

Boa Energia: Base de Dados Pública de Consumo Residencial com Qualidade de Dados

Victor Hayashi, Haroldo Amaral, Tiago Fujii, André de Souza e Reginaldo Arakaki

Resumo— A transformação do setor energético traz desafios e oportunidades de inovação e pesquisa. As peculiaridades do cenário brasileiro demandam por estudos específicos para a consolidação de tendências como precificação dinâmica de energia. Contudo, a inexistência de bases de dados de consumo é um grande desafio para a pesquisa no setor. Para garantir a coleta de dados com qualidade, uma arquitetura tolerante a falhas é proposta para lidar com indisponibilidades momentâneas de conexão com a Internet. Resultados iniciais da coleta de dados de um mês em uma residência em conjunto com a implementação de algoritmos para predição de consumo são apresentados.

Palavras-Chave— Qualidade de Dados, Internet das Coisas, Tolerância a Falhas, Consumo de Energia Residencial.

Abstract— The transformation of the energy sector brings challenges and opportunities for innovation and research. The peculiarities of Brazilian scenario demand specific studies to consolidate trends such as dynamic energy pricing. However, the lack of consumption databases is a major challenge for research in the sector. In order to guarantee data collection with quality, a fault-tolerant architecture is proposed to deal with momentary Internet unavailabilities. Initial results of data collection of one month in a residence together with implementation of consumption prediction algorithms are presented.

Keywords— Data Quality, Internet of Things, Fault-tolerant systems, Residential Energy Consumption.

I. INTRODUÇÃO

O setor energético brasileiro está em processo de transformação. Mudanças de cunho tecnológico e regulatório estão trazendo novos desafios e oportunidades de inovação tecnológica e pesquisa. Ao passo que tendências como geração distribuída, precificação dinâmica, mercado livre de energia e mobilidade elétrica se consolidam, é fato o impacto em escala vindouro nas indústrias e vidas de milhões de pessoas.

Devido às suas peculiaridades, o cenário brasileiro demanda estudos específicos. Considere o desafio da distribuição de energia devido à extensão do território nacional, aspectos culturais e sociais, o fato da matriz energética ser majoritariamente renovável (hidroelétricas), o histórico de formação de preços no setor e a existência de um calendário de mudanças grandes no setor numa janela de tempo relativamente pequena. A implementação de novos modelos de tarifação, como a tarifa branca [1] e a precificação dinâmica em granularidade horária [2] impactarão diretamente os padrões de consumo do setor residencial, responsável por cerca de 28% do consumo total de energia em 2018 [3].

Pesquisas e estudos de consumo residencial enfrentam diversos desafios. Há carência de bancos de dados públicos referente ao consumo de energia residencial no Brasil, uma vez que a grande maioria de bancos de dados públicos refere-se a residências norte-americanas ou europeias [4, 5]. Mesmo os bancos de dados disponíveis, conforme a Tabela I, possuem problemas relacionados à qualidade dos dados. Aspectos de exatidão e periodicidade são particularmente importantes: a precisão está relacionada ao quanto os dados coletados pelos sensores estão próximos da realidade, enquanto a completude está relacionada à ausência de lacunas temporais na sequência de dados coletados [6].

Os problemas que causam a baixa qualidade dos dados de consumo possuem diversas origens. Em um projeto piloto realizado entre 2011 e 2014 na Alemanha, problemas reportados nas seis residências participantes variam desde a desconexão acidental do cabo de rede dos medidores inteligentes, até a instabilidade da conexão com a internet [6]. Outros fatores encontrados na literatura [7] são falhas do equipamento elétrico de medição e da rede de comunicação. A baixa qualidade de dados dificulta o desenvolvimento de algoritmos com base nestes dados. Por exemplo, a precisão de algoritmos de predição é diminuída [4, 6]. A rede de energia pode ser impactada se os algoritmos de predição forem empregados na diminuição do risco de sua operação (ao aumentar a previsibilidade da demanda) e na tomada de decisão de consumidores residenciais (com o intuito educativo de diminuir picos de consumo).

