

Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Coordenação de Pós-Graduação em Ciência da Computação

Dissertação de Mestrado

Estratégias de agrupamento de consumidores  
residenciais para o melhoramento de ações de  
eficiência energética

Harllan Andryê Bezerra Silva

Campina Grande, Paraíba, Brasil

Maio – 2017

Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Coordenação de Pós-Graduação em Ciência da Computação

Estratégias de agrupamento de consumidores  
residenciais para o melhoramento de ações de  
eficiência energética

Harllan Andryê Bezerra Silva

Dissertação submetida à Coordenação do Curso de Pós-Graduação em  
Ciência da Computação da Universidade Federal de Campina Grande -  
Campus I como parte dos requisitos necessários para obtenção do grau  
de Mestre em Ciência da Computação.

Área de Concentração: Ciência da Computação

Linha de Pesquisa: Eficiência Energética

Andrey Elísio Monteiro Brito

(Orientador)

Campina Grande, Paraíba, Brasil

©Harllan Andryê Bezerra Silva, 12/05/2017

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

S586e Silva, Harllan Andryê Bezerra.  
Estratégias de agrupamento de consumidores residenciais para o melhoramento de ações de eficiência energética / Harllan Andryê Bezerra Silva. – Campina Grande, 2017.  
84 f. : il. color.

Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2017.  
"Orientação: Prof. Dr. Andrey Elísio Monteiro Brito".  
Referências.

1. Eficiência Energética. 2. Agrupamento de Dados. 3. Consumidor Residencial.  
I. Brito, Andrey Elísio Monteiro. II. Título.

CDU 621.31(043)

## Resumo

O consumo de energia elétrica vem crescendo a cada dia. Precisamos utilizar a energia elétrica de forma consciente, pois os recursos naturais que são utilizados para a geração de energia podem acabar devido ao seu uso ineficiente. O crescimento populacional das últimas décadas, o aparecimento de mais aparelhos eletrônicos e eletrodomésticos geram um consumo excessivo de energia. Devido ao crescimento no consumo de energia elétrica é necessária a implantação de programas de eficiência energética, que se dá através da introdução de novas tecnologias, incentivo à mudança de hábito do próprio consumidor e uso racional de energia elétrica. O foco deste trabalho é no setor residencial, que é o segundo maior consumidor de energia elétrica no Brasil, e como há consumidores que compartilham características e padrões de carga semelhantes, isso possibilita o uso de agrupamento de dados. Pensando nisso é proposto o uso de agrupamento para auxiliar programas de eficiência energética na análise dos dados dos consumidores e na criação de grupos representativos de uma população. A criação de grupos ajuda a concessionária de energia a fornecer ofertas comerciais ou recomendações específicas para grupos específicos, diminuir a complexidade das análises que teriam que ser feitas em uma população e obter relacionamentos personalizados, mais eficazes e equitativos entre os fornecedores de energia e seus clientes. O agrupamento irá proporcionar a aplicação de soluções que ajudem o consumidor a utilizar energia elétrica de forma eficiente, a partir do momento em que ele recebe informações sobre seu consumo e como ele poderá utilizar essas informações, sabendo o que elas irão proporcionar como resultado. Este trabalho iniciou-se com a investigação de medidas de dissimilaridade para representar a semelhança entre perfis de consumo de energia elétrica (um dos fatores utilizados para os agrupamentos) e entre as três medidas utilizadas a distância Euclidiana se destacou com os melhores resultados nos experimentos feitos, seja variando a quantidade de observações das séries ou a base de dados. Após isso foram feitos agrupamentos utilizando 4 fatores extraídos da base de dados e assim criados 15 cenários de agrupamentos a partir da combinação desses fatores. Por meio dos resultados desses agrupamentos foi possível reduzir a quantidade de cenários por serem semelhantes e também escolher os cenários (fatores) mais relevantes a serem considerados quando se quer criar grupos de consumidores residenciais.

## **Abstract**

The consumption of electric energy has been increasing every day. We need to use electric power in a conscious way, because the natural resources that are used for the generation of energy can end up due to its inefficient use. The population growth of the last decades, the appearance of more electronic devices and appliances generate an excessive consumption of energy. Due to the growth in the consumption of electric energy, it is necessary to implement energy efficiency programs, which are carried out through the introduction of new technologies, an incentive to change the consumer's habit and rational use of electric energy. The focus of this work is on the residential sector, which is the second largest consumer of electricity in Brazil, and since there are consumers who share similar characteristics and load patterns, this allows the use of data grouping. Thinking about that, the use of clustering to support energy efficiency programs in the analysis of consumer data and in the creation of representative groups of a population is proposed. Groups creation helps the utility to provide commercial offers or specific recommendations for specific groups, reduce the complexity of the analyzes that would have to be done in a population, and get personalized, more effective and equitable relationships between energy suppliers and their customers. The clustering will provide the application of solutions that help the consumer to use electricity efficiently, from the moment he receives information about his consumption and how he can use that information, knowing what they will provide as a result. This work began with the investigation of measures of dissimilarity to represent the similarity between profiles of electric energy consumption (one of the factors used for the clustering) and among the three measures used the Euclidean distance stood out with the best results in the experiments made, either by varying the number of observations of the series or the database. After that, clusters were made using 4 factors extracted from the database and thus 15 clustering scenarios were created from the combination of these factors. Through the results of these clustering it was possible to reduce the number of scenarios to be similar and also to choose the most relevant scenarios to consider when creating groups of residential consumers.

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Contexto . . . . .	1
1.2	Justificativa . . . . .	4
1.3	Objetivos . . . . .	5
1.4	Organização do documento . . . . .	6
<b>2</b>	<b>Trabalhos relacionados</b>	<b>8</b>
2.1	Eficiência energética e ciência do consumo . . . . .	8
2.2	Agrupamento de curvas de carga . . . . .	11
2.3	Considerações finais . . . . .	13
<b>3</b>	<b>Fundamentação teórica</b>	<b>15</b>
3.1	Agrupamento de dados . . . . .	15
3.2	Medidas de dissimilaridade . . . . .	17
3.3	Métricas de validação . . . . .	21
<b>4</b>	<b>Materiais e métodos</b>	<b>25</b>
4.1	Construção da base de dados . . . . .	25
4.2	Cenários propostos e configurações de grupos . . . . .	30
4.3	Agrupamento . . . . .	33
4.3.1	Perfil de consumo . . . . .	34
4.3.2	Algoritmo de agrupamento adaptativo . . . . .	35
<b>5</b>	<b>Resultados e discussão</b>	<b>41</b>
5.1	Similaridade entre os perfis de consumo . . . . .	41

---

5.2	Agrupamentos de consumidores de energia elétrica . . . . .	47
5.2.1	Analisando 1 grupo de cada cenário relevante . . . . .	52
5.2.2	Análise subjetiva e a métrica silhueta . . . . .	59
5.3	Considerações finais . . . . .	63
<b>6</b>	<b>Conclusão</b>	<b>65</b>
6.1	Como o sistema poderia ser usado na prática? . . . . .	67
6.2	Limitações e Trabalhos futuros . . . . .	68
<b>A</b>	<b>Posse de aparelhos e Hábitos de consumo de energia elétrica</b>	<b>75</b>
<b>B</b>	<b>Avaliação de grupos feita por especialistas</b>	<b>79</b>
<b>C</b>	<b>Avaliação de grupos</b>	<b>83</b>

# Lista de Figuras

1.1	Matriz elétrica brasileira. . . . .	2
2.1	Desagregação de equipamentos. . . . .	10
2.2	Economias residenciais devido ao <i>feedback</i> sobre o consumo de energia. . .	11
2.3	Exemplo de curvas de carga típicas. . . . .	12
3.1	Processo de mineração de dados. . . . .	16
3.2	Exemplo do algoritmo <i>k-means</i> . . . . .	17
3.3	Exemplo da Distância Euclidiana. . . . .	19
3.4	Caminho a seguir usando DTW. . . . .	19
3.5	Exemplo da medida de distância DTW. . . . .	20
3.6	Ilustração dos elementos envolvidos no cálculo da silhueta. . . . .	23
3.7	Curvas de carga típica e reais. . . . .	24
4.1	Percentual de consumidores por faixa de consumo. . . . .	27
4.2	Percentual de aparelhos existentes nas residências. . . . .	28
4.3	Percentual de consumidores por classes de recomendação. . . . .	29
4.4	Funcionamento do gerador de carga. . . . .	30
4.5	Exemplo dos dados normalizados. . . . .	33
4.6	Necessidade da normalização dos perfis de consumo. . . . .	34
4.7	Modelo do algoritmo adaptativo. . . . .	36
4.8	Fluxograma do <i>feedback</i> automático. . . . .	40
5.1	Valores da silhueta e MAE com intervalo de confiança de 95% (31 observações). . . . .	42
5.2	Valores da silhueta e MAE com intervalo de confiança de 95% (7 observações). . . . .	43

---

5.3	Valores da silhueta e MAE com intervalo de confiança de 95% (24 observações). . . . .	44
5.4	Agrupamento de 20 curvas de carga (1 dia - 24 observações). . . . .	45
5.5	Agrupamento das curvas de carga da base REFIT (144 observações). . . . .	46
5.6	Agrupamento das curvas de carga da base REFIT (24 observações). . . . .	47
5.7	Exemplo de um agrupamento com 2 e 3 grupos. . . . .	51
5.8	Localização geográfica de cada residência de um grupo do cenário 1. . . . .	53
5.9	Perfis de consumo de um grupo do cenário 15. . . . .	56
5.10	Perfis de consumo de um grupo do cenário 11. . . . .	58
5.11	Agrupamento do cenário 1. . . . .	61
5.12	Agrupamento do cenário 1 - localização geográfica. . . . .	62
B.1	Classificação dos grupos do 1º conjunto de grupos avaliado. . . . .	81
B.2	Classificação dos grupos do 2º conjunto de grupos avaliado. . . . .	82

# Lista de Tabelas

4.1	Classes das recomendações. . . . .	28
4.2	Recomendações e suas classes. . . . .	38
4.3	Dados obtidos do questionário. . . . .	38
5.1	Limites do intervalo de confiança para a silhueta e MAE (31 observações). . . . .	43
5.2	Limites do intervalo de confiança para a silhueta e MAE (7 observações). . . . .	43
5.3	Limites do intervalo de confiança para a silhueta e MAE (24 observações). . . . .	44
5.4	Valor da silhueta e MAE para o agrupamento com 20 curvas de carga. . . . .	45
5.5	Valor da silhueta e MAE (base REFIT - 144 observações). . . . .	46
5.6	Valor da silhueta e MAE (base REFIT - 24 observações). . . . .	46
5.7	Grupos formados nos cenários 2 e 9. . . . .	49
5.8	Cenários semelhantes em cada configuração de grupo. . . . .	50
5.9	Cenários e sua relevância. . . . .	52
5.10	Informações sobre as residências de um grupo do cenário 1. . . . .	53
5.11	Informações sobre as residências de um grupo do cenário 9. . . . .	54
5.12	Informações sobre as residências de um grupo do cenário 11. . . . .	57
5.13	Aparelhos em cada residência de um grupo do cenário 13. . . . .	58
5.14	Classificação dos grupos do cenário 1. . . . .	63
B.1	Comparação entre a classificação feita por especialistas e pela Silhueta (1º conjunto de grupos). . . . .	81
B.2	Comparação entre a classificação feita por especialistas e pela Silhueta (2º conjunto de grupos). . . . .	82

# Capítulo 1

## Introdução

### 1.1 Contexto

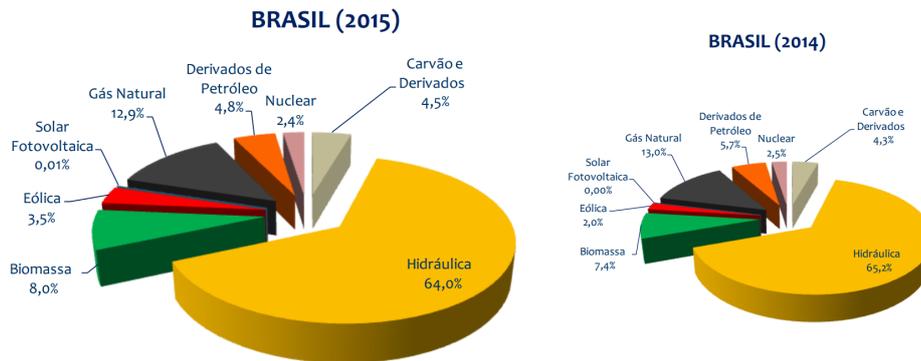
O consumo de energia elétrica vem crescendo a cada dia. Precisamos utilizar a energia elétrica de forma consciente, pois os recursos naturais que são utilizados para a geração de energia podem acabar devido ao seu uso ineficiente. O consumo sustentável e consciente de energia reduz os impactos negativos ambientais e econômicos. Em particular, os impactos negativos ambientais são causados pela utilização de recursos naturais não renováveis, como por exemplo o petróleo, gás natural e carvão, e a emissão de gases de efeito estufa pelas fontes geradoras de energia elétrica.

O Brasil dispõe de uma matriz elétrica de origem predominantemente renovável, com destaque para a geração hidráulica que responde por 64% da oferta interna. As fontes renováveis representam 75,5% da oferta interna de eletricidade no Brasil, que é a resultante da soma dos montantes referentes à produção nacional mais as importações, que são essencialmente de origem renovável (ENERGIA, 2016).

A principal fonte de geração de energia elétrica no Brasil é a hidráulica, que apresentou uma redução de 3,7% no ano de 2015 na comparação com o ano anterior. O consumo final de eletricidade no país em 2015 registrou um recuo de 1,8%, mas para suprir a redução do uso das hidrelétricas houve a expansão da geração térmica, especialmente das usinas movidas a carvão mineral e biomassa, cujas participações na matriz elétrica, na comparação de 2015 contra 2014, cresceram. A Figura 1.1 ilustra esse cenário.

O avanço da participação de recursos renováveis na matriz elétrica brasileira, apenas de

Figura 1.1: Matriz elétrica brasileira.



