

Uma proposta de desagregação de energia baseada no Modelo Oculto de Markov

Igor H. D. Fontes¹, Cristopher G. S. Freitas¹, Eduardo G. dos Santos¹,
Andre L. L. Aquino¹

¹Instituto de Computação – Universidade Federal de Alagoas

{ihdf, cristopher, egs, alla}@laccan.ufal.br

Abstract. *The energy disaggregation allows us to separate the aggregate energy consumption of a household into its contributing appliances. In this work, we used an energy disaggregation method based on the building of Hidden Markov Models that describes the power demand of appliances and we proposed the use of Baum-Welch algorithm to the parameter estimation for the refrigerator and washing machine model. Each model was obtained and evaluated using the Tracebase and REDD Dataset. The results presents a general model for each appliance, which were constructed using until 7 training instances, so that the metrics presents a better convergence. Lastly, each model was adjusted to disaggregate the energy.*

Resumo. *A desagregação de energia nos permite separar o consumo agregado de energia de uma residência em seus aparelhos contribuintes. Nesse trabalho, usamos um método de desagregação de energia baseado na construção de Modelos Ocultos de Markov que descrevem a demanda de energia de aparelhos de uso doméstico e propusemos o uso do algoritmo de Baum-Welch para a estimativa dos parâmetros do modelo do refrigerador e da lavadora. Cada modelo foi obtido e avaliado com os datasets Tracebase e REDD. Os resultados apresentam um modelo geral para cada aparelho, os quais foram construídos usando até 7 instâncias de treinamento, de modo que a métrica apresenta uma melhor convergência. Por fim, cada modelo foi ajustado para desagregar a energia.*

1. Introdução

A busca de alternativas eficientes para o desenvolvimento sustentável e a conservação dos recursos naturais tem sido um assunto marcante na academia e na indústria. Dessa forma, a manutenção da eficiência energética e o monitoramento do consumo de equipamentos eletroeletrônicos é essencial para a gestão dos recursos naturais no século XXI. Com isso, a desagregação da demanda de energia total de um cliente, de modo a descrever o consumo individual de cada aparelho, surge como uma alternativa para melhorar a forma como utilizamos a energia elétrica. Por fim, também podemos identificar comportamentos irregulares da rede elétrica e escalonar a utilização de máquinas.

A comunidade científica tem proposto várias técnicas para a desagregação da energia. Inicialmente, [Hart 1992] propôs a utilização sequencial de aparelhos de modo a permitir a detecção da assinatura de cada aparelho à medida que eles são ativados; [Lam et al. 2007] sugeriram a criação de bases de dados, com diferentes assinaturas de

vários aparelhos possíveis, para a construção de modelos que generalizam o comportamento de um dispositivo. No entanto, esses métodos para o monitoramento da demanda individual de energia não se mostram aplicáveis em grande escala, já que demandam alto custo e tempo de execução. Desse modo, soluções para o monitoramento não-intrusivo da carga de um aparelho, ou NILM (*Non-intrusive load monitoring*) [Hart 1992], são desafiadoras para a comunidade científica. Atualmente, a dificuldade de identificação de sinais individuais de aparelhos em meio ao sinal agregado é facilitada por técnicas de inteligência artificial. Além disso, o fato de utilizarmos apenas o sinal agregado para o monitoramento não-intrusivo já constitui um fator marcante de economia de dados e custos operacionais, pois basta utilizarmos um único ponto de monitoramento no ambiente a ser analisado para aquisição dos dados necessários para desagregação de energia.

Entre os modelos de desagregação existentes, podemos destacar aqueles baseados nos Modelos Ocultos de Markov, ou HMM (*Hidden Markov Model*), os quais viabilizam a construção de modelos probabilísticos que descrevem o comportamento geral de cada tipo de aparelho utilizado no ambiente a ser monitorado [Parson et al. 2014] [Parson et al. 2012]. Esses modelos utilizam processos não-supervisionados para detectar, no sinal agregado, assinaturas de aparelhos específicos em funcionamento e, assim, construir modelos usados na desagregação de energia.

Os modelos ocultos de Markov também têm sido usados para formular previsões do consumo de energia com a utilização de métodos alternativos para o reconhecimento de padrões, em que é possível verificar a aplicação do algoritmo de Baum-Welch para a inferência dos parâmetros de um HMM, como proposto em [Kwak et al. 2018]. [Kong et al. 2016] também propuseram o uso de métodos alternativos de estimação de parâmetros para modelos ocultos de Markov fatoriais.

