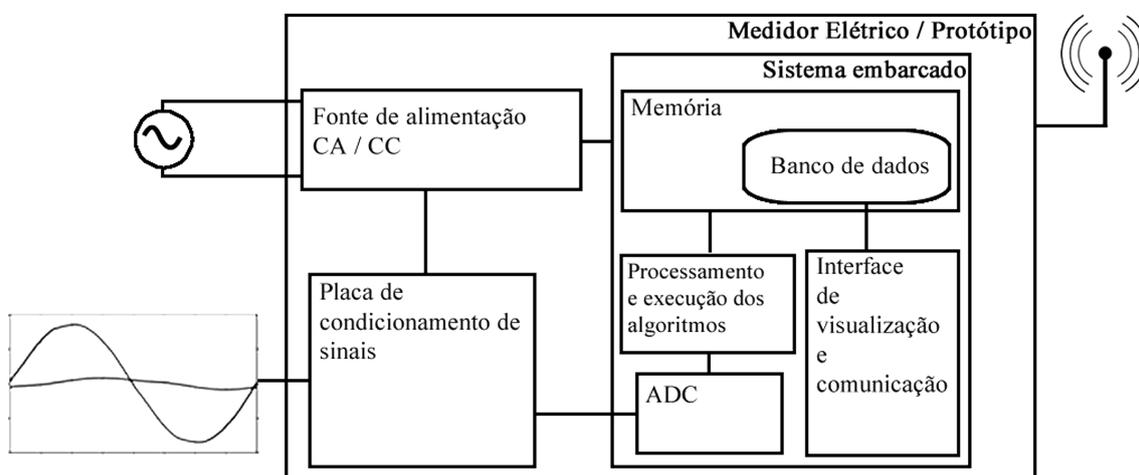


Figura 1. Diagrama do protótipo do medidor de energia.

O BBB é responsável pela aquisição e conversão analógica para digital do sinal condicionado, que é feito com resolução de 12 bits. Em seguida, o BBB processa os cálculos das grandezas elétricas e computa os atributos. O BBB armazena os dados na memória do dispositivo, entretanto há possibilidade de enviar os dados para a nuvem.

Foram coletados 460 minutos de oscilografias de tensão e corrente para 28 equipamentos elétricos (classes do problema), apresentados na Tabela 2 (a numeração das classes foi baseada nos padrões adotados por [Souza et al. 2018]). A coleta das formas

de onda de tensão e corrente foi feita durante a estabilização de potência da carga ligada, conforme exemplificado na Figura 2.

Tabela 2. Equipamentos com oscilografias coletadas para compor a base de dados.

Classe	Equipamento	Classe	Equipamento
1	Lâmpada incandescente 60W	17	Lâmpada mista
2	Ar condicionado	20	Secadora de roupas regulável
3	Refrigerador	21	Refrigerador regulável
4	Micro-ondas	22	Liquidificador
5	TV CRT	23	Panificadora elétrica
6	TV LCD	24	Desktop PC
8	Chuveiro	26	Processador de alimentos
9	Carga RL	27	Freezer
10	Carga não linear 1	28	Forno elétrico
11	Carga não linear 2	29	Portão de garagem
13	Ferro de passar	30	Laptop (notebook)
14	Máquina de lavar roupas	32	Secadora de roupas
15	Secador de cabelos	33	Refrigerador regular
16	Lâmpada fluorescente	34	Aspirador de pó

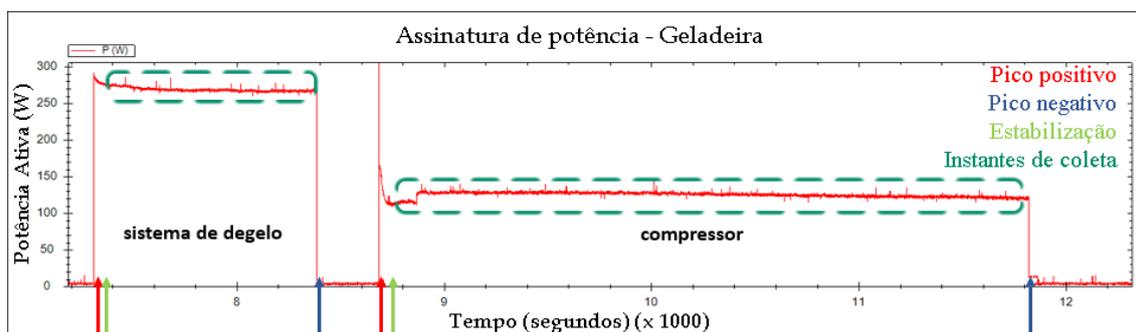


Figura 2. Exemplo do instante da coleta de dados.