Fonte: Adaptado de Energia (2016).

0,9% em relação a 2014, é devido à queda da geração térmica a derivados de petróleo e ao incremento das gerações a biomassa e eólica, apesar da redução da oferta hidráulica, que do ano de 2012 até 2015 caiu em 20,5% devido às condições hidrológicas desfavoráveis. O consumo de eletricidade no setor residencial reduziu em 0,7% no ano de 2015 em relação ao ano anterior, devido ao aumento da tarifa e queda da atividade econômica. Porém essa redução é pouca comparada com o aumento no consumo de eletricidade desse setor ao longo do tempo, pois de 2006 a 2014 o consumo aumentou em 12,4% (ENERGIA, 2016).

É preocupante o crescimento do uso de energia elétrica, pois o consumo excessivo pode causar impactos negativos ao meio ambiente e para a sociedade. O crescimento populacional das últimas décadas, o aparecimento de mais aparelhos eletrônicos e eletrodomésticos geram um consumo excessivo de energia. Esta realidade pode ser revertida com tecnologias eficientes e mudanças nos nossos hábitos de consumo.

Com a necessidade de aumentar a sustentabilidade e diminuir custos aparece o conceito de eficiência energética que, segundo a agência internacional de energia (*International Energy Agency* - IEA), é uma forma de gerir e restringir o crescimento no consumo de energia (AGENCY, 2016). Com a eficiência energética buscamos conservar os recursos não renováveis, otimizar o uso de energia e combater os desperdícios.

Com o objetivo de promover a eficiência energética no Brasil, o governo implantou em 1985 o Programa Nacional de Conservação de Energia Elétrica - PROCEL<sup>1</sup>, coordenado pelo Ministério de Minas e Energia. Esse programa é bastante conhecido pelo Selo Procel de

<sup>1</sup><http://www.procelinfo.com.br/>

economia de energia encontrado em equipamentos e eletrodomésticos, indicando a eficiência dos mesmos, ajudando aos consumidores na compra de equipamentos mais eficientes.

Mesmo com algumas ações do governo, o desperdício de energia elétrica ainda existe e algumas pessoas não sabem o que fazer para utilizar a energia elétrica de forma eficiente, pois não basta apenas colocar selos de eficiência em equipamentos ou trocar lâmpadas. No setor residencial o desperdício de energia pode ser causado por diversos fatores, como por exemplo: mau uso de determinado equipamento, defeito em algum aparelho, problema na instalação elétrica da residência, entre outros.

Hoje, quando recebemos a conta de energia em nossas casas, só sabemos qual o valor da conta e quanto em kWh foi gasto durante o mês. Desta forma a falta de informação aos consumidores dificulta o uso eficiente da energia elétrica, devido à diversos fatores que causam o desperdício. Vários estudos (BATRA et al., 2014; TRUNG et al., 2014; BATRA, 2015) estão sendo feitos na área de eficiência energética com o objetivo de ajudar as pessoas a economizar energia. Com esse objetivo, estão sendo desenvolvidos medidores inteligentes (smart meters) que captam os dados da rede elétrica e retornam para o usuário, que por meio de uma interface em seu computador pode verificar seu consumo de energia detalhado em um período desejado e detectar de alguma forma anomalias que estão (ou podem estar) causando o desperdício de energia.

Medidores inteligentes constituem uma tecnologia que é bastante promissora, porém no Brasil ainda não temos medidores inteligentes em escala residencial. Mesmo não tendo estes medidores inteligentes no Brasil em escala residencial, esse tipo de serviço é fornecido por empresas que utilizam um equipamento instalado junto ao medidor de energia elétrica já existente na residência.

Mesmo com o surgimento de programas do governo no Brasil na área de eficiência energética, o consumo de energia continua aumentando. Houve diminuição no desperdício de energia com programas de eficiência energética, mas mesmo assim as pessoas continuam desperdiçando energia por não terem informações suficientes que lhe auxiliem no uso da energia elétrica de forma eficiente. Com base nisso temos o problema principal deste estudo, fornecer informações que ajudem os consumidores no uso eficiente da energia elétrica.

## 1.2 Justificativa

Como discutido anteriormente, o setor residencial é o segundo maior consumidor de energia elétrica no Brasil (ENERGIA, 2016) – cerca de 21,3% do consumo total em 2015, no qual este trabalho tem como alvo esse setor, devido ao crescimento no consumo de energia elétrica e pela facilidade de conseguir as informações necessárias para a elaboração deste trabalho.

Devido a esse crescimento no consumo é necessária a implantação de programas de eficiência energética, que se dá através da introdução de novas tecnologias, incentivo à mudança de hábito do próprio consumidor em relação ao seu consumo e uso racional de energia elétrica. Consumir energia de maneira racional e eficiente significa evitar desperdícios de energia, sem prejudicar o nível de conforto e qualidade de vida do usuário (CASTRO, 2015).

Sabendo que há muitos consumidores que compartilham características e padrões de carga semelhantes (KANG; LEE, 2015), essas semelhanças levantam a possibilidade do uso de agrupamento de dados. Pensando nisso é proposto o uso de agrupamento para auxiliar programas de eficiência energética na análise dos dados dos consumidores e na criação de grupos representativos de uma população, no qual criar grupos ajuda a:

- Diminuir a complexidade das análises que teriam que ser feitas em uma população, pois não é preciso analisar cada residência, apenas os grupos formados, uma vez que são compostos por residências semelhantes de acordo com suas características;
- Concessionária de energia (ou a alguma empresa da área) a fornecer ofertas comerciais ou recomendações específicas para grupos específicos, a partir do momento que ela conhece o comportamento real dos seus clientes;
- Obter relacionamentos personalizados, mais eficazes e equitativos entre os fornecedores de energia e clientes (ZEDAN et al., 2010; RÄSÄNEN; RUUSKANEN; KOLEHMAINEN, 2008).

Essas recomendações também são chamadas na literatura de *feedback* (VINE; BUYS; MORRIS, 2013) e dependendo de como o problema é modelado, o *feedback* pode ativar um desejo de economia de custos ou para minimizar o impacto ambiental. O *feedback* pode ser

combinado com outros instrumentos, como os incentivos de preços, fixação de metas ou uma competição (FISCHER, 2008).

Basicamente, uma pessoa deve saber sobre a natureza do problema, as opções existentes e suas respectivas consequências, a fim de julgá-los em termos de normas e motivos (FISCHER, 2008). Em geral, a literatura conclui que o *feedback* pode reduzir o consumo de energia elétrica em casas de 5% a 20% (HOUDE et al., 2013; VINE; BUYS; MORRIS, 2013; ZVINGILAITE; TOGEBY, 2015). A redução de consumo de energia se dá ao longo do tempo (gradativamente), pois, geralmente, envolvem mudanças de hábito (HOUDE et al., 2013).

No final deste estudo as contribuições alcançadas foram: mostrar que o uso de agrupamento é viável, quais fatores (informações) são mais relevantes para criar grupos e mostrar formas de utilização da informação, mesmo na presença de informação limitada sobre consumidores reais. Uma vez que os fatores relevantes para o agrupamento de residências foi definido, estes fatores são úteis também para a recomendação de itens para os consumidores residenciais, que tipicamente é feito sem agrupamentos, mas usando fatores para a identificação de consumidores que poderiam ter apoiado recomendações semelhantes.

### 1.3 Objetivos

Este trabalho tem o objetivo de apresentar uma solução para programas de eficiência energética, buscando entender e avaliar o processo de agrupamento quando aplicado nesse contexto e isso é importante pois quando feito na prática, com dados reais, pode ajudar na criação de grupos representativos de uma população e na análise dos dados dos consumidores residenciais.

O agrupamento irá proporcionar a aplicação de soluções que ajudem o consumidor a utilizar energia de forma eficiente, a partir do momento em que ele recebe informações sobre seu consumo e como ele poderá utilizar essas informações, sabendo o que elas irão proporcionar como resultado. Algumas das dificuldades da atuação do consumidor residencial na busca por conservação de energia é a falta de informação e apoio, uma vez que o consumidor consciente de seus impactos possui dificuldades em aplicar o que conhece sobre conservação de energia, identificar e quantificar a energia poupável (FOURNIER, 2010).

Como objetivos específicos, temos:

- Avaliar estratégias de agrupamento dentro do contexto de eficiência energética utilizando alguns fatores, tais como: perfis de consumo, aparelhos existentes nas residências, localização geográfica e *feedback*<sup>2</sup> do consumidor;
- Verificar quais fatores (ou combinação deles) são mais relevantes para criar grupos;
- Analisar como a abordagem proposta pode ser utilizada na prática.

A avaliação das estratégias de agrupamento será por meio de uma análise subjetiva dos grupos formados e também observando o valor da métrica silhueta (métrica para validação de agrupamentos). Essa análise subjetiva mostra se os grupos formados fazem sentido, ou seja, se eles são compostos por residências com características semelhantes (consumos parecidos, mesmo aparelhos, entre outras).

## 1.4 Organização do documento

Este trabalho possui seis capítulos organizados da seguinte forma:

No Capítulo 2 (Trabalhos relacionados) são apresentados os trabalhos encontrados na literatura na área de eficiência energética, que tratam sobre agrupamentos de curvas de carga e como auxiliar as pessoas na utilização da energia elétrica de forma eficiente.

No Capítulo 3 (Fundamentação teórica) são abordadas algumas técnicas que foram utilizadas para a elaboração desta dissertação, como o algoritmo de agrupamento, as medidas de dissimilaridade utilizadas entre as curvas de carga (perfis de consumo), assim como as métricas de validação para avaliar a qualidade dos grupos obtidos.

No Capítulo 4 (Material e Método) são apresentados detalhes da construção da base de dados, os cenários propostos e configurações de grupos utilizando alguns fatores para gerar os agrupamentos, assim como o método utilizado para encontrar a medida de dissimilaridade que representa a semelhança entre as curvas de carga.

---

<sup>2</sup>*Feedback* em relação às recomendações passadas para cada consumidor, de acordo com as ações que estão dispostos a realizar para a redução do consumo.

---

No Capítulo 5 (Resultados e discussão) são apresentados os diversos experimentos realizados com as medidas de dissimilaridade para encontrar qual é a ideal para representar a semelhança entre as curvas de carga. Também são abordados os resultados dos agrupamentos utilizando alguns fatores extraídos da base de dados e com isso foi possível escolher os fatores mais relevantes a serem considerados quando se quer criar grupos de consumidores residenciais semelhantes.

No Capítulo 6 (Conclusão) são apresentadas as conclusões a respeito dos experimentos feitos, como o sistema (agrupamento) poderia ser utilizado na prática, assim como as limitações e trabalhos futuros.

# Capítulo 2

## Trabalhos relacionados

Uma vasta gama de trabalhos podem ser encontrados na área de eficiência energética com o objetivo de auxiliar as pessoas na utilização da energia elétrica de forma eficiente, fornecendo *feedback* a respeito do seu consumo por meio de informações obtidas das residências, tais como: valor do consumo, aparelhos existentes na residência, utilização da energia ao longo do tempo (perfis de consumo), entre outras.

### 2.1 Eficiência energética e ciência do consumo

A necessidade de utilizar energia de forma eficiente a nível mundial iniciou a partir da década de 1970, com os choques no preço do petróleo. Nesse momento, ficou claro que o uso das reservas de recursos fósseis teria custos crescentes, seja do ponto de vista econômico, seja do ponto de vista ambiental (ENERGÉTICA, 2010). Desde então, no Brasil, diversas medidas vêm sendo tomadas para diminuir o impacto ambiental e o desperdício de energia, dentre as quais foram criados alguns programas, como: Programa Brasileiro de Etiquetagem (PBE), Programa Nacional de Conservação de Energia Elétrica (PROCEL) e Programa Nacional de Conservação de Petróleo e Derivados (CONPET).

Mesmo com algumas ações do governo, o desperdício de energia elétrica ainda existe e algumas pessoas não sabem o que fazer para utilizar a energia elétrica de forma eficiente ou fazem ações que resultam em pouco impacto, devido à não disponibilidade de informações que os auxiliem na tomada de decisões a respeito do uso eficiente da energia elétrica.

Eficiência energética é consumir energia de forma consciente, sem desperdícios e res-

tringindo o seu crescimento, sem alteração na qualidade de vida e conforto do consumidor. Várias pesquisas estão sendo feitas com esse objetivo, auxiliando as pessoas no uso eficiente da energia elétrica fornecendo *feedback* sobre o consumo de energia. Basicamente, uma pessoa deve saber sobre a natureza do problema, as opções existentes e suas respectivas consequências, a fim de julgá-los em termos de normas e motivos (FISCHER, 2008).

Dependendo de como o problema é moldado, o *feedback* pode ativar um desejo de economia de custos ou para minimizar o impacto ambiental, e pode ser combinado com outros instrumentos, como os incentivos financeiros, a fixação de metas ou compromisso pessoal (FISCHER, 2008). Informações passadas para os consumidores com maior frequência (diária, semanal) são mais eficazes na redução do valor da conta ou na utilização eficiente da energia elétrica (FISCHER, 2008; VINE; BUYS; MORRIS, 2013; ZVINGILAITE; TOGEBY, 2015).

O agrupamento de consumidores proporciona o uso de *feedback* mais específico para o grupo e que segundo Fischer (2008), a utilidade da informação depende fortemente da forma como é apresentado e se é específico para as necessidades do grupo alvo. Em geral, a literatura conclui que o *feedback* pode reduzir o consumo de energia elétrica em casas de 5% a 20% (HOUDE et al., 2013; VINE; BUYS; MORRIS, 2013; ZVINGILAITE; TOGEBY, 2015), no qual essa economia é maior quando são fornecidas informações a nível de aparelho.