2. Materiais e métodos

2.1. Descrição da proposta

Nesse trabalho, analisamos os algoritmos disponíveis para a estimação dos parâmetros de um HMM e construímos, com o uso do algoritmo de Baum-Welch [Sammut and Webb 2011], um HMM para cada dispositivo elétrico comum aos diversos tipos de residências, de modo que os parâmetros de cada modelo puderam ser ajustados para descrever dispositivos de ambientes ainda não monitorados. Tais modelos foram usados como base para a desagregação da energia, já que facilitam a detecção de padrões característicos de cada tipo de dispositivo na série temporal da demanda total de energia de uma residência, por exemplo. Em suma, nossa proposta é utilizar o algoritmo de Baum-Welch para a estimação dos parâmetros de um HMM quando aplicado ao problema da desagregação de energia baseado em modelos gerais para cada tipo de aparelho. Abordagem até então não identificada na literatura.

2.2. Datasets

Para a avaliação da proposta apresentada, utilizamos duas bases de dados: i) Tracebase dataset, a qual contém a demanda energética de diversas instâncias de cada tipo de aparelho e ii) REDD Dataset, que contém dados agregado e desagregado de 6 casas diferentes. Os resultados apresentaram uma proposta de obtenção de um modelo geral para cada aparelho de forma que apenas 7 instâncias de treinamento foram necessárias para que a

métrica de avaliação convergisse. Destacamos que o algoritmo de Baum-Welch foi utilizado como um estimador para HMMs, representando uma alternativa para a criação de modelos gerais de aparelhos.

2.3. Seções

O restante do trabalho está organizado na seguinte forma: Na seção 3, descrevemos um Modelo Oculto de Markov e seu uso para a construção de modelos de aparelhos elétricos. Na seção 4, construímos modelos gerais para cada tipo de aparelho. Na seção 5, descrevemos os resultados e usamos dados do *REDD Dataset* para demonstrar o uso do modelo proposto para a desagregação de energia. Por fim, na seção 6, apresentamos a conclusão e trabalhos futuros, sendo as figuras e tabelas expostas de autoria própria.

3. Modelos de aparelhos elétricos usando Modelos Ocultos de Markov

A definição formal de um HMM inclui uma cadeia de Markov de variáveis discretas, a qual representa a sequência de estados ocultos, de forma que as probabilidades de transição de estados são definidas por uma matriz A . Além disso, cada HMM é composto por uma sequência de variáveis contínuas, as quais são chamadas observações e estão relacionadas pelos estados ocultos por uma probabilidade de emissão de observações. As variáveis discretas representam os estados operacionais de um aparelho, enquanto as contínuas, a demanda de energia observada do aparelho. Uma representação de um modelo oculto de Markov pode ser vista na Figura 1.

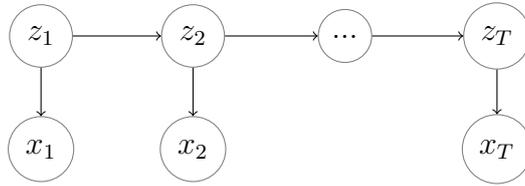


Figura 1. Um modelo oculto de Markov (as variáveis z e x representam os estados ocultos e as observações, respectivamente)

Considere a variável z_t como uma variável discreta que assume algum valor dos k estados possíveis para cada aparelho, como podem ser vistos na Figura 2. Já a variável contínua x_t representa o consumo de energia do aparelho (número decimal positivo) no instante t da série temporal usada para treinamento do modelo. Por exemplo, podemos dizer que uma máquina de lavar está no estado $z_3 = On$ e emite uma observação $x_3 = 1302.6$ W em $t=3$.