Com base no levantamento dos trabalhos mais pertinentes disponíveis na literatura, apresentados na Seção 2, foram inicialmente selecionados os 32 atributos apresentados na Tabela 1. Para analisar a qualidade e relevância desses atributos, foi empregado o método de florestas aleatórias (com 10 árvores de decisão) nos dados coletados e foram analisadas as regras de decisões criadas pelo algoritmo. A Tabela 3 apresenta a soma do número de ocorrências para cada atributo em todas as árvores de decisão; dos 32 atributos, somente 20 apareceram em alguma regra.

Observando a Tabela 3, conclui-se que nenhum atributo proveniente exclusivamente do sinal de tensão foi relevante para a caracterização dos dados. Portanto, foram removidos os atributos com os valores da distorção harmônica de tensão (DHT_U), do valor eficaz de tensão (U_{RMS}) e das componentes harmônicas de tensão (U_1, U_3, \dots, U_{19}), restando os 20 atributos apresentados na Tabela 4.

Como os vinte atributos selecionados são calculados a cada ciclo de 60 Hz (1/60 segundos), foram computados sete valores obtidos através das seguintes medidas estatísticas de localidade e dispersão, calculadas por minuto (3.600 leituras): média (avg),

Tabela 3. Ocorrência dos atributos nas árvores de decisão.

Atributo	Ocorrência	Atributo	Ocorrência	Atributo	Ocorrência
I_1	29	I_{RMS}	18	I_9	6
QF	28	DHT_I	18	I_{11}	6
Q	27	V	15	I_{15}	4
I_3	22	P	14	I_{13}	2
PF	22	I_5	14	I_{17}	2
I_f	21	VF	13	I_{19}	1
A	19	I_7	12		

Tabela 4. Atributos selecionados para compor a representação dos dados.

Atributo	Descrição
A	Potência aparente
P	Potência ativa
Q	Potência reativa
V	Potência residual
$PF(\lambda)$	Fator de potência
$QF(\lambda_Q)$	Fator de reatividade
$VF(\lambda_Q)$	Fator de assimetria
I_{RMS}	Valor eficaz da corrente
DHT_I	Distorção harmônica total da corrente
I_f	Corrente inativa
I_1 a I_{19}	Amplitude da harmônica de ordem n ($n = 1, 3, \dots, 19$) da corrente

desvio padrão (*std*), valor mínimo (*min*), valor máximo (*max*), primeiro quartil (q_1), segundo quartil (q_2) e terceiro quartil (q_3). Desta forma, a proposta final de caracterização das amostras para representar o padrão de consumo elétrico por equipamento (por minuto) é composta por 140 atributos, com o processo de coleta e representação das amostras sumarizado na Figura 3. Com 460 minutos de oscilografia para cada um dos 28 equipamentos elétricos, foram coletadas e rotuladas 12.880 amostras para compor a base de dados. Como contribuição adicional desta pesquisa, a base de dados foi publicamente disponibilizada em <https://github.com/wesleyangelino/nilmbr.git>.

4. Resultados e discussões

Após a caracterização, coleta e rotulação dos dados, as amostras duplicadas foram removidas e os valores dos atributos foram normalizados por padronização (média igual a zero e desvio padrão igual a um). Posteriormente, foram empregados métodos tradicionais de classificação para analisar se a representação proposta é suficientemente informativa para prover boa capacidade de predição aos modelos. Para isso, foi utilizada a técnica de validação cruzada através da tradicional validação k -fold estratificada, com $k = 5$.

Os seguintes métodos de classificação foram empregados: k -vizinhos mais próximos (kNN), máquinas de vetores de suporte (SVM) com *kernel* RBF, SVM com *kernel* linear e florestas aleatórias (RF). Como o desempenho desses métodos é sensível à parametrização, foi realizada uma busca em grade para escolher os melhores parâmetros em cada rodada da validação cruzada. Neste caso, para o kNN, foram analisados os valores de $k = \{1, 3, 5\}$; para a RF, o número de árvores de decisão $T = \{10, 20, \dots, 90, 100\}$; e para o SVM, foram utilizados os *kernels* Linear e RBF com $C = \{0, 001, 0, 01, \dots, 100, 1.000\}$ e $\gamma = \{0, 001, 0, 01, \dots, 100, 1.000\}$.