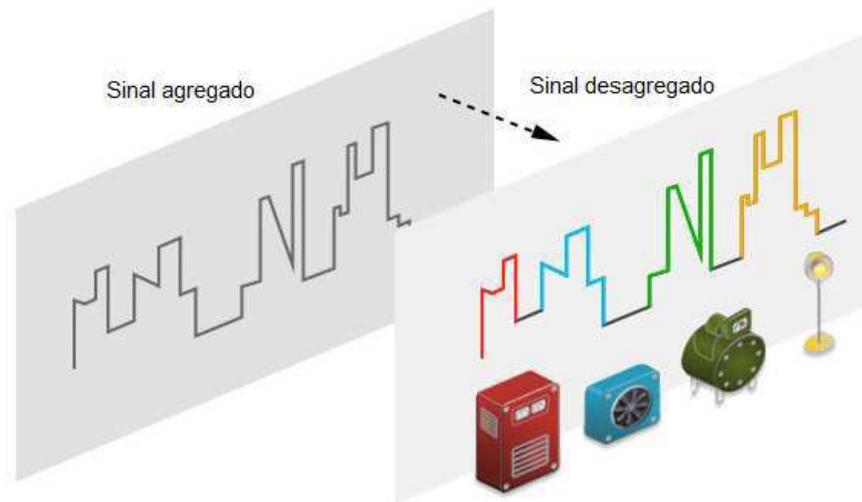
Para fornecer dados a nível de aparelho é utilizada uma técnica de monitoramento de carga não intrusivo (NIALM, do inglês *Nonintrusive Appliance Load Monitoring*) (BATRA, 2015; BATRA et al., 2014; HART, 1985), que é um processo para analisar as mudanças de tensão e corrente elétrica em uma residência e deduzir quais aparelhos estão sendo utilizados na casa. Para obter esses dados é utilizado um único sensor junto ao medidor de energia da residência, que coleta os dados agregados da rede elétrica e envia para um servidor, no qual possui um algoritmo baseado no processo de NIALM que é executado com o propósito de detectar eventos<sup>1</sup> e desagregar os equipamentos conectados à rede elétrica. A Figura 2.1 ilustra um exemplo de desagregação.

Fornecer aos consumidores dados desagregados do seu consumo de energia elétrica é

---

<sup>1</sup>Um evento se refere a uma variação na corrente elétrica e esta possivelmente reflete uma alteração no estado de um equipamento, por exemplo: passar do estado ligado para o desligado.

Figura 2.1: Desagregação de equipamentos.



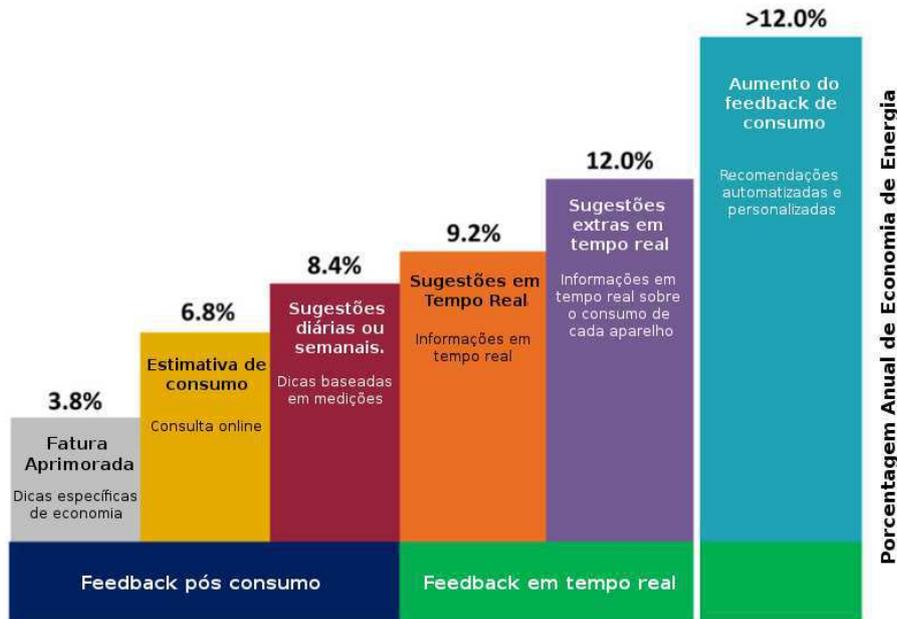
Fonte: Adaptado de Efficiency (2015).

mais eficaz do que fornecer dados agregados (e.g., consumo mensal). Pesquisas feitas de 1995 a 2013 mostram que informações a nível de aparelho alcançou um maior percentual de economia de energia anual (>12%) em relação à informações baseadas em dados agregados (ARMEL et al., 2013). A Figura 2.2 ilustra os resultados dessas pesquisas em relação aos efeitos de fornecer aos consumidores *feedback* sobre o seu consumo de eletricidade, no qual o *feedback* pós consumo pode resultar em uma economia de até 8,4% e essa economia pode ser maior se fornecido um *feedback* em tempo real e a nível de aparelho (e.g., consumo de cada aparelho).

Fornecer informações desagregadas pode, de fato, levar o consumidor a economizar energia elétrica (BATRA; SINGH; WHITEHOUSE, 2015), por essa razão um dos objetivos dessa dissertação é analisar o agrupamento utilizando dados a nível de aparelho, além de diminuir a complexidade das análises que teriam que ser feitas em uma população.

O *feedback* é uma estratégia promissora para promover a conservação de energia e que é mais eficaz quando é combinado com intervenções (metas, comparações, ...) e quando é aplicado com maior frequência (BEKKER et al., 2010; HOUDE et al., 2013; KARLIN; ZINGER; FORD, 2015).

Os trabalhos apresentados mostram que a informação a respeito do consumo de eletricidade é uma estratégia promissora para ajudar os consumidores no uso eficiente da energia

Figura 2.2: Economias residenciais devido ao *feedback* sobre o consumo de energia.

Fonte: Adaptado de Armel et al. (2013).

elétrica e esse tipo de informação foi utilizada no presente trabalho, passando recomendações sobre economia de energia para os grupos de consumidores formados no processo de agrupamento descrito no capítulo 4.

## 2.2 Agrupamento de curvas de carga

O uso de curvas de carga (ou perfis de consumo) ajuda a analisar e identificar o comportamento dos consumidores em relação à utilização da energia elétrica ao longo do tempo e também possibilita prever o crescimento da demanda de eletricidade. A tarefa de agrupamento de perfis de consumo tem como objetivo descobrir um conjunto de modelos de perfis ou padrões que representem o mais fielmente possível o conjunto de dados original (IGLESIAS; KASTNER, 2013; FERREIRA et al., 2013).

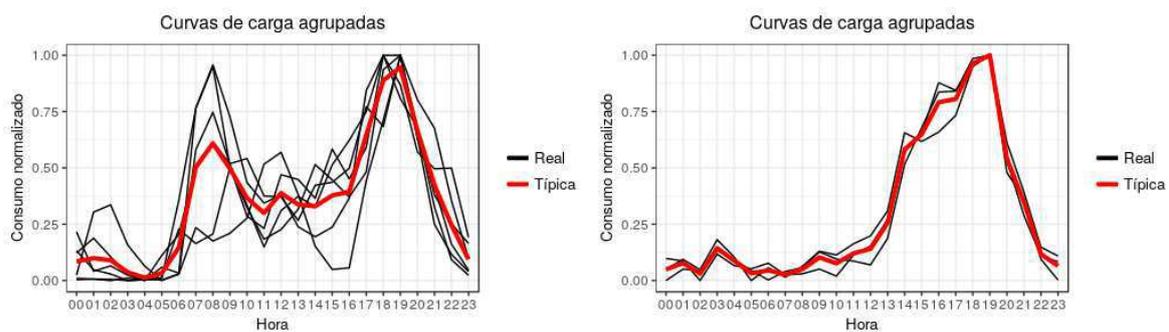
Uma das dificuldades desse tipo de agrupamento é encontrar uma medida de dissimilaridade entre as séries (perfis). Na literatura encontramos alguns trabalhos sobre comparação entre medidas de dissimilaridade e as mais usadas são: correlação de Pearson, distância euclidiana e alinhamento temporal dinâmico (DTW, do inglês *Dynamic Time Warping*) (IGLE-

SIAS; KASTNER, 2013; WILLIAMS, 2013; BENÍTEZ et al., 2014). Entre essas três medidas, a distância euclidiana é a mais utilizada, pois mostrou-se ser melhor para representar a semelhança entre as curvas de carga, devido à ter um erro médio absoluto menor (KIM; KO; CHOI, 2011) e formar grupos mais homogêneos (IGLESIAS; KASTNER, 2013; LAVIN; KLABJAN, 2015).

Teive et al. (2015) agrupou curvas de carga diárias de consumidores industriais do estado de Santa Catarina para gerar tipologias de curvas de carga típicas utilizando o algoritmo *k-means* para criar os grupos, porém não foi informado no artigo qual medida de dissimilaridade foi utilizada entre as curvas de carga. Na Figura 2.3 temos um exemplo dessas curvas, onde a curva típica (em vermelho) é a média entre as curvas de carga que fazem parte do mesmo grupo e as demais (em preto) são os dados reais dos consumidores. Essa curva de carga típica representa o comportamento do uso de eletricidade ao longo do tempo para o grupo formado.

Também foram aplicados questionários nas indústrias para obter algumas informações como: horário de funcionamento, porte da empresa, consumo mensal, entre outras. Os dados dos questionários são relacionados com as curvas típicas e isso pode ajudar a classificar um novo consumidor que não possui dados de medição, aplicando um questionário e depois atribuindo esse consumidor a uma tipologia de carga.

Figura 2.3: Exemplo de curvas de carga típicas.



Outro trabalho que também usou o agrupamento de curvas de carga para criar tipologias é o de Andrade et al. (2013). Nesse trabalho utilizou-se dados de consumidores industriais e comerciais com o objetivo de apresentar uma metodologia para descoberta de conjuntos (*clusters*) mínimos e significativos desses consumidores, onde cada classe consumidora era descrita por um número reduzido de tipologias. O agrupamento também pode ser utilizado

para outros fins, como é o caso de Müller e Franco (2014) que usou o agrupamento de curvas de carga para a previsão de cargas e para reduzir o conjunto de dados que imponham menores cargas computacionais ao algoritmo de previsão, no qual foi utilizado o algoritmo *k-means* para o agrupamento e a distância euclidiana como medida de distância entre as curvas.

Segundo Lavin e Klabjan (2015) as características qualitativas, como tipo de construção, localização e assim por diante, não são suficientes no agrupamento de consumidores para comparação, porque elas não captam suas tendências de uso de energia. Ao contrário do agrupamento de perfis de consumo que, de um modo geral, é comumente usado para classificar os consumidores de energia (SMITH; WONG; RAJAGOPAL, 2012), prever a demanda futura de energia (TEIVE et al., 2015; MÜLLER; FRANCO, 2014) ou detectar valores atípicos (LI; BOWERS; SCHNIER, 2010).

A análise do consumo da energia elétrica ao longo do tempo, usando perfis de consumo, proporciona estabelecer relações entre consumo e hábitos dos consumidores. Neste âmbito, a demanda ou hábitos de consumo podem ser resumidos por modelos energéticos que nos levam a relacionamentos personalizados, mais eficazes e equitativos entre os fornecedores de energia e clientes (ZEDAN et al., 2010; RÄSÄNEN; RUUSKANEN; KOLEHMAINEN, 2008), assim como alcançar uma melhoria no funcionamento de uma rede inteligente (*smart grid*) (KANG; LEE, 2015).

O agrupamento de curvas de carga (ou perfis de consumo de energia elétrica) foi utilizado neste trabalho de dissertação para criar grupos de residências com curvas parecidas, ou seja, utilizamos curvas do setor residencial diferente dos trabalhos aqui apresentados, que utilizaram curvas de consumidores dos setores industrial e comercial. Além disso, elaboramos alguns experimentos para descobrir a melhor medida de dissimilaridade entre correlação, euclidiana e DTW, que são bastantes utilizadas na literatura.

## 2.3 Considerações finais

Neste capítulo foram abordados alguns trabalhos referentes a área de pesquisa que ajudaram na elaboração desta dissertação. Pode-se observar que a necessidade de utilizar energia de forma eficiente é um assunto que vem sendo discutido a bastante tempo e que vários pesquisadores estão procurando formas de ajudar os consumidores a poupar energia, utilizando-a

de forma eficiente.

O uso de *feedback* a respeito do consumo de energia elétrica é uma estratégia promissora para promover a conservação de energia (KARLIN; ZINGER; FORD, 2015), no qual é necessário o uso de *feedback* mais específico para obter melhores resultados, ou seja, redução do consumo utilizando a energia de forma eficiente. O agrupamento de consumidores proporciona isso, a partir do momento em que serão criados grupos distintos, cada um com suas características e assim poderão ser passadas informações específicas para cada grupo de acordo com suas necessidades.

Perfis de consumo de energia elétrica são informações que ajudam a entender como o consumidor utiliza a energia ao longo do tempo, por exemplo: maior consumo em horário de pico, que torna a fatura de energia mais cara. Sem esse nível de informação fica difícil oferecer algum tipo de *feedback* que auxilie o consumidor na tomada de decisão para diminuir seu consumo no horário de pico, por exemplo. Porém, dados de consumo nesse nível são difíceis de obter, pois há a necessidade de ter dispositivos que colem os dados das residências, como os medidores inteligentes, que ainda não é uma realidade no Brasil.

Durante essa revisão não foram encontrados trabalhos que utilizam diversos fatores para criar grupos de consumidores residenciais ou que fale sobre a dificuldade de obter tais informações. Por isso é possível entender a necessidade do desenvolvimento deste trabalho voltado para a análise de estratégias de agrupamento dentro do contexto de eficiência energética.