TABELA I. BASES DE DADOS DE CONSUMO RESIDENCIAL

Dataset	Localização	# Residências
UK-DALE [8]	Reino Unido	5
Smart* [9]	América do Norte	400+
REDD [10]	EUA	6

Dado o exposto, fica fundamentada a oportunidade de construção de uma base de dados com consumo residencial no Brasil. Uma proposta de solução de coleta de dados que permita garantir a integridade dos dados coletados será útil para fomentar o desenvolvimento de algoritmos.

Desta forma, o objetivo deste artigo é a proposição de uma arquitetura de coleta de dados que permita a construção de bases de dados de consumo energético residencial no Brasil, com o diferencial de garantir sua integridade. Como prova de conceito, a coleta de dados em granularidade horária de uma residência é realizada, e os dados são disponibilizados de forma pública. O desenvolvimento de algoritmo de predição com redes neurais para maior previsibilidade da curva horária para o consumidor residencial é realizado como um caso de uso possível da base de dados.

Victor T. Hayashi, Tiago Y. Fujii e Reginaldo Arakaki, PCS – Departamento de Engenharia de Computação e Sistemas Digitais, Universidade de São Paulo - USP, São Paulo-SP, e-mails: {victor.hayashi, tiago.fujii, reg}@usp.br; Haroldo Amaral e André de Souza, LSISPOTI - Laboratório de Sistemas de Potência e Técnicas Inteligentes, Universidade Estadual Paulista - UNESP, Bauru-SP, e-mails: Haroldo.amaral@alumni.usp.br, andre.souza@unesp.br.

II. SOLUÇÃO PROPOSTA

Os eventos que podem levar a uma falha de lacuna na sequência de dados foram organizados em uma árvore de falhas, conforme ilustrado na Figura 1. Os eventos mais críticos listados na literatura estão no primeiro nível da árvore de falhas (i.e. uma queda de energia, falha no medidor ou falha de comunicação). Observe que, por estarem ligados a uma porta OU, qualquer um dos três eventos causa a falha em análise.

A solução proposta busca tratar da falha de lacuna temporal na sequência de dados coletados, em especial a falha de comunicação. Na Figura 1, pode-se observar que esta falha é causada por indisponibilidade da rede local, da conexão com a internet ou de indisponibilidade do servidor de computação em nuvem, responsável pelo armazenamento dos dados.

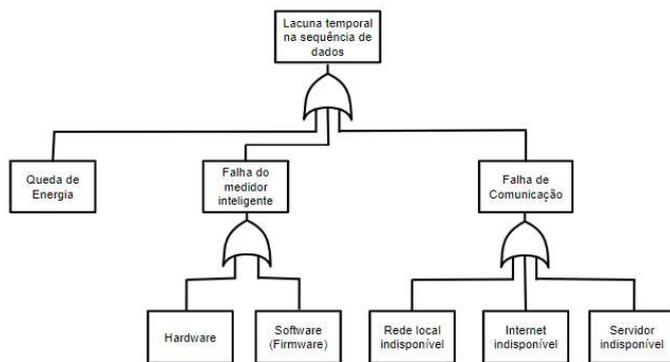


Fig. 1. Árvore de Falhas para lacuna na sequência de dados

Os fatores que levam a uma falha de comunicação estão relacionadas à instabilidade do nível de conectividade em uma arquitetura de Internet das Coisas, e afetam diretamente sua disponibilidade [11]. Uma justificativa inicial pode ser observada na Figura 2, que mostra o número de falhas de conexão com a internet por dia em uma residência, no período de 23 de abril de 2019 a 02 de setembro de 2019. As falhas de indisponibilidade da internet ocorreram durante o processo de coleta de dados de consumo residencial do trabalho [4]. Observa-se que o número de dias em que não há falha de conexão corresponde a uma minoria.

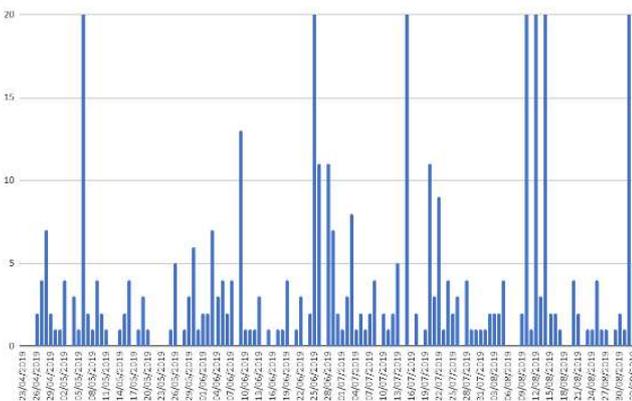


Fig. 2. Número de falhas de conexão na coleta de dados do trabalho [4] (23/04/2019 a 02/09/2019)