Cada HMM pode ser descrito por três parâmetros. O primeiro deles corresponde à probabilidade do dispositivo estar no estado k da variável oculta z_t em $t = 1$, que pode ser representada pelo vetor

$$\pi_k = p(z_1 = k). \quad (1)$$

Além de π , temos a matriz A , a qual representa a probabilidade de transição de um estado i em $t - 1$ para o estado j em t tal que

$$A_{i,j} = p(z_t = j | z_{t-1} = i). \quad (2)$$

O último dos parâmetros é dado pela probabilidade de emissão de um observação x descrita pela função governada por ϕ . Nesse caso, é assumida uma distribuição gaussiana

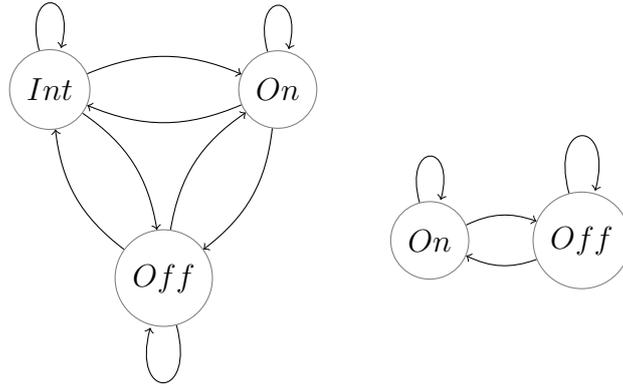


Figura 2. Modelos de 3 e 2 estados, respectivamente

[Kim et al. 2011], pois foi previamente mostrada como uma boa aproximação para a demanda de energia de um aparelho tal que

$$x_t | z_t, \phi \sim N(\mu_{z_t}, \sigma_{z_t}). \quad (3)$$

As três equações mencionadas acima podem ser usadas para calcular a verossimilhança do HMM

$$p(x, z | \theta) = p(z_1 | \pi) \prod_{t=2}^T p(z_t | z_{t-1}, A) \prod_{t=1}^T p(x_t | z_t, \phi), \quad (4)$$

em que $\theta = \{\pi, A, \phi\}$ são os parâmetros do modelo de uma dada instância do aparelho analisado.

A fim de aprender os parâmetros dos HMMs, distribuições a priori são atribuídas aos parâmetros do modelo. Para as probabilidades iniciais e para a matriz de transição, temos uma inicialização com distribuições uniformes, ou seja, as transições são admitidas como equiprováveis. Já para as probabilidades de emissão de observáveis, selecionamos os parâmetros de cada gaussiana (média e desvio padrão) com base num conhecimento prévio acerca do consumo médio (em W) de cada aparelho, os quais podem ser vistos na Tabela 1.

Tabela 1. Parâmetros iniciais para as distribuições normais que constituem a descrição das probabilidades de emissão de observáveis para cada aparelho

Aparelho	μ_{off}	σ_{off}	μ_{int}	σ_{int}	μ_{on}	σ_{on}
Refrigerador	0	10	100	31.27	-	-
Máquina de lavar	0	10	150	10	1350	100

Os parâmetros adotados para as distribuições normais a priori em cada estado z foram escolhidos de modo que o estado Off tenha médias de consumo próximas a 0 W e desvio padrão pequeno se comparado aos desvios-padrão escolhidos para os estados On

e Int (no caso dos aparelhos modelados com 3 estados). Tais escolhas foram motivadas pelo fato de que, no estado de menor média de consumo (Off), espera-se que os tipos de aparelhos modelados exijam uma demanda próxima de 0 W, enquanto, nos estados de maior média de consumo, cada aparelho possui uma média característica e um desvio padrão maior do que no estado Off, devido à maior variabilidade na demanda de consumo entre as diversas instâncias de um tipo de aparelho.

Desse modo, ao invés de utilizarmos métodos de inferência variacional, como proposto por [Parson et al. 2014], os quais são mais complexos e demandam maior tempo de implementação, usamos o algoritmo de Baum-Welch para estimar os parâmetros dos modelos de cada instância de um dado tipo de aparelho com o uso do pacote para modelagem probabilística chamado Pomegranate [Schreiber 2017], implementado em Python. Na Figura 3, podemos ver um esquema do processo de treinamento do modelo proposto.