# Capítulo 3

## Fundamentação teórica

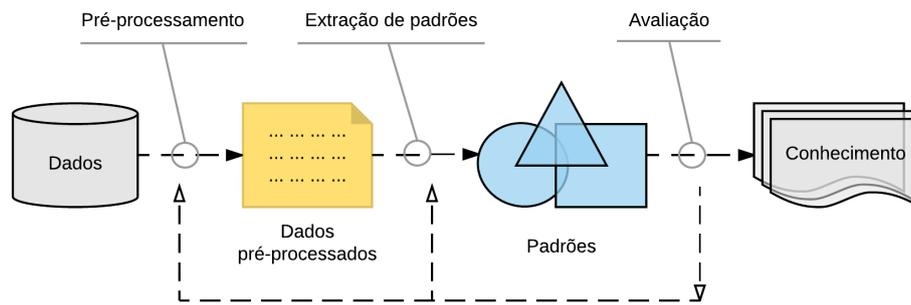
Neste capítulo são abordadas algumas técnicas que foram utilizadas para a elaboração desta dissertação. Inicialmente é descrito o processo de mineração de dados, mostrando suas etapas e qual algoritmo de agrupamento foi utilizado. Também são abordadas as medidas de dissimilaridade utilizadas entre as curvas de carga (perfis de consumo), assim como as métricas de validação para avaliar a qualidade dos grupos obtidos e comparar os resultados de alguns agrupamentos.

### 3.1 Agrupamento de dados

Analisar e compreender grandes quantidades de dados é uma tarefa difícil de se fazer manualmente e isso motivou diversos estudos a utilizarem técnicas automáticas de extração de conhecimento de bases de dados, as quais estão dentro da área de Mineração de Dados (MD). O processo de extração do conhecimento é iterativo e interativo, identificando padrões úteis e compreensíveis na base de dados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996; REZENDE, 2003). Na Figura 3.1 temos as etapas do processo de MD.

Esse processo de MD é dividido em algumas etapas (REZENDE, 2003): pré-processamento, extração de padrões, pós-processamento e a utilização do conhecimento. No pré-processamento há a preparação dos dados que devem ser tratados para a sua utilização, como por exemplo, para ser utilizado em um algoritmo de agrupamento, que é o foco deste trabalho. Dependendo dos dados o tratamento pode ser a retirada de valores incorretos ou ausentes da base, a normalização dos dados, a adequação dos dados no formato esperado

Figura 3.1: Processo de mineração de dados.



Fonte: Adaptado de Fayyad et al. (1996).

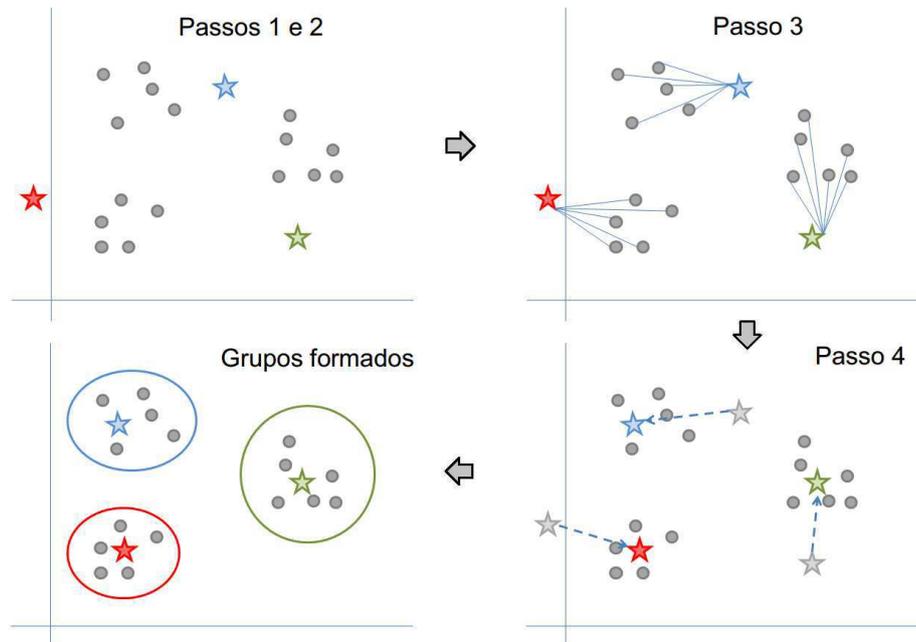
pelo algoritmo, entre outros.

Na extração de padrões é realizada a escolha do algoritmo, assim como sua configuração para extração do conhecimento. Após isso, temos a etapa de pós-processamento, no qual serão analisados os resultados obtidos na etapa anterior, seja por meio de uma avaliação subjetiva ou alguma métrica que valide os resultados. Por último temos a utilização do conhecimento, que foi extraído nas etapas anteriores e será utilizado em algum processo de tomada de decisão.

O agrupamento de dados, que faz parte da etapa de extração de padrões, tem como objetivo identificar grupos (*clusters*) em um conjunto de dados de acordo com a similaridade entre os objetos desse conjunto. Para o agrupamento dos dados deste trabalho foi utilizado o *k-means*, que é um algoritmo de aprendizagem não supervisionada e bastante utilizado na literatura (CHICCO, 2012; LAVIN; KLABJAN, 2015), além de ser escalável e poder convergir para grandes conjuntos de dados (SMITH; WONG; RAJAGOPAL, 2012). Basicamente, o algoritmo *k-means* pode ser descrito nos seguintes passos e ilustrado na Figura 3.2:

- **Passo 1:** Escolhe a quantidade  $k$  de grupos;
- **Passo 2:** Escolhe o valor, aleatoriamente, dos  $k$  centróides (pontos centrais dos grupos – na Figura 3.2 são as estrelas) iniciais;
- **Passo 3:** Calcula a distância (Euclidiana) entre cada elemento (objeto) e os centróides. Os elementos são alocados ao centróide mais próximo a ele;
- **Passo 4:** Calcula a média da distância entre os elementos dentro de cada grupo formado. Essa média será o novo centróide de cada grupo;

- **Passo 5:** Repete os passos 3 e 4 até não ocorrer alterações nos valores dos centróides.

Figura 3.2: Exemplo do algoritmo *k-means*.

Fonte: Adaptado de Pagans (2015).

Formalmente, para criar um agrupamento é fornecido um conjunto de dados  $D = \{d_1, d_2, \dots, d_n\}$ , resultando em um agrupamento com  $k$  grupos definido como:  $G = \{g_1, g_2, \dots, g_k\}$  com  $k \leq n$ , tal que (XU; WUNSCH II, 2005):

- $\bigcup_{i=1}^k g_i = D$  (todos os elementos pertencem ao conjunto de dados  $D$ );
- $g_i \neq \emptyset, 1 \leq i \leq k$  (os grupos contém pelo menos 1 elemento);
- $g_i \cap g_j = \emptyset, i, j = 1, \dots, k$  e  $i \neq j$  (cada elemento pertence exclusivamente a um único grupo)

## 3.2 Medidas de dissimilaridade

Na literatura existem diversas medidas de distância com o objetivo de definir o grau de semelhança entre dois objetos (IGLESIAS; KASTNER, 2013; LIAO, 2005), no qual as mais

utilizadas são: Correlação de Pearson, Distância Euclidiana e alinhamento temporal dinâmico (DTW, do inglês *Dynamic Time Warping*). Neste trabalho foram usadas medidas de dissimilaridade (ou medidas de distância) na descoberta de perfis de consumo semelhantes, as quais são descritas a seguir.

A Correlação de Pearson foi utilizada na comparação entre as séries temporais (perfis de consumo) para verificar a similaridade entre elas, no qual o cálculo da correlação foi feito de uma série para todas e todas para uma, construindo assim uma matriz de distância  $N \times N$ , onde  $N$  é a quantidade de séries da base de dados. O cálculo da correlação pode ser observado na Equação 3.1 (IGLESIAS; KASTNER, 2013).

$$corr_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.1)$$

Onde:

- $corr_{xy}$ : é a correlação entre a série X e a série Y;
- $\bar{x}$  e  $\bar{y}$ : a média da amostra das séries X e Y, respectivamente;
- $i$ : é cada observação da série temporal.

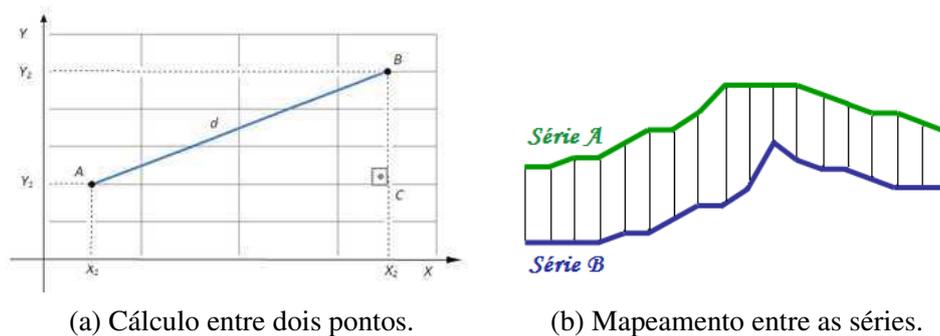
Essa correlação é calculada entre pares de pontos de duas séries temporais e elas precisam ter a mesma quantidade de observações, ou seja, ter o mesmo tamanho. O mesmo acontece com a Distância Euclidiana, que é a soma da raiz quadrada da diferença entre dois pontos (MÜLLER; FRANCO, 2014), por exemplo: dadas duas séries temporais  $A = (A_1, \dots, A_n)$  e  $B = (B_1, \dots, B_n)$  de tamanho N, a distância entre elas é definida de acordo com a Equação 3.2.

$$dis_{euclidiana}(A, B) = \sqrt{\sum_{i=1}^n (A_i - B_i)^2} \quad (3.2)$$

Na Figura 3.3 temos um noção intuitiva da Distância Euclidiana, onde na imagem da esquerda temos o que seria o cálculo entre os dois pontos A e B, e o cálculo na imagem da direita pode ser visto como a soma das linhas entre os pares de pontos das séries.

O DTW é um algoritmo para medir a semelhança entre duas séries temporais que podem variar no tempo (IGLESIAS; KASTNER, 2013). Encontra-se o alinhamento não-linear ideal

Figura 3.3: Exemplo da Distância Euclidiana.



(a) Cálculo entre dois pontos.

(b) Mapeamento entre as séries.

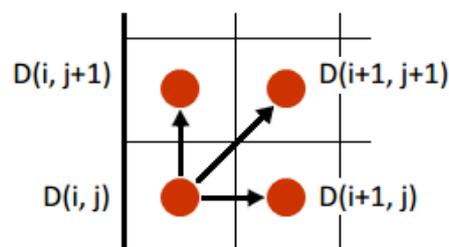
Fonte: Adaptado de Tsiporkova (2016).

entre duas séries, que pode ser visto como encontrar o caminho mais curto para ir a partir do ponto (1,1) até o ponto (n,m) da matriz.  $N$  é o comprimento de um série temporal e  $M$  é o comprimento da outra série, por exemplo:  $A = (A_1, \dots, A_m)$  e  $B = (B_1, \dots, B_n)$ .

Nessa medida de distância, Inicialmente usa-se a distância euclidiana entre os pares de pontos das séries (distância de um para todos e todos para um) e constrói uma matriz  $N \times M$ . Utilizando essa matriz, precisamos encontrar o caminho da distância mínima, no qual para isso é construído outra matriz contendo as distâncias mínimas (custos) para chegar a um ponto específico. Para isso algumas restrições são necessárias:

1. O caminho deve iniciar do ponto (1,1) e terminar em (N, M);
2. Não podemos voltar no tempo, a partir do ponto (i, j) nós podemos ir para frente (i+1, j), para cima (i, j+1) ou na diagonal (i+1, j+1). O caminho mais curto para os pontos adjacentes, como ilustra a Figura 3.4, no qual o ponto (1,1) começa do canto inferior esquerdo da matriz.

Figura 3.4: Caminho a seguir usando DTW.



Atendendo a essas restrições, serão obtidos os custos e com isso calculamos o caminho mínimo, no qual a ideia é retroceder do último ponto (N, M) até (1, 1) minimizando o custo, ou seja:

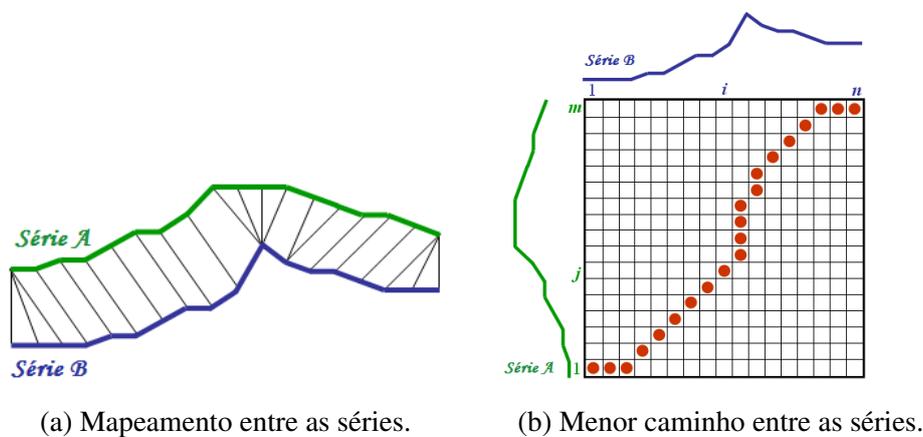
$$D(i, j) = \min\{C(i-1, j-1), C(i-1, j), C(i, j-1)\} + \text{distance}(i, j) \quad (3.3)$$

Onde:

- $C(i, j)$ : é o custo acumulado entre  $i$  e  $j$ ;
- $\text{distance}(i, j)$ : distância euclidiana dos pontos  $i$  e  $j$ .

Na Figura 3.5 temos uma noção intuitiva do DTW, onde na imagem da esquerda temos um exemplo de mapeamento entre duas séries e na direita um exemplo do menor caminho (em vermelho) encontrado entre as séries.

Figura 3.5: Exemplo da medida de distância DTW.



Fonte: Adaptado de Tsiporkova (2016).

As medidas de distância que definem a semelhança entre pontos, são classificadas entre medidas de similaridade e medidas de dissimilaridade (OLIVEIRA, 2013), no qual a que usamos neste trabalho são medidas de dissimilaridade. A diferença é que:

- Na medida de similaridade quanto maior a semelhança entre os elementos, maior o valor da medida e da mesma forma quanto menor a semelhança, menor o valor da medida;

- Na medida de dissimilaridade ocorre o contrário, quanto maior a semelhança, menor o valor e quanto menor a semelhança, maior é o valor da medida.