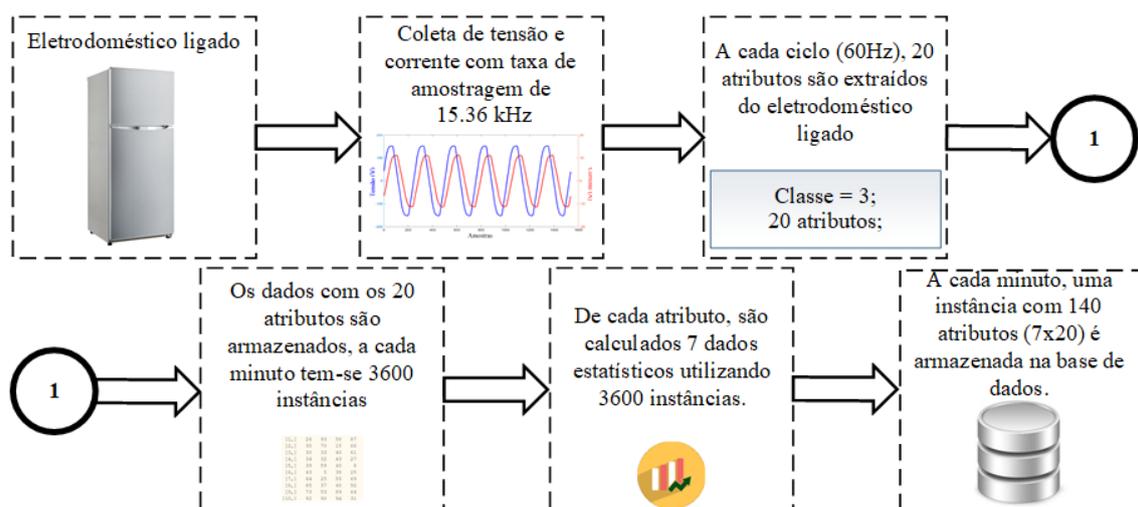


Figura 3. Diagrama da coleta de uma amostra.

Todos os métodos foram implementados em Python 3.7, utilizando funções da biblioteca `scikit-learn` [Pedregosa et al. 2011]. Os experimentos foram executados em um computador com processador i7-8750H e memória RAM de 16GB. Para comparar os desempenhos, foram utilizadas as métricas tradicionais: macro revocação, macro precisão e macro F-medida. Os resultados obtidos estão apresentados na Tabela 5, sendo que as duas últimas linhas (t_{train} e t_{test}) correspondem aos tempos médios de treinamento e teste, respectivamente.

Tabela 5. Resultados obtidos na tarefa de categorização de cargas usando a representação proposta neste trabalho.

Classificador	kNN	RF	SVM (RBF)	SVM (Linear)
F-medida	1,000	1,000	1,000	0,999
Precisão	1,000	1,000	1,000	0,999
Revocação	1,000	1,000	1,000	0,999
$t_{train}(s)$	-	0,405	0,795	0,260
$t_{test}(s)$	1,005	0,063	0,675	0,652

De acordo com os resultados obtidos, é evidente que a caracterização dos dados proposta neste trabalho foi muito efetiva na representação das amostras, pois os três métodos de classificação não lineares empregados obtiveram 100% de acerto nos experimentos conduzidos, enquanto o SVM Linear obteve 99,9%. Portanto, pode-se afirmar que com a representação proposta, os dados são linearmente separáveis. No contexto de identificação de cargas, todos os 28 equipamentos elétricos foram reconhecidos corretamente com base nos seus padrões de cargas elétricas nas 12.880 amostras coletadas. Esse desempenho surpreendente levantou a suspeita de que os métodos estavam sobreajustando os modelos aos dados. Para certificar que isso não ocorreu, foram geradas as curvas de aprendizado apresentadas na Figura 4. Para isso, 20% dos dados foram separados para teste e o restante foi usado para o treinamento incremental de cada método, sendo este incremento de 1% em 1% dos dados. Conforme pode ser visto, a acurácia aumenta consideravelmente até o valor unitário quando o treinamento ocorre com cerca de 20% dos dados (aproximadamente 2.200 amostras). Portanto, conclui-se que os modelos não estão sobreajustados e que os dados possuem um padrão que pode ser descrito com uma base de regras construída com cerca de 20% das amostras.