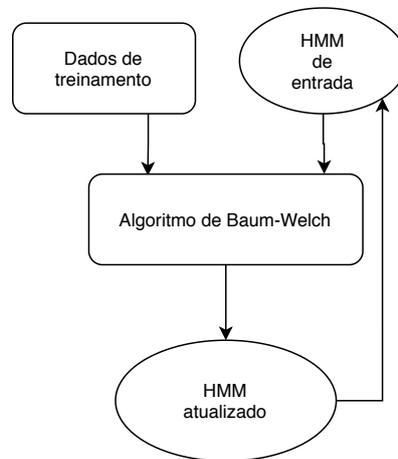


Figura 3. Esquema do treinamento de cada HMM com o uso de Baum-Welch

A fim de termos uma generalização sobre múltiplas instâncias, temos que cada conjunto de parâmetros θ é usado para construir um modelo geral com parâmetros Θ que representam cada tipo de aparelho por inteiro. Isso pode ser feito por meio da retirada de amostras das distribuições a posteriori sobre os parâmetros θ das instâncias de cada aparelho e da determinação de distribuições que se ajustem a estas amostras.

Para as probabilidades iniciais e para a matriz de transição de cada modelo geral, usamos a média dos parâmetros π e A de cada instância disponível para cada tipo de aparelho. Por exemplo, para cada instância disponível do aparelho refrigerador, treinamos um HMM e usamos tais médias mencionadas como parâmetros para descrever o modelo geral desse tipo de aparelho.

Para as probabilidades de emissão de observáveis, extraímos amostras de tamanho fixo de cada distribuição normal obtida dos modelos individuais treinados, de modo que tivemos dados advindos de várias instâncias de um único tipo de aparelho. Posteriormente, usamos máxima verossimilhança para estimar os parâmetros da distribuição normal que melhor se adequa às amostras retiradas.

Vale salientar que a estimativa dos parâmetros de um HMM com o uso do algoritmo de Baum-Welch envolve multiplicações de probabilidades, o que pode provocar a ocorrência de *underflow* durante a execução do algoritmo para séries temporais muito

grandes. Desse modo, a implementação incluiu o uso da técnica de escalagem para prevenir a ocorrência desses erros [Stamp 2017].

4. Construção dos modelos gerais

Para avaliação da criação de um modelo geral para cada tipo de aparelho, utilizamos o *Tracebase dataset* [Reinhardt et al. 2012], que contém dados de múltiplas instâncias de cada tipo de aparelho com taxa de amostragem de $1Hz$. Em média, para cada tipo de aparelho contido nesse conjunto de dados, temos, pelo menos, 8 instâncias disponíveis. Cada instância contém séries temporais da demanda de energia individual do aparelho, as quais serão usadas para treinamento do modelo geral de cada tipo de dispositivo.

Para validação dos modelos gerais construídos com o *Tracebase dataset*, utilizamos o conjunto de dados REDD [Kolter and Johnson 2011], o qual contém informações detalhadas da demanda energética de várias casas e é, frequentemente, utilizado para fins de desagregação de energia. O REDD possui sinais agregado e desagregado de 6 casas diferentes de um período de, aproximadamente, um mês, ou seja, os sinais do consumo total da residência e da demanda individual de cada aparelho contribuinte, em que a taxa de amostragem dos dados é de $1/3Hz$. Desse modo, pudemos avaliar o quanto um dado modelo geral ajustado descrito por Θ explica a demanda real de um aparelho.

A fim de avaliar os modelos gerais ajustados, os quais denotamos por AT, utilizamos dados da casa 2 do *REDD Dataset* para executar a busca por janelas de tempo para treinamento. Foi feito um pré-processamento dos dados de modo que a taxa de amostragem foi alterada de $1/3Hz$ para $1/60Hz$ de forma que as janelas pudessem ser identificadas mais facilmente. Usaremos o exemplo do refrigerador para comparar os modelos disponíveis com os propostos nesse trabalho.

O log da função de verossimilhança $p(x|\Theta)$ foi a métrica utilizada para a validação do modelo geral de cada aparelho, em que x é a série temporal de teste e Θ é o conjunto de parâmetros de cada modelo geral. O cálculo dessa função foi feita com o uso do algoritmo *forward*, implementado com o uso do pacote de Python, Pomegranate [Schreiber 2017].