A matriz de dissimilaridade possui as seguintes propriedades:

- Simetria:  $d(x, y) = d(y, x)$ , ou seja, distância entre x e y é a mesma entre y e x;
- Se  $d(x, y) \neq 0$  então  $x \neq y$ , ou seja, são elementos diferentes;
- Se  $x = y$ , então  $d(x, y) = 0$  (diagonal da matriz).

Em relação a essas três medidas de dissimilaridade aqui apresentadas, a distância Euclidiana obteve melhores resultados quando o objetivo foi encontrar perfis de consumo semelhantes. Alguns experimentos foram feitos até chegar a essa conclusão, utilizando o *k-means* para agrupar os perfis, no qual eles estão apresentados no capítulo 5 deste trabalho.

### 3.3 Métricas de validação

A fim de avaliar a qualidade dos grupos obtidos e comparar os resultados de alguns agrupamentos, duas métricas foram utilizadas: o índice de silhueta (ROUSSEEUW, 1987) e o Erro Médio Absoluto (do inglês, *Mean Absolute Error* - MAE).

O índice de silhueta é uma métrica não supervisionada para validação de agrupamentos, que verifica a coesão (densidade dos grupos) e separação (distância inter *cluster*). O valor da silhueta pode variar entre -1 e 1, em que os valores próximos a 1 indicam que o elemento está bem adaptado ao seu grupo e diferente dos grupos vizinhos, 0 (ou próximo de 0) indica que o elemento está indefinido entre os grupos e próximo a -1 indica que o elemento pode estar no grupo errado. O valor da silhueta é maior quando os grupos são densos e separados. Esses valores são obtidos por meio da Equação 3.4.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3.4)$$

Também pode ser representado da seguinte forma (ROUSSEEUW, 1987):

$$s(i) = \begin{cases} 1 - a(i)/b(i), & \text{se } a(i) < b(i) \\ 0, & \text{se } a(i) = b(i) \\ b(i)/a(i) - 1, & \text{se } a(i) > b(i) \end{cases} \quad (3.5)$$

Onde:

- $a(i)$ : é a dissimilaridade média do elemento  $i$  em relação a todos os outros elementos do grupo (*cluster*) que contém  $i$ .
- $b(i)$ : é a dissimilaridade média do elemento  $i$  em relação a todos os elementos do grupo mais próximo (vizinho de  $i$ , ao qual ele não pertence).
- $a(i) < b(i)$ : se a dissimilaridade média dentro do grupo que  $i$  pertence é menor que a dissimilaridade entre outros grupos, então o elemento  $i$  está no grupo correto.
- $a(i) > b(i)$ : se a dissimilaridade média dentro do grupo que  $i$  pertence é maior que a dissimilaridade entre outros grupos, então o elemento  $i$  está no grupo errado. Provavelmente ele deveria estar no grupo vizinho mais próximo,  $b(i)$ .

Se um *cluster* contém apenas um elemento, a silhueta é zero ( $s(i) = 0$ ). O cálculo é feito para cada elemento, em que a medida da qualidade geral do agrupamento é a média aritmética dos índices de silhueta (*average silhouette width*), conforme a Equação 3.6.

$$s(i) = \frac{1}{n} \sum_{i=1}^n \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3.6)$$

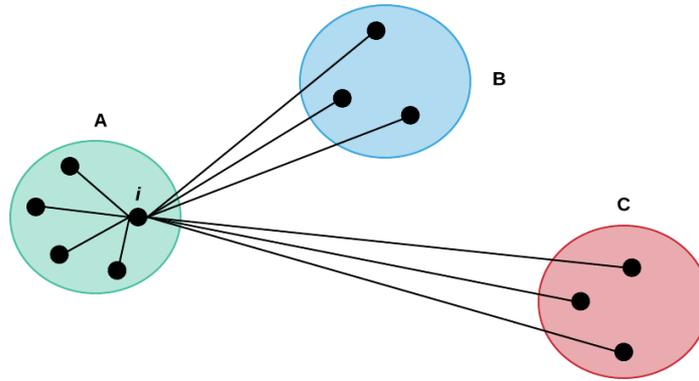
A Figura 3.6 ilustra um exemplo de como é representado o cálculo da silhueta. Nesse caso o  $a(i)$  é a média da distância entre  $i$  e os elementos do grupo A, o  $b(i)$  é a média das distâncias entre  $i$  e os elementos do vizinho mais próximo, que é o grupo B.

É necessário calcular a média para todos os grupos para saber qual o vizinho mais próximo de  $i$  (menor distância). Com isso temos:

$$b(i) = \min d(i, G) \quad (3.7)$$

Onde:

Figura 3.6: Ilustração dos elementos envolvidos no cálculo da silhueta.



Fonte: Adaptado de Rousseeuw (1987).

- $d(i, G)$ : distância média entre  $i$  e um grupo  $G$  qualquer. No qual,  $G$  é diferente do grupo que  $i$  pertence,  $G \neq G_i$ .

Arbelaitz et al. (2013), fez um estudo comparativo avaliando 30 índices de validação de agrupamentos em diferentes ambientes com características diferentes, como por exemplo, dimensionalidade dos dados e quantidade de grupos, além de utilizar 3 algoritmos de agrupamento: *k-means*, *Ward* e *Average-linkage*. Dos 30 índices avaliados, nenhum se mostrou superior aos demais em todos experimentos feitos, porém, o índice de silhueta obteve os melhores resultados na maioria dos testes, demonstrando ser uma métrica apropriada para validação de agrupamentos.

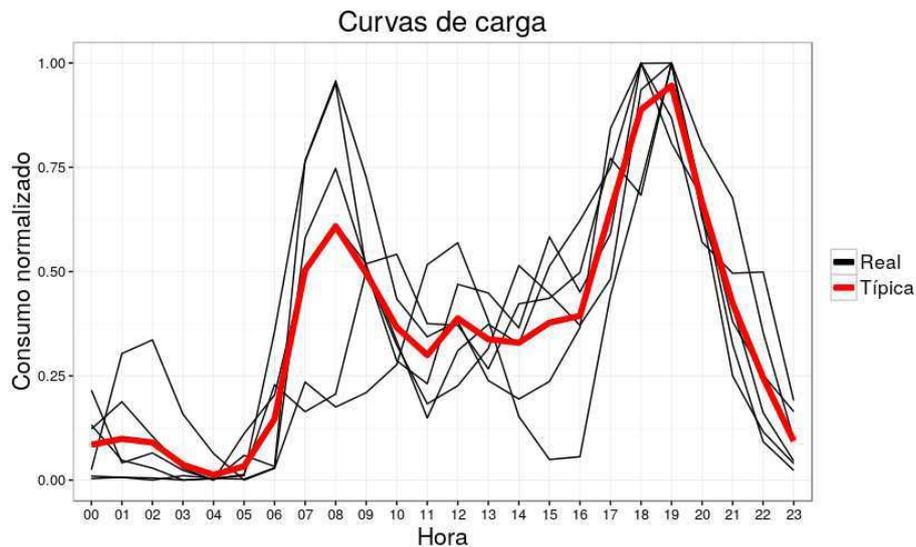
O erro médio absoluto (MAE) foi utilizado para as curvas de carga, no qual o erro foi calculado observando as curvas de carga típicas e as curvas de carga reais. Na Figura 3.7 temos uma noção intuitiva da curva típica (vermelha) e as reais (pretas), no qual a curva típica é a média entre os perfis de consumo que fazem parte do mesmo grupo e as demais curvas são os dados reais dos consumidores. A curva típica representa o comportamento do uso de eletricidade ao longo do tempo para o grupo formado. O cálculo do MAE pode ser observado na Equação 3.8 (KIM; KO; CHOI, 2011).

$$MAE = \frac{\sum_{i=1}^k \sum_{j=1}^{r_i} \frac{\sum_{t=1}^{obs} |x_t^{tp_i} - x_t^{rl_{pj}}|}{obs}}{\sum_{i=1}^k r_i} \quad (3.8)$$

Onde:

- $k$ : quantidade de *clusters*;
- $r_i$ : quantidade de residências (perfis) no  $i^{th}$  *cluster*;
- $obs$ : quantidade de observações;
- $tlp_i$ : curva de carga típica (*typical load profile*) do  $i^{th}$  *cluster*;
- $x_t^{tlp_i}$ : observação no tempo  $t$  da curva típica de carga do  $i^{th}$  *cluster*;
- $rlp_{ij}$ : curva de carga real (*real load profile*) da  $j^{th}$  residência pertencente ao  $i^{th}$  *cluster*.

Figura 3.7: Curvas de carga típica e reais.



O MAE foi utilizado porque a curva de carga típica é a que representa todo o *cluster*. Se o *cluster* for bem compacto, a curva típica irá ficar bem próxima das curvas reais e terá um erro menor do que em um *cluster* disperso.

# Capítulo 4

## Materiais e métodos

Neste capítulo são apresentados detalhes da construção da base de dados, coletando informações por meio de um questionário e utilizando uma base de dados sintética com curvas de carga residenciais. Também são abordados os cenários propostos e configurações de grupos utilizando alguns fatores para gerar os agrupamentos com o algoritmo *k-means*, assim como o método utilizado para encontrar a medida de dissimilaridade que representa a semelhança entre as curvas de carga. Os experimentos foram feitos usando a linguagem de programação R, no qual todos os dados e informações necessárias para replicar os resultados estão no GitHub<sup>1</sup>.

### 4.1 Construção da base de dados

Uma das dificuldades encontradas neste trabalho foi a inexistência de uma base de dados com consumidores residenciais para avaliar a solução proposta. Como discutido no Capítulo 1, no Brasil não há medidores inteligentes em escala residencial que possam fornecer informações detalhadas sobre o consumo elétrico das residências e por isso se torna necessário a utilização de outros meios para conseguir essas informações. Atualmente a forma de obter alguma informação sobre o consumo de energia é por meio do consumo mensal obtido pela leitura do medidor feito por um agente da concessionária de energia elétrica, porém essa informação não é fornecida pela concessionária por questões de privacidade.

Devido à essa dificuldade, foram utilizadas informações de relatórios de Pesquisas de

---

<sup>1</sup><https://github.com/Harllanandrye/EnergyEfficiency-clustering->

Posse de equipamentos e Hábitos de uso (PPHs) realizados pela Eletrobrás e PROCEL (ELETROBRAS, 2007), para gerar uma população sintética de consumidores. Para isso, foi aplicado um questionário com o objetivo de obter informações de consumidores residenciais, utilizando como base as questões do questionário aplicado pela Eletrobrás/PROCEL nas PPHs. A participação das pessoas para responder o questionário foi voluntária e anônima, ou seja, não inclui nenhum tipo de informação pessoal que possa identificá-los.

No questionário foram introduzidas questões relativas às condições socioeconômicas dos consumidores, os tipos de equipamentos utilizados nas residências e outras características que pudessem representar o consumidor, tais como: localização geográfica, valor e consumo médio mensal de energia, renda média da família, adota medidas para reduzir o consumo, forma de identificar o consumo dos equipamentos, ações para reduzir o consumo, entre outras características. O questionário aplicado para a construção da população sintética pode ser encontrado no Apêndice A desta dissertação.

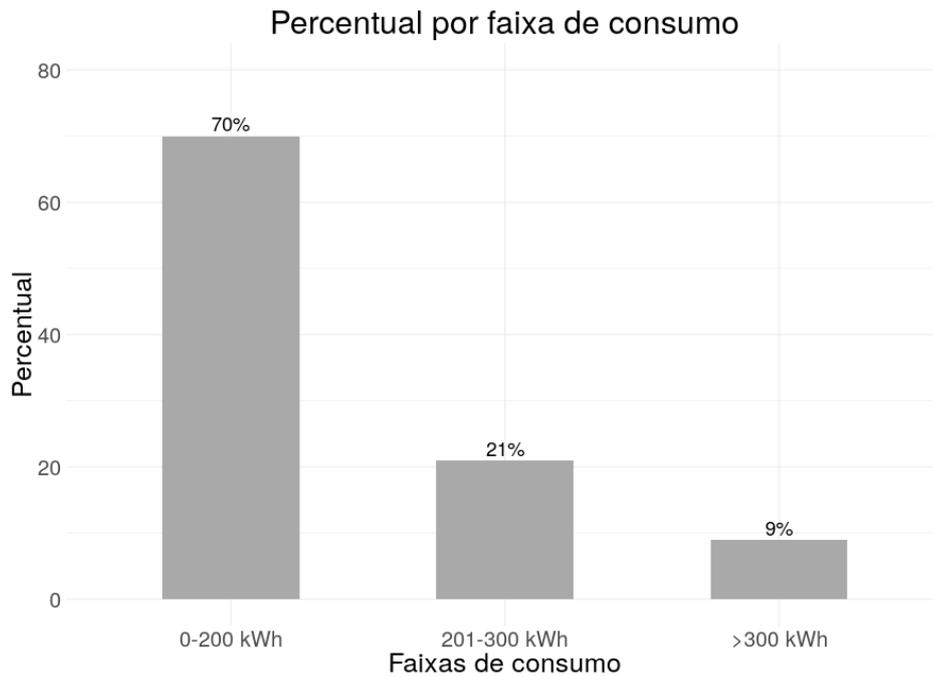
Entre essas informações coletadas, algumas parecem mais relevantes e, assim, foram utilizadas para a formação dos grupos de consumidores semelhantes, a saber: localização, consumo mensal, aparelhos existentes na residência e ações para reduzir o consumo. A localização geográfica e consumo permitem agrupar consumidores de uma mesma região com consumos parecidos.

Para facilitar as análises foram criadas algumas faixas de consumo mensal, no qual as residências que têm valores de consumo na mesma faixa, são semelhantes. Podemos considerar 3 faixas de consumo de acordo com Anuário Estatístico de Energia Elétrica 2016 (ENERGÉTICA, 2016), feito pela Empresa de Pesquisa Energética (EPE) com o apoio do Ministério de Minas e Energia, que são: 0-200 kWh, 201-300 kWh e maior que 300 kWh. Na Figura 4.1 temos a distribuição de consumidores, da nossa base de dados, por faixa de consumo.