A Figura 3 ilustra a validação experimental da instabilidade por nível de serviço através do monitoramento do ping (um computador local instalado na casa foi responsável pelo monitoramento). A conectividade internet diz respeito à conexão da residência com o provedor de internet (falha de internet indisponível na Figura 1), enquanto a conectividade local à conexão da rede local WiFi 802.11, e a conectividade do módulo está relacionada com a conexão do módulo com a rede local da residência (estes dois últimos estão relacionados a falha de rede local indisponível na Figura 1). Já a falha de servidor indisponível diz respeito à instabilidade dos serviços de computação em nuvem, como ilustrado no gráfico inferior da Figura 3. O monitoramento do ping foi realizado em residência na cidade de Santo André, no período de 30/junho/2019 a 13/julho/2019, em intervalos de 2 minutos entre as medidas.

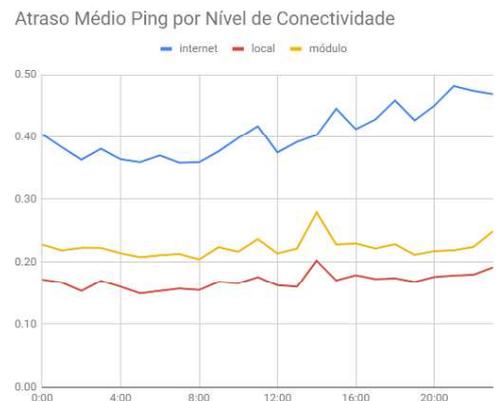


Fig. 3. Atraso médio do ping (30/06/2019 a 13/07/2019)

Uma arquitetura comumente usada para a construção das bases de dados consiste no esquema ilustrado na Figura 4. O medidor inteligente de energia fica conectado a uma rede sem fio através de um hotspot (1), que disponibiliza a rede local sem fio. Este hotspot também está conectado à internet, e é através dessa conexão que os dados de consumo passam para a retaguarda cloud, onde são persistidos (2). Essa operação se repete periodicamente (por exemplo, a cada um segundo). O problema ocorre durante indisponibilidades momentâneas da rede: indisponibilidade da rede local (relacionado à qualidade do dispositivo de rede), servidor na nuvem (garantido por SLA – Service Level Agreement) ou da conexão com a internet (como observado nas Figuras 2 e 3, ocorre certa instabilidade nas duas residências em estudo). Esses períodos de instabilidade possuem como consequências as lacunas temporais na sequência de dados coletados, portanto prejudicando o aspecto de completude da qualidade de dados.

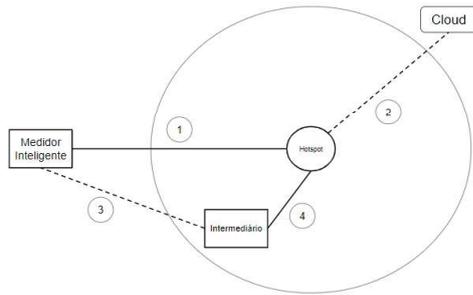


Fig. 4. Arquitetura Cloud para Coleta de Dados

A “Arquitetura Edge” proposta apresenta uma redundância (vide Figura 5). Em caso de indisponibilidade do servidor em nuvem, rede local ou da conexão com a internet, o medidor inteligente conecta-se a um intermediário (3), responsável por armazenar temporariamente os dados de consumo. Quando a infraestrutura de comunicação estiver disponível novamente, este intermediário conecta-se ao *hotspot* (4) e encaminha os dados armazenados para a cloud. Este intermediário pode estar implementado como um módulo de software no próprio medidor [4] ou em um *middleware* local, como um Raspberry Pi 3 (vide módulo “Morpheus” de [12]).