Com isso, os dados do *Tracebase dataset* foram selecionados de modo que cada tipo de aparelho tivesse, pelo menos, 8 instâncias de treinamento. Daí, os dados de uma dessas instâncias foram usados como conjunto de teste e os demais, usados para a construção do modelo geral. Cada modelo geral foi comparado por meio do *cross validation* com dois *benchmarks*. Um desses *benchmarks* foi treinado com dados que contêm a instância de teste (modelo ST), o qual representa um limite superior para $p(x|\Theta)$, já que modelo geral Θ é treinado com o uso dos dados de teste, ou seja, a probabilidade do conjunto de dados x pertencer a um tipo de aparelho descrito pelo modelo geral Θ é maior se comparada a GT e AT; o outro é treinado com dados de um único aparelho diferente do de teste (modelo NT), ou seja, o treinamento de NT não favorece a descrição de um tipo de aparelho em geral, mas de uma instância particular. Desse modo, NT representa um limite inferior para $p(x|\Theta)$. Vale salientar que basta calcularmos a métrica de NT e ST uma única vez, já que os dados para o treinamento desses modelos são advindos de instâncias únicas θ_n do aparelho analisado.

Finalmente, os modelos gerais (modelos GT) foram criados com o uso de dados de 2 a 7 instâncias de cada tipo de aparelho, de modo que foi possível demonstrar a

viabilidade da construção desses modelos com o uso de poucas instâncias de treinamento. Cada conjunto de treinamento foi reparticionado 15 vezes e a média dos logs de $p(x|\Theta)$ é calculada para fins de validação do modelo.

Por fim, realizamos ajustes no modelo geral para que ele possa descrever o padrão de consumo de um aparelho contido numa residência ainda não analisada com o uso de dado agregado (demanda total de energia do ambiente). Com isso, o método utilizado para ajustar cada modelo geral Θ foi baseado na identificação de janelas de tempo no dado agregado x que captassem intervalos nos quais apenas o aparelho analisado muda de estado operacional, como proposto por [Parson et al. 2014]. Essa captação pode ser feita por meio do cálculo de $p(x|\Theta)$, que representa a probabilidade de um intervalo de dado agregado ter sido gerado por uma instância específica de um aparelho descrito pelo modelo geral Θ . Antes de calcular $p(x|\Theta)$, devemos subtrair a carga base da janela de dados escolhida, ou seja, subtrair a demanda mínima de energia no dado agregado, de forma que o sinal de consumo do aparelho de interesse seja evidenciado no dado agregado após o ajuste: $x_{a:b}^* = x_{a:b} - \min(x_{a:b})$, em que $x_{a:b}^*$ é a janela de dados após a subtração e $x_{a:b}$, a janela original entre os instantes a e b . As janelas $x_{a:b}^*$ que apresentarem $p(x_{a:b}^*|\Theta)$ maior do que um valor limite v previamente calculado serão usadas para ajuste do modelo geral Θ por meio do algoritmo de Baum-Welch.

O valor limite v corresponde ao menor valor de $p(x_{a:b}|\Theta)$ calculado entre todas as instâncias θ do aparelho disponíveis no conjunto de treinamento dos modelos gerais. O tamanho da janela de tempo para cada aparelho foi escolhida como sendo a maior janela de tempo em que um aparelho daquele tipo estivesse em funcionamento no *Tracebase dataset*. No caso do refrigerador, a janela foi de $200min$. Dessa forma, podemos identificar assinaturas de aparelhos não vistos anteriormente e ajustar os parâmetros do modelo geral Θ ao aparelho de interesse.

5. Resultados

Para os modelos gerais, o aparelho modelado com o uso de 2 estados foi o refrigerador, enquanto a máquina de lavar foi modelada com o uso de 3 estados. A escolha da quantidade de estados se baseou no padrão de demanda energética de cada dispositivo, nos componentes elétricos internos e nos ciclos de funcionamento mais comuns, como exposto em outros métodos de desagregação de energia que utilizam Modelos Ocultos de Markov [Parson et al. 2014].

Como exemplo de validação do modelo proposto, implementamos o modelo geral da máquina de lavar e determinamos os parâmetros do HMM que descrevem esse tipo de aparelho. Na Figura 4, podemos ver a distribuição gaussiana obtida para a emissão de observáveis no estado operacional *On* da máquina de lavar. Tal distribuição reflete as características esperadas para um modelo geral, as quais incluem uma média de consumo próxima do valor estabelecido a priori expostos na Tabela 1. Além disso, temos um desvio-padrão maior se comparado às emissões no estado *Off*, já que há uma menor variabilidade no consumo da máquina de lavar no estado *Off*.