Existe a faixa de consumo entre 0 e 100 kWh que não foi considerada, pois apenas 3 consumidores da população utilizada no trabalho tem consumo médio mensal inferior à 100 kWh, o que torna a amostra muito pequena, por isso, resolvemos agrupar essa faixa com a subsequente, resultando na faixa 0-200 kWh.

Os dados referentes à posse de aparelhos informam quais aparelhos estão presentes na residência, bem como suas quantidades, no qual foram considerados 9 aparelhos na pesquisa

Figura 4.1: Percentual de consumidores por faixa de consumo.



por apresentarem uma grande participação no consumo final de energia elétrica (ELETROBRAS, 2007), que são: Chuveiro elétrico, Ar-condicionado, Máquina de lavar roupas, Microondas, Ventilador, Forno elétrico, Lava louça, Ferro elétrico e Geladeira. Na Figura 4.2 temos o percentual de aparelhos existentes nas residências para cada um dos 9 selecionados na realização da pesquisa.

Em relação às ações para reduzir o consumo de energia, foram definidas algumas classes (ou ações) para as recomendações, no qual foram colocadas no questionário para a pessoa (consumidor) que está respondendo informar qual delas estaria disposto a seguir. As classes estão na Tabela 4.1, no qual as abreviações definidas para cada classe serão utilizadas no algoritmo adaptativo, descrito na seção 4.3.

As respostas da pesquisa em relação às ações que as pessoas estariam dispostos a realizar para reduzir o consumo de energia estão ilustradas na Figura 4.3, com o valor percentual de cada resposta.

Além dessas informações obtidas do questionário, foi utilizada uma base de dados sintética de perfis de consumo de energia elétrica, pois não temos medidores inteligentes para obter informações reais das residências. O perfil de consumo remete para o histórico do consumo elétrico ao longo do tempo (por exemplo, consumo a cada hora) e é um dos fatores

Figura 4.2: Percentual de aparelhos existentes nas residências.

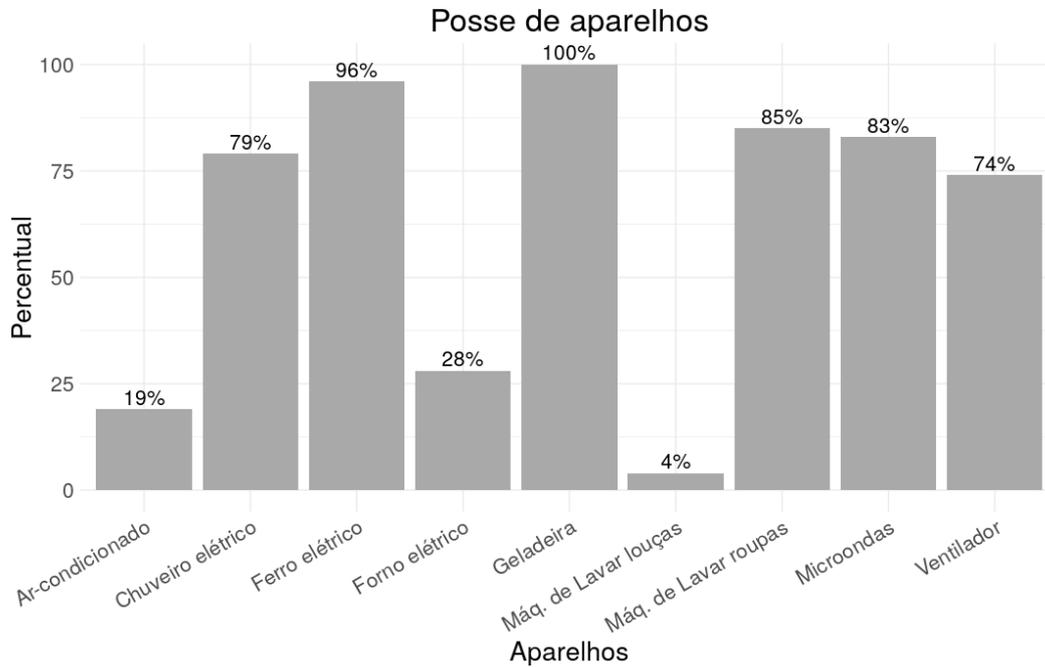


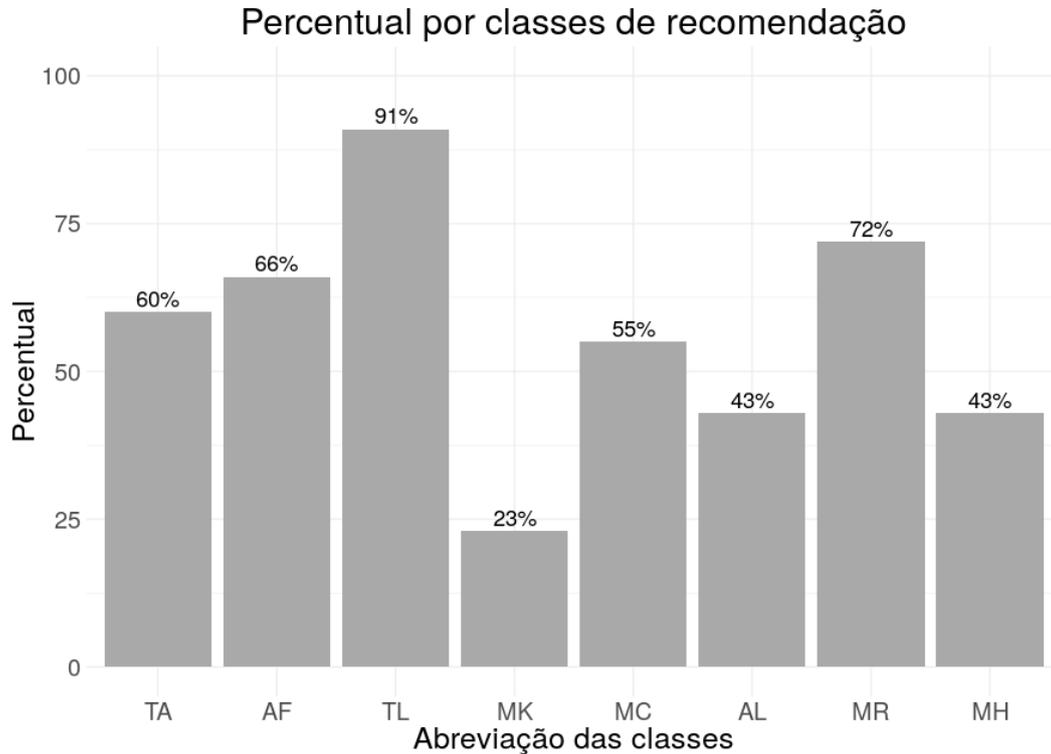
Tabela 4.1: Classes das recomendações.

Classe	Descrição	Abreviação
Trocar aparelhos	Trocar aparelhos elétricos antigos por mais novos e eficientes.	TA
Aprender sobre aparelhos	Aprender funções e funcionamento de seus aparelhos elétricos.	AF
Trocar lâmpadas	Trocar lâmpadas por modelos mais eficientes.	TL
Mudanças na casa	Realizar mudanças no interior da casa (cor da parede, posições de aparelhos, ...).	MK
Mudanças no conforto	Mudanças de hábitos que alteram o seu conforto (como diminuir temperatura do chuveiro elétrico).	MC
Atividades de limpeza	Adicionar atividades de limpeza específicas de alguns aparelhos elétricos.	AL
Mudanças na rotina	Mudar rotina de uso de alguns aparelhos (como acumular roupas sujas para lavar uma única vez).	MR
Mudanças de horário	Mudar horários de algumas atividades da sua rotina.	MH

utilizados neste trabalho para a formação de grupos de consumidores semelhantes.

Esses perfis foram gerados utilizando um gerador de carga que tem como entrada dados de consumo de aparelhos e sua saída é a curva de carga agregada de uma residência, como pode ser observado na Figura 4.4. Esses perfis sintéticos foram gerados antes de ser aplicado o questionário, portanto os aparelhos que foram utilizados na entrada do gerador não são os mesmos do questionário. Embora os aparelhos obtidos por meio do questionário não reflitam

Figura 4.3: Percentual de consumidores por classes de recomendação.

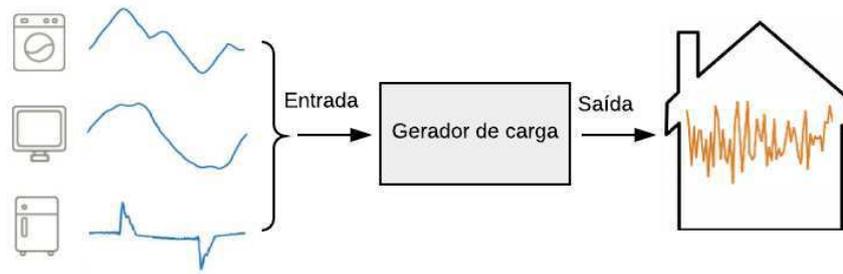


completamente os dados das PPHs (ELETROBRAS, 2007), a carga gerada sinteticamente corresponde à esses dados por serem mais realistas.

Essa base de dados com perfis sintéticos foi gerada utilizando informações sobre posse de aparelhos presentes nas PPHs e os aparelhos utilizados foram: lâmpada, geladeira, televisor, ferro elétrico, ar-condicionado e chuveiro elétrico. Nas PPHs há algumas tabelas com a quantidade de aparelhos de acordo com as faixas de consumo, por exemplo: residências com consumo entre 0-200kWh tem em média menos do que 1 ferro elétrico. Com base nisso foram escolhidas as quantidades de aparelhos em cada perfil, por exemplo: algumas residências da 1ª faixa de consumo não possuem ferro elétrico e outras possuem apenas 1. A quantidade escolhida de aparelhos em cada perfil de consumo gerado foi aleatória.

Para cada residência ter um perfil de consumo que possa ser próximo da realidade, combinamos os dados do questionário com esses perfis da seguinte forma: Com o valor do consumo em kWh fornecido pelo consumidor, comparamos esse valor com todos os consumos em kWh dos 148 perfis da base de dados. Se o valor de consumo obtido do questionário for igual ao valor de consumo do perfil, então atribuímos aquele perfil a residência. Após esse mapeamento, cada residência passou a ter um perfil de consumo. Tomou-se o cuidado de não

Figura 4.4: Funcionamento do gerador de carga.



permitir que duas residências tivessem o mesmo perfil de consumo, já que cada residência foi comparada com cada perfil da base.

Existem outras bases de dados disponíveis publicamente com perfis de consumo residenciais (BATRA et al., 2014), porém a quantidade de residências nessas bases é pequena e isso influenciaria no mapeamento entre as curvas de carga e os dados do questionário, diminuindo a nossa base de dados. Por isso foram utilizados os dados sintéticos.

Em resumo, concluímos a formação da base de dados, composta por 47 residências, no qual os dados foram obtidos por meio de um questionário, aplicado na cidade de Campina Grande - PB, e de dados de perfis de consumo sintéticos, assim temos dados socioeconômicos, posse de aparelhos e como é a utilização da energia ao longo do tempo. Vale destacar a dificuldade da realização do processo de coleta de dados, pois para o seu sucesso havia a dependência da disposição voluntária do respondente em fornecer seus dados sobre posse de aparelhos e hábitos de consumo.

## 4.2 Cenários propostos e configurações de grupos

Para descobrir quais são os fatores (ou combinação deles) mais representativos para criar grupos com consumidores semelhantes, foram criados 15 cenários de acordo com os 4 fatores propostos para o agrupamento. Os fatores utilizados e os cenários criados estão descritos a seguir.

### Fatores:

- Características diversas: localização geográfica e valor do consumo;

- *Feedback*: relacionado as recomendações passadas para o consumidor;
- NIALM: informação dos 9 aparelhos;
- Perfil: curvas de carga sintéticas mapeadas com os dados do questionário.

**Cenários:**

1. Características diversas;
2. Características diversas + *Feedback* (Adaptativo);
3. Características diversas + *Feedback* + NIALM (Adaptativo);
4. Características diversas + *Feedback* + Perfil (Adaptativo);
5. Características diversas + *Feedback* + NIALM + Perfil (Adaptativo);
6. Características diversas + NIALM;
7. Características diversas + Perfil;
8. Características diversas + NIALM + Perfil;
9. *Feedback* (Adaptativo);
10. *Feedback* + NIALM (Adaptativo);
11. *Feedback* + Perfil (Adaptativo);
12. *Feedback* + NIALM + Perfil (Adaptativo);
13. NIALM;
14. NIALM + Perfil;
15. Perfil.

Os cenários onde aparece “(Adaptativo)” são aqueles que irão utilizar o algoritmo de agrupamento adaptativo (subseção 4.3.2) devido ao uso do *feedback* do consumidor. Os outros não precisam de adaptação, pois as características e seus pesos não mudam, como acontece no *feedback*.

Além dos cenários, foram definidas algumas configurações de grupos, que estão relacionadas a quantidade de grupos formados em cada agrupamento. Sabemos que a quantidade de grupos influencia no resultado do agrupamento e por isso foi escolhido utilizar de 3 a 8 grupos, para todos os 15 cenários, totalizando 90 experimentos (15 cenários x 6 configurações de grupos). Essa quantidade de grupos é devido ao tamanho da base de dados, pois aumentando a quantidade podemos ter grupos com poucas residências ou até mesmo apenas 1 residência sendo o próprio grupo.

As etapas do agrupamento com os cenários e as configurações de grupos são:

1. Escolhe o cenário de agrupamento;
2. Escolhe a quantidade de grupos que serão formados, entre 3 e 8;
3. Pré-processa os dados, como explicado na seção 4.3;
4. Cria os grupos de consumidores, utilizando o algoritmo de agrupamento;
5. Obtem os resultados;
6. Retorna ao passo 1, escolhendo um novo cenário. Repete esse processo para todos os 15 cenários e para todas as configurações de grupos, 3 a 8.