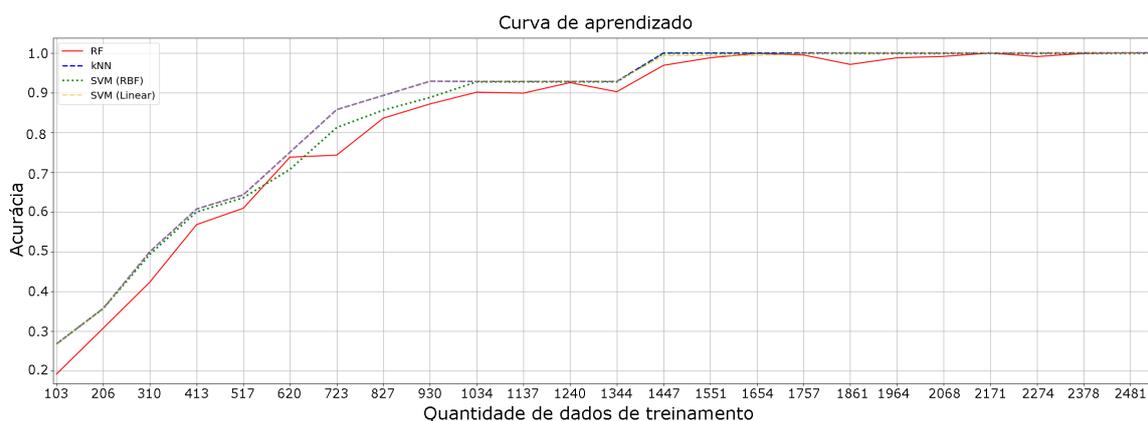


Figura 4. Curvas de aprendizado dos métodos de classificação.

Ao analisar cuidadosamente as amostras por equipamento, é possível afirmar que a predição perfeita é um comportamento esperado neste tipo problema, pois grande parte das cargas elétricas apresenta pequena variação de potência durante o tempo de funcionamento e sobretudo porque foram coletados os dados de todos os estágios de operação dos 28 equipamentos elétricos. Portanto, com base no desempenho apresentado, conclui-se que é possível realizar a efetiva desagregação de cargas em cenários reais e em tempo quase real (menos de 0,000025 segundo de resposta por amostra considerando a RF). Esses resultados abrem um grande leque de possibilidades para explorações de novos recursos e empregos de técnicas NILM em medidores cognitivos.

Comparando esses resultados com os reportados na literatura [Kelly and Knottenbelt 2015, Souza et al. 2018] que, tradicionalmente oferecem acurácia entre 70% e 98% no melhor caso, este trabalho apresentou uma proposta de representação de dados que proporcionou um aumento considerável nos indicadores de desempenho. Outro destaque corresponde ao algoritmo de aprendizado de máquina geralmente recomendado: o kNN. Este trabalho, com a proposta da nova caracterização de dados, evidenciou que na verdade os métodos de classificação tradicionais obtêm desempenho similar, sendo que a RF apresentou melhores resultados em termos de tempo computacional, enquanto que o kNN apresentou o maior esforço em tempo de teste. Desta forma, é conclusivo que a RF pode ser recomendada como algoritmo de classificação padrão para a desagregação de consumo com base nas cargas de equipamentos elétricos.

5. Conclusão

Este trabalho apresentou uma nova abordagem de caracterização de atributos com o objetivo de identificar equipamentos elétricos em um sistema de desagregação de consumo de energia elétrica. Com base na revisão de diversos trabalhos da literatura e na análise dos dados, foram selecionados 20 atributos relevantes, dos quais foram derivadas 7 medidas estatísticas, criando assim uma representação de dados composta por 140 atributos.

Para validar a caracterização de dados proposta neste artigo, foi construída uma base de dados composta por 12.880 amostras rotuladas em 28 possíveis classes (equipamentos elétricos), totalizando 460 instâncias por equipamento. Cada instância corresponde a 140 atributos estatísticos computados a partir de leituras de tensão e corrente em um minuto de funcionamento de cada equipamento elétrico.