Na Figura 5, podemos ver os resultados obtidos na validação do modelo geral da máquina de lavar, na qual podemos perceber que, à medida que o número de instâncias usadas para construir o modelo geral GT proposto aumenta, a média dos logs de $p(x|\Theta)$ cresce cada vez mais lentamente até aproximadamente convergir. Desse modo, verifica-

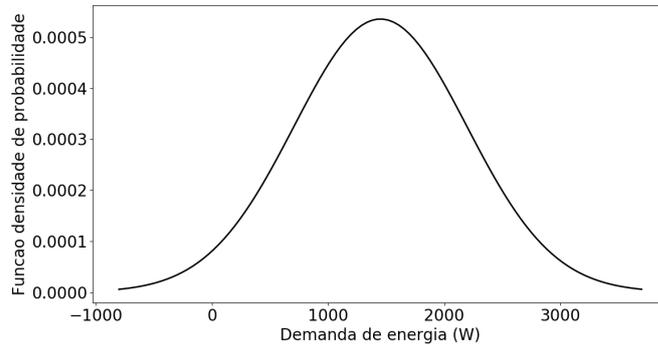


Figura 4. Função densidade de probabilidade para a emissão de observáveis para o estado O_n do modelo geral da máquina de lavar

mos que, mesmo em aparelhos com mais de 2 ciclos de funcionamento (como a máquina de lavar), é possível construir modelos gerais com até 7 instâncias de treinamento. Com isso, podemos evitar o uso de enormes bancos de dados para o treinamento de modelos gerais e, assim, reduzir o custo de obtenção desses modelos.



Figura 5. logs médios obtidos da validação cruzada do modelo da máquina de lavar para conjuntos de treinamento de tamanhos crescentes

Para criação de modelos que descrevam um aparelho específico de um ambiente não visto anteriormente, ajustamos os modelos gerais por meio do processo descrito anteriormente com dados do *REDD Dataset*. Para a validação do ajuste para fins de desagregação de energia, utilizamos o modelo geral do refrigerador e ajustamos os parâmetros desse modelo.

Na Figura 6, vemos a função densidade de probabilidade da emissão de observáveis do refrigerador no estado *Off* antes(modelo GT) e depois do ajuste(modelo AT). Podemos perceber que houve uma mudança na média e no desvio-padrão de forma que o modelo AT possui um desvio-padrão menor do que o de GT, já que AT foi ajustado para descrever um refrigerador específico, enquanto GT, um modelo geral do refrigerador.

Para comparar os modelos, utilizamos o log da função de verossimilhança $p(x_{teste}|\Theta)$ como métrica, em que x_{teste} é o dado desagregado do refrigerador (consumo individual) e Θ_i é um dos modelos de referência i denotados por GT, AT, ST. O modelo ST corresponde ao modelo treinado com dados desagregados, ou seja, com os dados de

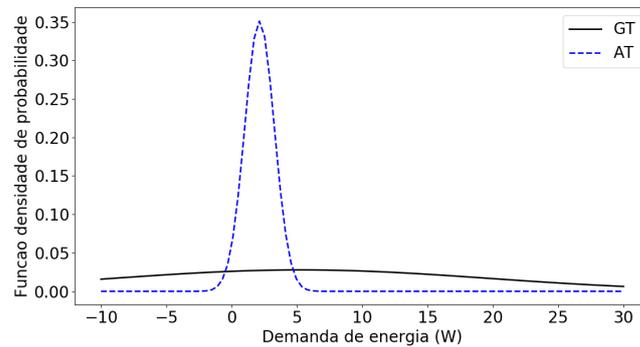


Figura 6. Funções densidade de probabilidade para a emissão de observáveis no estado *Off* do refrigerador antes e depois do ajuste

teste. Dessa forma, $p(x_{teste}|\Theta_{ST})$ corresponde a um limite superior para os modelos GT e AT.

Na Figura 7, observamos que o log para os modelos GT e AT propostos nesse trabalho possuem valores próximos aos do modelo ST, de modo que podem estimar o comportamento da instância de teste do refrigerador por meio dos parâmetros de cada um desses modelos. Assim, verificamos que é viável a utilização de modelos gerais ajustados com base em Modelos Ocultos de Markov para desagregação de energia com o uso do algoritmo de Baum-Welch.