Por meio dos resultados dos experimentos foram escolhidos os fatores mais relevantes, no qual essa relevância é em relação à quantidade de vezes que determinado cenário aparece nas diferentes configurações de grupos e a dificuldade de obter determinado fator (informação), por exemplo, as características diversas são informações fáceis de obter observando apenas a conta de energia, diferente do perfil de consumo que é necessário um equipamento para capturar as informações da rede elétrica. Essa relevância dos fatores está detalhada no capítulo 5.

Um outro experimento foi feito para avaliar o uso da métrica silhueta dentro do contexto abordado com base nas respostas de alguns especialistas da área de eficiência energética, porém os resultados foram inconclusivos e esse experimento está detalhado no Apêndice B.

### 4.3 Agrupamento

Os agrupamentos criados nos experimentos feitos para essa pesquisa, utilizaram 4 fatores extraídos da nossa base de dados, que foram: Características diversas, *Feedback*, NIALM e Perfil. Esses fatores estão detalhados na seção 4.2. Essas informações serviram como entrada para o algoritmo *k-means*, mas antes de passar os dados para o algoritmo foi necessário normalizar os dados para uma padronização (tarefa de pré-processamento), pois isso minimiza os problemas oriundos do uso de unidades e dispersões distintas entre as variáveis (RODRIGUES et al., 2015).

A normalização evita que uma dimensão se sobreponha em relação às outras. A técnica de normalização utilizada neste trabalho foi a normalização segundo a amplitude (Equação 4.1), que utiliza os valores máximo e mínimo das variáveis para normalizar linearmente os dados entre 0 e 1.

$$valor_{normalizado} = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (4.1)$$

onde  $i$  é um valor que será normalizado da variável  $x$ ,  $\min(x)$  e  $\max(x)$  são os valores mínimos e máximos, respectivamente, da variável em questão. Na Figura 4.5 temos um exemplo dessa normalização, no qual temos os dados originais da nossa base de dados (esquerda) e os dados normalizados (direita), referente às características diversas. Para a localização não transformamos as coordenadas geográficas para cartesianas, apenas normalizamos os dados de acordo com a Equação 4.1, mas o impacto disso nos resultados é muito pequeno pois o agrupamento considera moradores apenas de uma mesma cidade.

Figura 4.5: Exemplo dos dados normalizados.

	Latitude ↕	Longitude ↕	Consumo ↕
1	-7.233777	-35.92631	170
2	-7.236530	-35.90415	207
4	-7.250077	-35.89750	112
5	-7.205927	-35.87462	323
6	-7.226925	-35.90636	271
7	-7.252768	-35.90212	105
8	-7.231723	-35.87795	120
9	-7.236049	-35.90828	142

(a) Dados originais.

	Latitude ↕	Longitude ↕	Consumo ↕
1	0.6202670	0.17865652	0.23497268
2	0.5827300	0.47862403	0.33606557
4	0.3980175	0.56869010	0.07650273
5	1.0000000	0.87843180	0.65300546
6	0.7136936	0.44866519	0.51092896
7	0.3613259	0.50617317	0.05737705
8	0.6482731	0.83333785	0.09836066
9	0.5892884	0.42265934	0.15846995

(b) Dados normalizados.

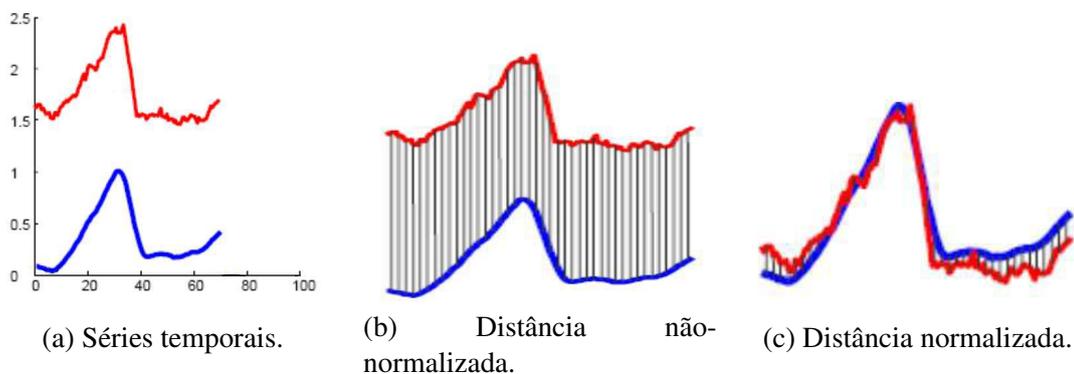
### 4.3.1 Perfil de consumo

Algumas das dificuldades ou incertezas enfrentadas na tarefa de agrupamento (*clustering*) envolvem a seleção e ajuste de critérios de agrupamento, algoritmos, número inicial de *clusters*, as características mais adequadas, medidas de dissimilaridade, técnicas de validação, etc (IGLESIAS; KASTNER, 2013). Entre elas, encontrar uma medida de dissimilaridade (semelhança) entre séries temporais não é uma tarefa trivial, pois depende do contexto e dos dados utilizados.

Pensando nisso, foram feitos alguns experimentos para descobrir qual a melhor medida de dissimilaridade (Correlação, Euclidiana ou DTW) para representar a semelhança entre duas séries temporais, no nosso caso, perfis de consumo de energia elétrica. O cálculo da dissimilaridade entre as séries é uma etapa do pré-processamento dos dados, no qual tem como resultado uma matriz de distância que será passada para o algoritmo de agrupamento.

Antes de calcular a dissimilaridade foi necessário normalizar os dados para que os perfis representassem apenas a demanda e não sua grandeza (CARVALHO; NETO, 2010), pois o objetivo deste trabalho é criar grupos com curvas de carga parecidas e não demandas, por exemplo: se duas residências que possuem perfis de consumo parecidos, porém em níveis de demanda diferentes, fossem analisadas sem a normalização, certamente não seriam sujeitas de comparação já que haveria uma distância, no valor de consumo, entre elas. A normalização foi feita segundo a amplitude (Equação 4.1) e na Figura 4.6 temos uma intuição visual da necessidade da normalização das séries.

Figura 4.6: Necessidade da normalização dos perfis de consumo.



Fonte: Adaptado de (ALENCAR, 2007).

Após normalizar os perfis, alguns experimentos foram feitos para comparar as três medidas de dissimilaridade, criando agrupamentos variando a quantidade de observações das séries temporais (perfis) e a quantidade de curvas de carga. Os perfis de consumo da base de dados contém informações de 1 mês de cada residência, com observações a cada 1 hora. Com isso variamos a quantidade de observações da seguinte forma: 31 observações referente à 1 mês (consumo diário), 7 observações (1 semana) e 24 observações referente à 1 dia (hora a hora).

Além disso, em cada quantidade de observações também variamos a quantidade de curvas de carga (de 10 em 10) em cada agrupamento, entre 10 e 148 da nossa base de dados sintética, escolhidas aleatoriamente para cada execução do algoritmo. Foram feitos 30 experimentos, variando a quantidade de curvas em cada quantidade de observações, resultando em 30 valores médios para cada métrica de validação. Como em cada agrupamento formado (10-148) temos o valor de cada métrica, totalizando 15 valores numéricos, foram utilizadas as médias desses resultados para cada um dos 30 experimentos. Esses experimentos foram feitos para saber se a quantidade de observações e curvas de carga influenciam na escolha da melhor medida de dissimilaridade para representar a semelhança entre as séries.

Outra base de dados também foi utilizada para comparar as medidas de distâncias, chamada REFIT (MURRAY et al., 2015), que contém dados de 1 mês de consumo de 20 residências do Reino Unido (do inglês, *United Kingdom* - UK). Essa base foi escolhida por conter dados de perfis de consumo com observações menores que 1 hora, diferente da nossa base sintética. Isso é importante para analisar se com uma menor granularidade, por exemplo a cada 10 minutos, os resultados são semelhantes quando tenho observações a cada hora.

Para verificar os resultados desses experimentos foi utilizada a métrica de validação de *cluster* chamada Silhueta (do inglês, *Silhouette Width*), calculado o erro médio absoluto do agrupamento e também foi feita uma análise visual dos grupos formados. As métricas de validação estão detalhadas no capítulo 3.

### 4.3.2 Algoritmo de agrupamento adaptativo

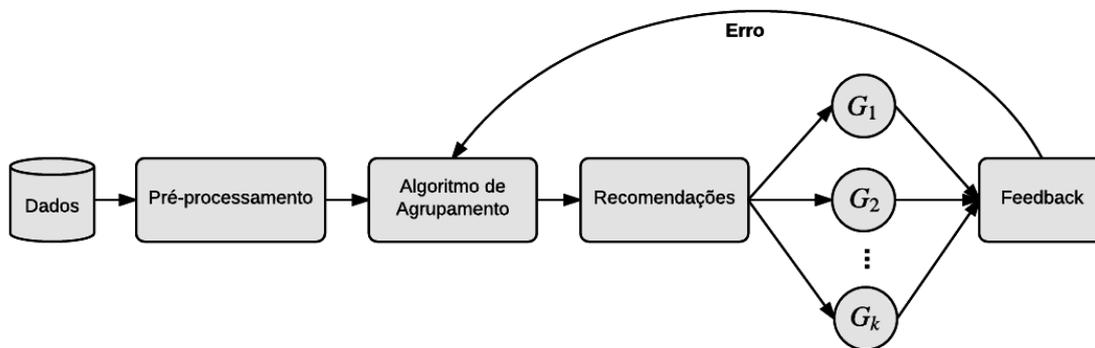
Como mencionado anteriormente, um dos fatores que foram utilizados para criar grupos de consumidores semelhantes foi o *feedback* em relação às recomendações passadas para eles. Essas recomendações estão relacionadas as ações que os consumidores estão dispostos a rea-

lizar para reduzir seu consumo de energia e com o *feedback* criamos grupos de consumidores com interesses comuns, que são aqueles consumidores que sempre ou geralmente dão seu *feedback* dizendo se as recomendações foram boas (gostou) ou não.

Como o comportamento do consumidor pode mudar ao longo do tempo e alguma recomendação pode não ser boa para determinado consumidor, foi utilizado um algoritmo de agrupamento adaptativo (RIBEIRO, 2016) que irá absorver essas mudanças no comportamento do consumidor de acordo com o *feedback* e as classes de recomendação.

Esse algoritmo também utiliza o *k-means* para criar os grupos, no qual a diferença está na atualização dos grupos a cada conjunto de recomendações passadas para eles, seguindo a ideia de uma rede neural com *backpropagation*, que a cada repetição do algoritmo as conexões entre os neurônios vão sendo evoluídas e alteradas com o intuito de minimizar o erro de aprendizagem (RUSSELL; NORVIG, 2009). Na Figura 4.7 podemos observar os passos para a execução do algoritmo adaptativo.

Figura 4.7: Modelo do algoritmo adaptativo.



No pré-processamento os dados coletados são organizados em uma tabela que contém o identificador do consumidor e um vetor de características (localização geográfica, *feedback*, ...). Esses dados são normalizados para ficarem na mesma escala e são atribuídos pesos iguais, conforme a Equação 4.2, para as características na primeira execução do algoritmo; logo após são criados os grupos com o *k-means*. Com os grupos criados, são passadas recomendações para cada um deles e cada consumidor dos grupos dá seu *feedback* a respeito das recomendações recebidas.

$$Peso = \frac{1}{N^{\circ} \text{ de características}} \quad (4.2)$$

Após o *feedback* é calculado o erro e atribuído novos pesos às características. O valor do erro é calculado observando quantas recomendações o consumidor marcou como útil de todas as recomendações passadas para ele, ou seja, o erro é a diferença entre o *feedback* esperado e o recebido. O cálculo pode ser observado na Equação 4.3, no qual o *feedback* esperado será sempre 1. Com os novos pesos o algoritmo é executado novamente para criar novos grupos ou melhorar os existentes, no qual o esperado é que o algoritmo convirja obtendo um erro próximo de zero.

$$Erro = 1 - \frac{n^{\circ} \text{ de rec. com feedback positivo}}{\text{total de recomendações passadas}} \quad (4.3)$$

Para encontrar e atribuir os novos pesos durante cada execução do algoritmo, foi utilizado o método de descida do gradiente (Equação 4.4), seguindo a mesma abordagem de uma rede neural. Considerando uma função de erro  $E[\vec{w}]$  com um vetor de pesos  $\vec{w}$ , queremos encontrar um vetor de pesos que minimiza  $E$ , até obter o menor erro possível. A função de erro é o erro médio calculado para cada consumidor, com base na Equação 4.3.

$$\nabla E[\vec{w}] = \left[ \frac{\partial E}{\partial w_0}, \frac{\partial E}{\partial w_1}, \dots, \frac{\partial E}{\partial w_n} \right] \quad (4.4)$$

Usando esse método atribuímos os novos pesos às características, como pode ser observado na Equação 4.5. A taxa de aprendizagem  $\eta$  determina o quanto  $E[\vec{w}]$  influencia nos pesos, no qual o valor máximo determinado para  $\eta$  foi de 0,5. Durante a execução do algoritmo esse valor pode diminuir 0,1 quando o erro médio aumenta, ou incrementar 0,1 quando ocorre o contrário; se o erro não sofre muita alteração, a taxa de aprendizagem permanece a mesma.

$$\vec{w} = \vec{w} - \eta \nabla E[\vec{w}] \quad (4.5)$$

Como não temos um sistema real para passar recomendações para os consumidores e obter as respostas deles, foram utilizadas as informações do questionário a respeito das ações que os consumidores estão dispostos a realizar e com isso simulamos as respostas deles.

Foram definidas algumas classes para as recomendações (Tabela 4.1), possibilitando o mapeamento entre as recomendações passadas e as classes que o consumidor está disposto a seguir, de acordo com as informações fornecidas no questionário.