Foram empregados três algoritmos tradicionais de classificação: *k*-vizinhos mais próximos, florestas aleatórias e máquinas de vetores de suporte. Todos obtiveram de-

sempenho surpreendente e foram capazes de prever corretamente o rótulo de todas as amostras das partições de teste em um processo de validação cruzada com 5 partições.

Com base nos resultados obtidos, ficou evidente que a caracterização dos dados proposta ofereceu excelente representação dos dados, que refletiu na significativa melhora no processo de desagregação de consumo e classificação de equipamentos com base nas suas características elétricas. Em termos de esforço computacional, as florestas aleatórias apresentaram o melhor desempenho, com tempo de resposta menor que 0,000025 segundo por amostra.

Esses resultados abrem um grande leque de possibilidades para explorações de novos recursos e empregos de técnicas NILM em medidores cognitivos, evidenciando que os métodos de classificação tradicionais podem ser seguramente empregados em cenários reais, mesmo que demandem baixo tempo de resposta. Alternativamente, como a predição foi perfeita para todos os métodos avaliados, pode ser considerada a criação de uma base de regras a partir da análise de uma amostragem representativa dos dados.

Com esse recurso de classificação embarcado, será possível criar um novo conceito de medidores de energia elétrica, permitindo auxiliar o consumidor a analisar e compreender o consumo de energia elétrica por equipamento; obter estratégias e sugestões de economia de energia; produzir relatórios que permitam o gerenciamento e favoreçam o uso consciente e sustentável da energia elétrica; prever e detectar equipamentos com falhas; entre muitas outras aplicações.

Para trabalhos futuros, é sugerido aumentar ainda mais os dados de validação, tanto em termos de quantidade de leituras quanto na variabilidade de equipamentos elétricos. Como etapa seguinte, sugere-se que o método de predição seja embarcado no medidor e avaliado em um cenário real controlado.

Agradecimentos

Os autores são gratos à CAPES (Processo 88887.334579/2019-0), CNPq e à FAPESP pelo apoio financeiro e bolsas que, direta ou indiretamente, fomentaram este projeto de pesquisa.

Referências

- Abubakar, I., Khalid, S., Mustafa, M., Shareef, H., and Mustapha, M. (2017). Application of load monitoring in appliances' energy management - a review. *Renewable and Sustainable Energy Reviews*, 67:235–245.
- Azzini, H. A. D., Cypriano, J. G. I., de Souza, W. A., Monzani, R. C., da Silva, L. C. P., and Venerando, A. C. (2018). Comparação das técnicas de monitoramento de cargas para a desagregação do consumo de energia elétrica. In *Congresso Brasileiro de Automática*, pages 2525–8311.
- Barai, G. R., Krishnan, S., and Venkatesh, B. (2015). Smart metering and functionalities of smart meters in smart grid - a review. In *IEEE Electrical Power and Energy Conference*, pages 138–145.
- Baranski, M. and Voss, J. (2004). Genetic algorithm for pattern detection in nialm systems. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 4, pages 3462–3468 vol.4.
- Basu, K., Debusschere, V., Bacha, S., Maulik, U., and Bondyopadhyay, S. (2015). Nonintrusive load monitoring: A temporal multilabel classification approach. *IEEE Transactions on Industrial Informatics*, 11(1):262–270.