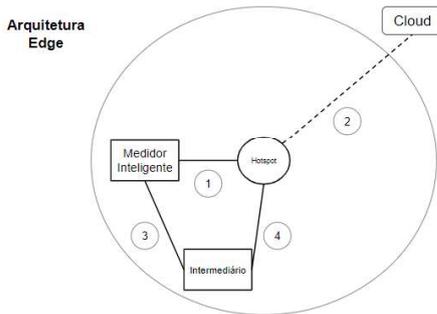


Fig. 5. Arquitetura Edge para Coleta de Dados

Conforme ilustrado na Figura 6, a arquitetura proposta “Data Harvesting” é semelhante ao implementado em redes de sensores sem fio para lidar com a indisponibilidade de conexão onde o sensor está localizado, como a ZebraNet [13]. Nesta arquitetura, o intermediário se conecta diretamente ao medidor inteligente e coleta os dados de consumo de energia (3), vide a ilustração superior da Figura 6. Já quando o intermediário entra no raio de cobertura do *hotspot*, este se conecta à internet e persiste os dados na nuvem (4), conforme a ilustração inferior da Figura 6. O intermediário deve estar implementado em um dispositivo físico móvel, como um *smartphone*.

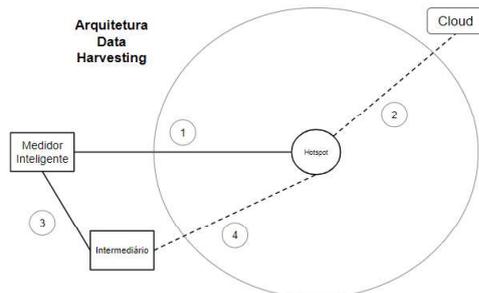


Fig. 6. Arquitetura Data Harvesting para Coleta de Dados

Com a granularidade horária para armazenamento de dados de 8 canais, cada medidor pode ter capacidade de armazenar estes dados de até um mês. Em um caso extremo, isto significa que a arquitetura pode tolerar até um mês de indisponibilidade de conexão. A Tabela II compara o aspecto de tolerância a falhas de comunicação citado, assim como a necessidade de um intermediário físico móvel e um último aspecto relativo à privacidade. Este aspecto está relacionado à possibilidade de implementar regras de autorização para envio dos dados privativos de consumo energético caso o dispositivo móvel físico seja um *smartphone*. Como é o intermediário que controla o envio de dados para a nuvem, configurações da granularidade dos dados enviados podem ser acessadas pelos titulares dos dados através de um aplicativo específico (a granularidade tem efeito sobre a privacidade dos dados de medidores inteligentes, conforme [14]).

TABELA II. COMPARATIVO ENTRE ARQUITETURAS PARA COLETA DE DADOS

Característica	Cloud	Edge	Data Harvesting
Tolerância a falhas de conexão	não	sim	sim
Intermediário físico móvel	não	não	sim

Como estudo de caso do uso da base de dados, foram desenvolvidos modelos multivariados de previsão do consumo energético a partir de seu histórico temporal e da temperatura interna da residência.

O modelo desenvolvido utiliza uma arquitetura *Perceptron Multicamadas* (MLP) e *Long Short Term Memory* (LSTM), possuindo como entradas os valores de temperatura e consumo das últimas 48 horas. Em suas camadas escondidas possui uma camada LSTM e uma camada de *dropout*, de modo a evitar o *overfitting* do modelo. A saída informa a previsão do valor de consumo para as próximas 24 horas.

III. RESULTADOS

Foram coletados dados do consumo energético total e por setores, além da temperatura interna de uma residência por um período de 28 dias. O intervalo entre as medições é de um segundo, e a granularidade dos dados armazenados é horária, de modo a adequar-se a modelos de previsão do consumo energético residencial. A arquitetura utilizada foi a “Arquitetura Edge”, com módulos de ESP8266 e uso de comunicação WiFi 802.11 (o detalhamento de aspectos de hardware foge do escopo deste artigo, vide seção IV).

A Tabela III mostra um comparativo entre o banco de dados desenvolvido neste trabalho com o REDD [3]. Pode-se observar que o Boa Energia possui uma menor taxa de amostragem e monitora os valores de corrente, temperatura interna e consumo energético de até 8 setores, os quais podem ser observados nas Figuras 7 e 8, de modo a auxiliar a implementação de algoritmos de previsão do consumo futuro. Embora o procedimento de sensoriamento ocorra a cada segundo, é realizado o *downsampling* para um período entre medidas de uma hora, de modo a reduzir efeitos de variabilidade do consumo instantâneo, sem afetar seu padrão médio.