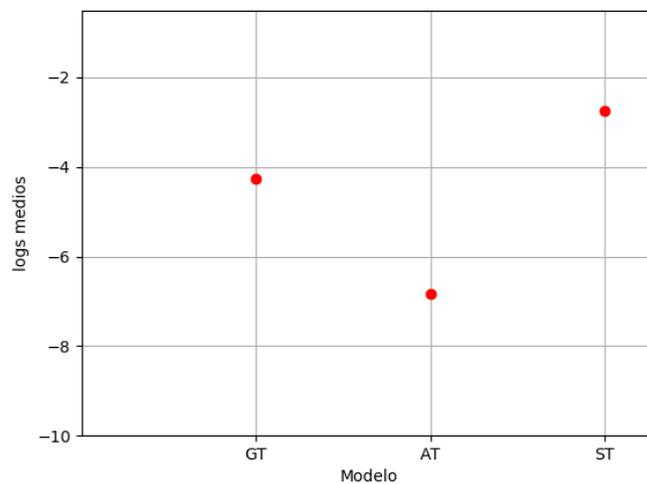


Figura 7. Média dos logs para diferentes modelos

6. Conclusão e trabalhos futuros

O modelo proposto representa uma alternativa de desagregação de energia baseada em Modelos Ocultos de Markov de forma que os parâmetros de cada modelo foram estimados com o uso do algoritmo de Baum-Welch. Os modelos gerais por aparelho validados nesse trabalho constituem uma forma eficaz de evitar processos que demandam grandes

quantidades de dados, além de contribuir para identificar assinaturas de aparelhos não-vistos anteriormente.

Em trabalhos futuros, pretendemos modelar outros aparelhos não contemplados nesse artigo, bem como adicionar comparações com propostas de desagregação e algoritmos já existentes por meio de aspectos quantitativos e qualitativos. o que inclui novas métricas de avaliação, tempo de execução e número de iterações. Ademais, analisar a sensibilidade dos parâmetros para as distribuições a priori de cada tipo de aparelho e definir um método eficaz para otimizar a desagregação de energia com base na escolha desses parâmetros.

Referências

- Hart, G. W. (1992). Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891.
- Kim, H., Marwah, M., Arlitt, M., Lyon, G., and Han, J. (2011). Unsupervised disaggregation of low frequency power measurements. In *Proceedings of the 2011 SIAM international conference on data mining*, pages 747–758. SIAM.
- Kolter, J. Z. and Johnson, M. J. (2011). Redd: A public data set for energy disaggregation research. In *Workshop on data mining applications in sustainability (SIGKDD)*, San Diego, CA, volume 25, pages 59–62.
- Kong, W., Dong, Z. Y., Ma, J., Hill, D. J., Zhao, J., and Luo, F. (2016). An extensible approach for non-intrusive load disaggregation with smart meter data. *IEEE Transactions on Smart Grid*, 9(4):3362–3372.
- Kwak, Y., Hwang, J., and Lee, T. (2018). Load disaggregation via pattern recognition: A feasibility study of a novel method in residential building. *Energies*, 11(4):1008.
- Lam, H. Y., Fung, G., and Lee, W. (2007). A novel method to construct taxonomy electrical appliances based on load signatures. *IEEE Transactions on Consumer Electronics*, 53(2):653–660.
- Parson, O., Ghosh, S., Weal, M., and Rogers, A. (2012). Non-intrusive load monitoring using prior models of general appliance types. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*.
- Parson, O., Ghosh, S., Weal, M., and Rogers, A. (2014). An unsupervised training method for non-intrusive appliance load monitoring. *Artificial Intelligence*, 217:1–19.
- Reinhardt, A., Baumann, P., Burgstahler, D., Hollick, M., Chonov, H., Werner, M., and Steinmetz, R. (2012). On the accuracy of appliance identification based on distributed load metering data. In *2012 Sustainable Internet and ICT for Sustainability (SustainIT)*, pages 1–9. IEEE.
- Sammut, C. and Webb, G. I. (2011). *Encyclopedia of machine learning*. Springer Science & Business Media.
- Schreiber, J. (2017). Pomegranate: fast and flexible probabilistic modeling in python. *The Journal of Machine Learning Research*, 18(1):5992–5997.
- Stamp, M. (2017). *Introduction to machine learning with applications in information security*. Chapman and Hall/CRC.