- Blum, A. L. and Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1):245 – 271. Relevance.
- Bouhouras, A. S., Gkaidatzis, P. A., Panagiotou, E., Poulakis, N., and Christoforidis, G. C. (2019). A nilm algorithm with enhanced disaggregation scheme under harmonic current vectors. *Energy and Buildings*, 183:392 – 407.
- Chang, H.-H. (2012). Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses. *Energies*, 5(11):4569–4589.
- Dong, M., Meira, P. C. M., Xu, W., and Freitas, W. (2012). An event window based load monitoring technique for smart meters. *IEEE Transactions on Smart Grid*, 3(2):787–796.
- Drenker, S. and Kader, A. (1999). Nonintrusive monitoring of electric loads. *IEEE Computer Applications in Power*, 12(4):47–51.
- Fernandes, R. A. S., da Silva, I. N., and Oleskovicz, M. (2013). Load profile identification interface for consumer online monitoring purposes in smart grids. *IEEE Transactions on Industrial Informatics*, 9(3):1507–1517.
- Figueiredo, M. B., de Almeida, A., and Ribeiro, B. (2011). An experimental study on electrical signature identification of non-intrusive load monitoring (nilm) systems. In *Adaptive and Natural Computing Algorithms*, pages 31–40, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Garcia, F. D., Marafao, F. P., d. Souza, W. A., and d. Silva, L. C. P. (2017). Power metering: History and future trends. In *IEEE Green Technologies Conference*, pages 26–33.
- Gillis, J. M., Alshareef, S. M., and Morsi, W. G. (2016). Nonintrusive load monitoring using wavelet design and machine learning. *IEEE Transactions on Smart Grid*, 7(1):320–328.
- Hart, G. W. (1992). Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891.
- Huang, T. D., Wang, W., and Lian, K. (2015). A new power signature for nonintrusive appliance load monitoring. *IEEE Transactions on Smart Grid*, 6(4):1994–1995.
- Kelly, J. and Knottenbelt, W. (2015). Neural nilm: Deep neural networks applied to energy disaggregation. In *ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 55–64.
- Lin, S., Zhao, L., Li, F., Liu, Q., Li, D., and Fu, Y. (2016). A nonintrusive load identification method for residential applications based on quadratic programming. *Electric Power Systems Research*, 133:241 – 248.
- Makonin, S., Popowich, F., and Gill, B. (2013). The cognitive power meter: Looking beyond the smart meter. In *IEEE Canadian Conference on Electrical and Computer Engineering*, pages 1–5.
- Marceau, M. and Zmeureanu, R. (2000). Nonintrusive load disaggregation computer program to estimate the energy consumption of major end uses in residential buildings. *Energy Conversion and Management*, 41(13):1389 – 1403.
- Opris, I. and Caracasian, L. (2013a). On the implementation of the functionalities of smart metering systems. In *International Symposium on Advanced Topics in Electrical Engineering*, pages 1–6.

- Opris, I. and Caracasian, L. (2013b). The relation between smart meters and electricity consumers. In *International Conference on Environment and Electrical Engineering*, pages 325–329.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Platt, G., West, S., and Moore, T. (2015). The real-world challenges and opportunities of distributed generation. In *IEEE Energy Conversion Congress and Exposition*, pages 1112–1116.
- Rathmair, M. and Haase, J. (2013). Load identification and management framework for private households. In *Annual Conference of the IEEE Industrial Electronics Society*, pages 5729–5734.
- Ruzzelli, A. G., Nicolas, C., Schoofs, A., and O’Hare, G. M. P. (2010). Real-time recognition and profiling of appliances through a single electricity sensor. In *IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, pages 1–9.
- Sadeghianpourhamami, N., Ruysinck, J., Deschrijver, D., Dhaene, T., and Develder, C. (2017). Comprehensive feature selection for appliance classification in nilm. *Energy and Buildings*, 151:98–106.
- Schirmer, P. A. and Mporas, I. (2019). Statistical and electrical features evaluation for electrical appliances energy disaggregation. *Sustainability*, 11(11).
- Shaw, S. R., Leeb, S. B., Norford, L. K., and Cox, R. W. (2008). Nonintrusive load monitoring and diagnostics in power systems. *IEEE Transactions on Instrumentation and Measurement*, 57(7):1445–1454.
- Souza, W. A., Marafao, F. P., Liberado, E. V., Diniz, I. S., and Serni, P. J. A. (2015). Power quality, smart meters and additional information from different power terms. *IEEE Latin America Transactions*, 13(1):158–165.
- Souza, W. A., Marafão, F. P., Liberado, E. V., Simões, M. G., and Silva, L. C. P. (2018). A nilm dataset for cognitive meters based on conservative power theory and pattern recognition techniques. *Journal of Control, Automation and Electrical Systems*, 29(6):742–755.
- Sultanem, F. (1991). Using appliance signatures for monitoring residential loads at meter panel level. *IEEE Transactions on Power Delivery*, 6(4):1380–1385.
- Teshome, D. F., Huang, T. D., and Lian, K. (2016). Distinctive load feature extraction based on fryze’s time-domain power theory. *IEEE Power and Energy Technology Systems Journal*, 3(2):60–70.
- Xiao, P. and Cheng, S. (2019). Neural network for NILM based on operational state change classification. *arXiv*.
- Zhang, X., Shen, J., Yang, T., Tang, L., Wang, L., Liu, Y., and Xu, P. (2019). Smart meter and in-home display for energy savings in residential buildings: a pilot investigation in shanghai, china. *Intelligent Buildings International*, 11(1):4–26.