A desagregação do consumo em setores pode ser utilizada para aprimorar a previsão do consumo total, além de permitir o

desenvolvimento de sistemas de recomendações de economia de energia personalizados, indicando os setores de maior consumo e mudanças de hábito para fora do horário de pico [15]. O monitoramento da temperatura interna mínima, média e máxima em granularidade horária possibilita uma análise do *tradeoff* entre conforto térmico do usuário e seu consumo energético, além de aprimorar a precisão de modelos de predição do consumo energético com a adição desta variável exógena [16].

TABELA III. COMPARATIVO ENTRE BASES DE DADOS BOA ENERGIA E REDD

	REDD [10]	Boa Energia
Localidade	EUA	Brasil
# sensores	20	8
intervalo médio entre medidas	4.0 s	1 h
# medidas realizadas	16,549,124	835,200
# dados persistidos	16,549,124	696
tamanho	259 MB	61,5 KB
início	18/04/2011	03/02/2020
fim	24/05/2011	02/03/2020
total de dias	36	29
parâmetros	potência aparente (VA)	corrente (A), temperatura (°C), consumo estimado (kVAh)
escopo de desenvolvimento de algoritmos	desagregação de energia	predição

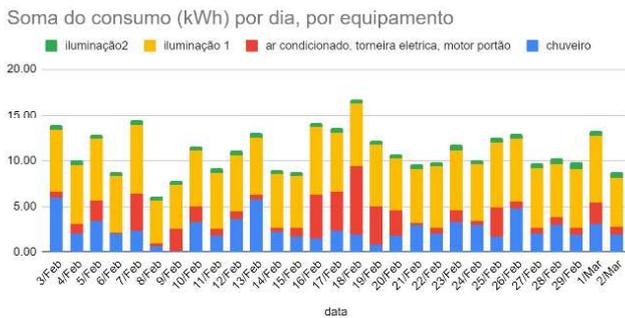


Fig. 7. Consumo por setor (03/02/2020 a 02/03/2020)



Fig. 8. Temperatura interna da residência (03/02/2020 a 02/03/2020)

Para análise do modelo de previsão, estudou-se o consumo do canal do ar condicionado de uma das residências, os quais podem ser observados na Figura 9.

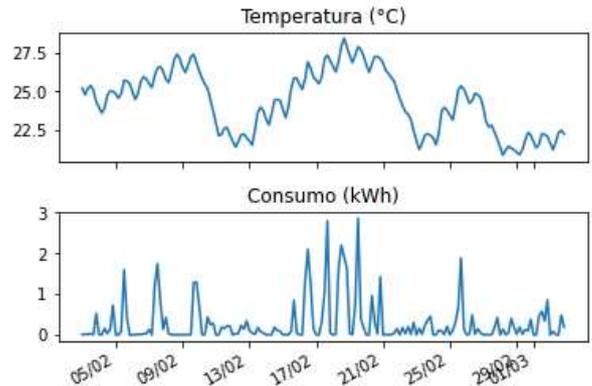


Fig. 9. Gráfico da temperatura interna da residência e consumo horário do ar condicionado

Na Figura 9, é possível observar como os dados de consumo do ar condicionado são esparsos, visto que em uma parte considerável do dia este encontra-se desligado. Para lidar com tal característica, realizou-se o *downsampling* dos dados de uma para três horas entre medições, aumentando assim a acurácia do modelo.

Treinando com os dados coletados de uma das residências por um período de 20 dias, e tomando como base de validação um período de 10 dias, obteve-se um erro médio absoluto de 0.1939 e erro médio quadrático de 0.1386. A Figura 10 mostra a evolução da perda nas bases de treinamento e validação durante a etapa de aprendizado do modelo.

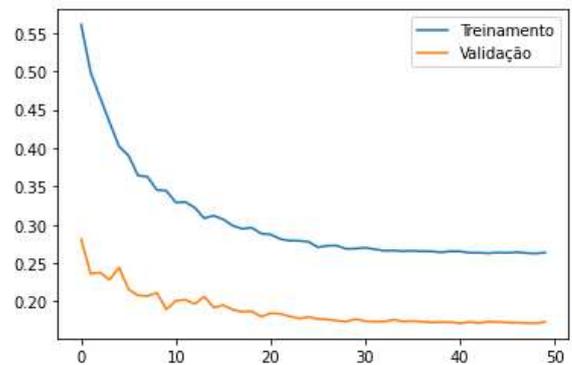


Fig. 10. Gráfico da métrica da perda em função da época de treinamento

A Figura 11 mostra um exemplo de previsão realizada pelo modelo a partir dos dados de consumo e temperatura das últimas 12 horas, podendo-se observar que o modelo ainda pode ser aperfeiçoado, entretanto, fugindo do escopo deste trabalho. Como aplicação, este pode ser utilizado para aprimorar a previsão do consumo energético total da residência. Para isto, também é necessário um tempo de amostragem maior, sendo esta uma das justificativas para o período de coleta do projeto, além de uma análise mais profunda dos parâmetros e arquitetura do modelo.

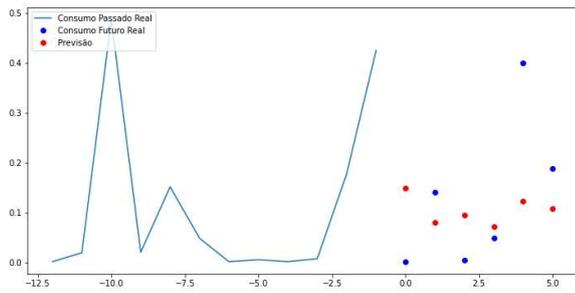


Fig. 11. Gráfico da previsão do consumo energético do ar condicionado

IV. CONSIDERAÇÕES FINAIS

A qualidade de dados de consumo energético e sua disponibilização é essencial para fomentar a pesquisa que endereça os diversos desafios do setor elétrico. Bancos de dados públicos como o Boa Energia podem fomentar estudos específicos para o cenário brasileiro, além do desenvolvimento de algoritmos de predição de consumo de energia residencial. Ao contrário dos bancos de dados presentes na literatura, ao incluir variáveis exógenas como temperatura interna da residência e o consumo de até 8 canais, permite a exploração de correlações e outros aprimoramentos ao desenvolvimento de algoritmos de predição. A contribuição do artigo está: na análise da falha de lacuna temporal na sequência de dados a partir de uma árvore de falhas, uma proposta de arquitetura de “Data Harvesting” que torna a coleta de dados tolerante a falhas intermitentes de conexão; e na disponibilização de uma base de dados de consumo energético por setores em granularidade horária e temperatura interna de uma residência, coletados durante um mês.

V. TRABALHOS FUTUROS

O aplicativo Android na arquitetura de “Data Harvesting” proposta está em desenvolvimento desde fevereiro de 2020. A disponibilização do hardware de baixo custo e publicação futura semelhante à [7], mas com disponibilização de 8 entradas em único módulo na arquitetura tolerante a falhas de comunicação está prevista, assim como a inclusão da medição de tensão para a computação da potência ativa pelo medidor de energia construído (dado que no estágio atual o medidor apenas realiza a medição de valores de corrente e estima a potência e consumo).

A instalação da solução em mais 2 residências foi realizada em março/2020, com previsão de instalar em mais 2 residências. A disponibilização da base de dados das 5 residências em conjunto com os resultados dos testes de usabilidade de aplicativos móveis e display integrado no medidor complementarão os dados de um mês de uma residência disponíveis no Github [17], e fazem parte de um projeto de conhecimento aberto em IoT [18]. A disponibilização dos dados por longo período (mais de um ano) permitirá análises de impactos da sazonalidade de estações do ano no consumo residencial.

A arquitetura proposta também pode ser usada na construção de bases de dados com diferentes periodicidades, que devem atender a variados casos de uso, como tarifação e qualidade de energia.

REFERÊNCIAS

- [1] ANEEL. “Tarifa Branca”. Disponível em: https://www.aneel.gov.br/sala-de-imprensa-exibicao/-/asset_publisher/XGPXSqMFHrE/content/tarifa-branca-e-opcao-para-consumidores-em-2020/656877?inheritRedirect=false. Acesso em 20 de janeiro de 2020.
- [2] MINISTRO DE ESTADO DE MINAS E ENERGIA. “PORTARIA No 301, DE 31 DE JULHO DE 2019”. Diário Oficial de 01.08.2019, seção 1, p. 3, v. 157, n. 147-A.
- [3] Empresa de Pesquisa Energética (EPE). “Anuário Estatístico de Energia Elétrica 2018”. Disponível em <http://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/anuario-estatistico-de-energia-eletrica>. Acesso em 20 de janeiro de 2020.
- [4] Haroldo Luiz Moretti do Amaral. “Desenvolvimento de uma nova metodologia para previsão do consumo de energia elétrica de curto prazo utilizando redes neurais artificiais e decomposição de séries temporais”. Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Energia e Automação Elétricas, São Paulo, 2019.
- [5] Y. Wang, Q. Chen, T. Hong, and C. Kang. “Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges,” IEEE Transactions on Smart Grid, vol. 10, no. 3, pp. 3125–3148, may 2019.
- [6] H. Ziekow, M. Strohbach, P. Kikiras, A. Leonardi, and P. Kikiras, “Dealing with Data Quality in Smart Home Environments Lessons Learned from a Smart Grid Pilot” [Online]. Disponível em: <https://www.researchgate.net/publication/296678374>. Acesso em 08 de abril de 2020.
- [7] Leonardi, Alessandro, Holger Ziekow, Martin Strohbach and Panayotis Kikiras. “Dealing with Data Quality in Smart Home Environments - Lessons Learned from a Smart Grid Pilot.” J. Sensor and Actuator Networks 5, 2016.
- [8] 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, Article number: 150007, 2015, DOI:10.1038/sdata.2015.7, 2015.
- [9] Sean Barker, Aditya Mishra, David Irwin, Emmanuel Cecchet, Prashant Shenoy, and Jeannie Albrecht, “Smart*: An Open Data Set and Tools for Enabling Research in Sustainable Homes”. Proceedings of the 2012 Workshop on Data Mining Applications in Sustainability (SustKDD 2012), Beijing, China, August 2012.
- [10] J. Zico Kolter and Matthew J. Johnson. “REDD: A public data set for energy disaggregation research”. In proceedings of the SustKDD workshop on Data Mining Applications in Sustainability, 2011.
- [11] R. Arakaki, V. Hayashi and W. Ruggiero, “Available and Fault Tolerant IoT System: Applying Quality Engineering Method”, ICECCE 2020 – 2nd International Conference on Electrical, Communication and Computer Engineering, Turkey, June 2020. No prelo.
- [12] V. Hayashi et. al, “Hedwig - Casa Conectada”. Escola Politécnica da Universidade de São Paulo, São Paulo, 2017.
- [13] Juang, Philo & Oki, Hidekazu & Wang, Yong & Martonosi, Margaret & Peh, Li-Shiuan & Rubenstein, Daniel. “Energy-efficient computing for wildlife tracking: Design tradeoffs and early experiences with ZebraNet”. ACM SIGPLAN Notices, 2002.
- [14] G. Eibl and D. Engel, “Influence of data granularity on smart meter privacy,” IEEE Transactions on Smart Grid, vol. 6, no. 2, pp. 930–939, 2015.
- [15] F. Luo, G. Ranzi, X. Wang, and Z. Y. Dong, “Service Recommendation in Smart Grid: Vision, Technologies, and Applications,” Proceedings of International Conference on Service Science, ICSS, vol. 0. Institute of Electrical and Electronics Engineers Inc., pp. 31–38, 2016.
- [16] D. J. Sailor and A. A. Pavlova, “Air conditioning market saturation and long-term response of residential cooling energy demand to climate change,” Energy, vol. 28, no. 9, pp. 941–951, 2003.
- [17] V. Hayashi. “Boa Energia” [Online]. Disponível em: <https://github.com/vthayashi/OKIoT/tree/master/BoaEnergia>. Acesso em 16 de abril de 2020.
- [18] V. Hayashi, V. Garcia, R. M. Andrade and R. Arakaki, “OKIoT Open Knowledge IoT Project: Smart Home Case Studies of Short-term Course and Software Residency Capstone Project”, IoTBDS 2020 – 5th International Conference on Internet of Things, Big Data and Security, Czech Republic, May 2020. No prelo.