Esse *feedback* simulado é feito automaticamente, comparando as classes das recomendações com as ações dos consumidores. Na Tabela 4.2 temos alguns exemplos de recomendações e as classes a qual fazem parte, similar ao arquivo de entrada do algoritmo para o *feedback* automático. Quando uma recomendação faz parte de uma classe é atribuído o valor 1, caso contrário, 0.

Tabela 4.2: Recomendações e suas classes.

Tipo	Descrição	TA	AF	TL	MK	MC	AL	MR	MH
geld	Evite que a geladeira fique exposta ao sol.	0	1	0	1	0	0	0	0
chuv	No verão mudar a chave do chuveiro para essa estação.	0	1	0	0	1	0	1	0
lamp	Trocar lâmpadas por fluorescentes.	1	1	1	0	0	0	0	0
mw	Evite o microondas para aquecer líquidos, prefira o fogão.	0	1	0	0	0	0	0	0

Além do arquivo com as recomendações, o algoritmo recebe como entrada os dados dos consumidores, coletados por meio do questionário. Na Tabela 4.3 temos um exemplo, no qual o valor 1 indica que o consumidor está disposto a realizar aquela ação para economizar energia.

Tabela 4.3: Dados obtidos do questionário.

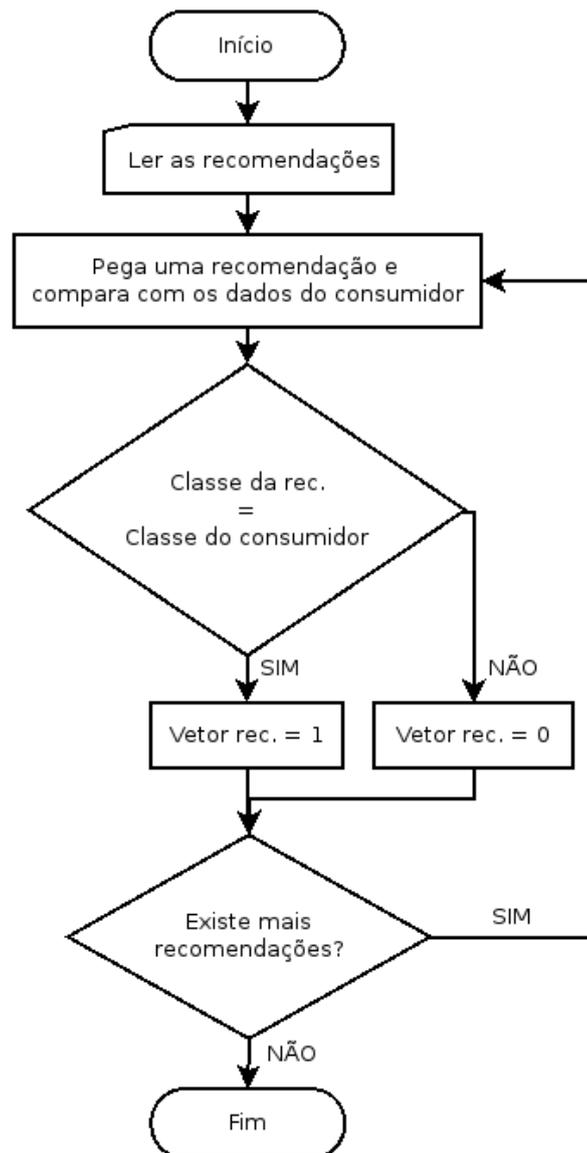
Consumidor	TA	AF	TL	MK	MC	AL	MR	MH
1	0	1	1	0	1	0	1	0
2	1	1	1	0	0	0	1	0
3	0	0	1	0	1	0	1	0
4	1	0	1	0	1	0	1	0

Cada consumidor possui um vetor de *feedback*, que é preenchido de acordo com as ações que eles estão dispostos a fazer, por exemplo: se for passada uma recomendação que envolve troca de aparelho e o consumidor estiver disposto a fazer isso, um campo do vetor será preenchido com valor igual a 1. Podemos representar esse exemplo assim: dado um

vetor de *feedback*  $\vec{v}_A = [geld, chuv, lamp, mw]$  para o consumidor A, no qual ele recebeu e aceitou uma recomendação que envolve troca de lâmpada (lamp), resultando no vetor  $\vec{v}_A = [0, 0, 1, 0]$ .

A execução do algoritmo do *feedback* automático (Figura 4.8) é feita para todos consumidores de cada grupo e os passos são:

1. Ler as recomendações passadas para o consumidor;
2. Pega uma recomendação das que foram passadas para ele;
3. Compara as classes dessa recomendação com as ações do consumidor (TA, AF, ...);
4. Se uma classe da recomendação for igual a do consumidor, então atribui valor 1 no vetor de *feedback*, informando que ele aceitou a recomendação;
  - (a) Caso contrário, atribui 0;
5. Passa para a próxima recomendação e volta ao passo 3, até não ter mais recomendações das que foram passadas para o consumidor.

Figura 4.8: Fluxograma do *feedback* automático.

# Capítulo 5

## Resultados e discussão

Neste capítulo são apresentados os resultados obtidos, aplicando a metodologia, descrita no capítulo anterior, na base de dados que contém informações de residências coletadas por meio de um questionário. Inicialmente são apresentados os diversos experimentos realizados com as três medidas de dissimilaridade para encontrar qual medida é ideal para representar a semelhança entre as curvas de carga. Os resultados mostram que a distância euclidiana representa melhor a semelhança entre as curvas do que a correlação e o DTW, motivando seu uso para o agrupamento dos perfis de consumo residenciais.

Também são abordados os resultados dos agrupamentos utilizando os 4 fatores extraídos da base de dados, no qual foram criados 15 cenários a partir da combinação desses fatores. Com os cenários propostos, foram feitas algumas análises com o objetivo de encontrar os cenários (ou fatores) mais relevantes. Além disso, é apresentada uma comparação entre a avaliação subjetiva do agrupamento e os valores da métrica silhueta, no qual a silhueta não se mostrou uma boa métrica para a validação dos grupos dentro do contexto desta dissertação.

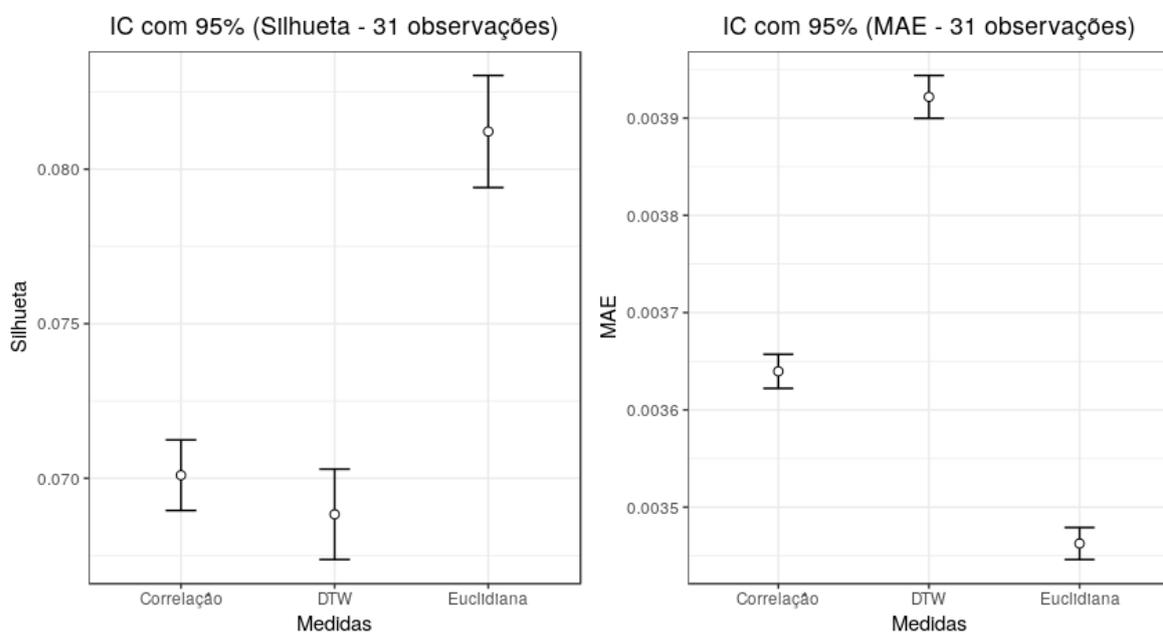
### 5.1 Similaridade entre os perfis de consumo

Encontrar uma medida de semelhança entre séries temporais não é uma tarefa trivial, pois depende do contexto e dos dados utilizados. Por isso foram feitos alguns experimentos para descobrir qual a melhor medida de dissimilaridade dentre Correlação, Euclidiana e DTW, para representar a semelhança entre perfis de consumo de energia elétrica.

A granularidade dos dados, ou seja, a quantidade de observações em uma série temporal,

pode influenciar ou não na escolha da medida de dissimilaridade. Pensando nisso, foram feitos agrupamentos variando a quantidade de observações das curvas de carga para saber se esse fator influencia na escolha da medida de dissimilaridade para representar a semelhança entre as curvas. Para os resultados desses agrupamentos foram calculados os intervalos de confiança (IC) de 95% para cada métrica de validação. Na Figura 5.1 temos os resultados das métricas Silhueta e MAE, variando a quantidade de curvas de carga com 31 observações (1 mês de dados, consumo diário).

Figura 5.1: Valores da silhueta e MAE com intervalo de confiança de 95% (31 observações).



Como a base de dados de perfis de consumo sintéticos possui 148 perfis, foram feitos agrupamentos a cada 10 perfis de consumo escolhidos aleatoriamente (explicado no capítulo anterior), para as três medidas de dissimilaridade. Na Tabela 5.1 temos as médias e os limites inferior e superior dos intervalos de confiança para os resultados da Figura 5.1, no qual a distância euclidiana foi a melhor, ou seja, obteve maior valor da silhueta e o menor erro médio absoluto (MAE – quanto menor esse valor, melhor é o resultado) entre as medidas de dissimilaridade utilizadas.

Outro experimento feito usando informações de consumo diário está ilustrado na Figura 5.2, com 7 observações referente a uma semana. O que pode ser observado é que a distância euclidiana continua com resultados melhores, no qual na Tabela 5.2 podemos verificar isso

Tabela 5.1: Limites do intervalo de confiança para a silhueta e MAE (31 observações).

	Silhueta			MAE		
	IC inferior	Média	IC superior	IC inferior	Média	IC superior
<b>Correlação</b>	0,069	0,07	0,071	0,00362	0,00364	0,00367
<b>DTW</b>	0,067	0,068	0,07	0,00389	0,00392	0,00394
<b>Euclidiana</b>	0,079	0,081	0,083	0,00344	0,00346	0,00348

observando os valores da média e dos limites dos intervalos de confiança das métricas.

Figura 5.2: Valores da silhueta e MAE com intervalo de confiança de 95% (7 observações).

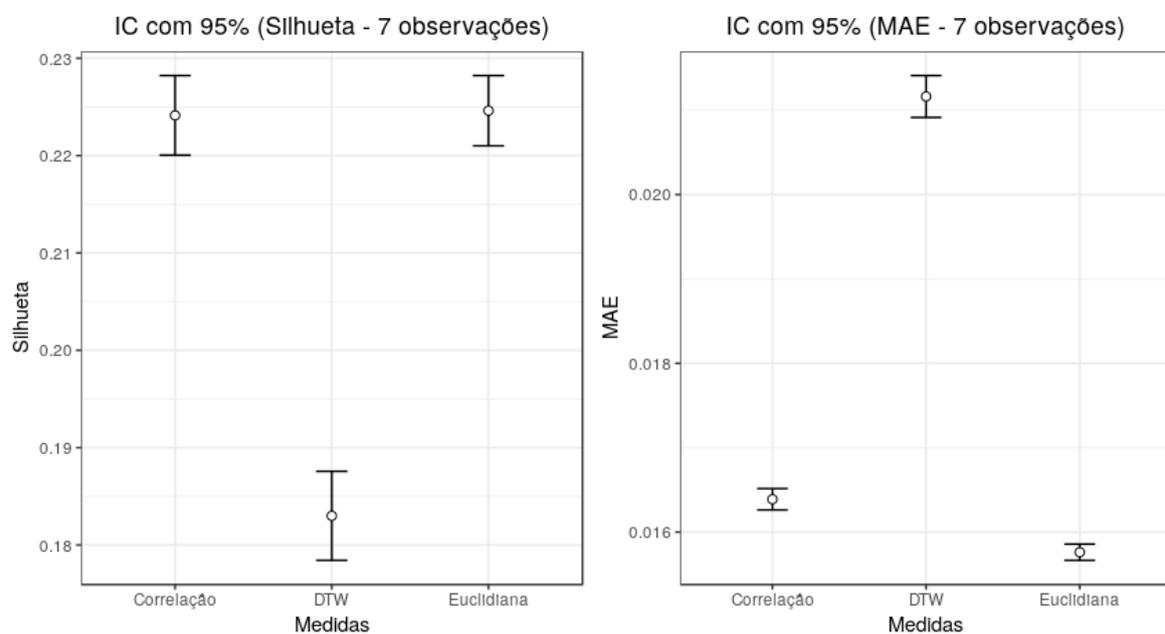


Tabela 5.2: Limites do intervalo de confiança para a silhueta e MAE (7 observações).

	Silhueta			MAE		
	IC inferior	Média	IC superior	IC inferior	Média	IC superior
<b>Correlação</b>	0,22	0,224	0,228	0,0162	0,0164	0,0166
<b>DTW</b>	0,178	0,183	0,187	0,0209	0,0212	0,0215
<b>Euclidiana</b>	0,221	0,225	0,228	0,0156	0,0157	0,0158

Aumentando a frequência de observações, a cada hora, temos os resultados na Figura 5.3 e na Tabela 5.3. Foram utilizados dados de 1 dia com observações a cada hora, resultando em 24 observações. Mesmo com dados a cada hora, a euclidiana se mantém como a melhor medida de dissimilaridade em relação à correlação e DTW.

Figura 5.3: Valores da silhueta e MAE com intervalo de confiança de 95% (24 observações).

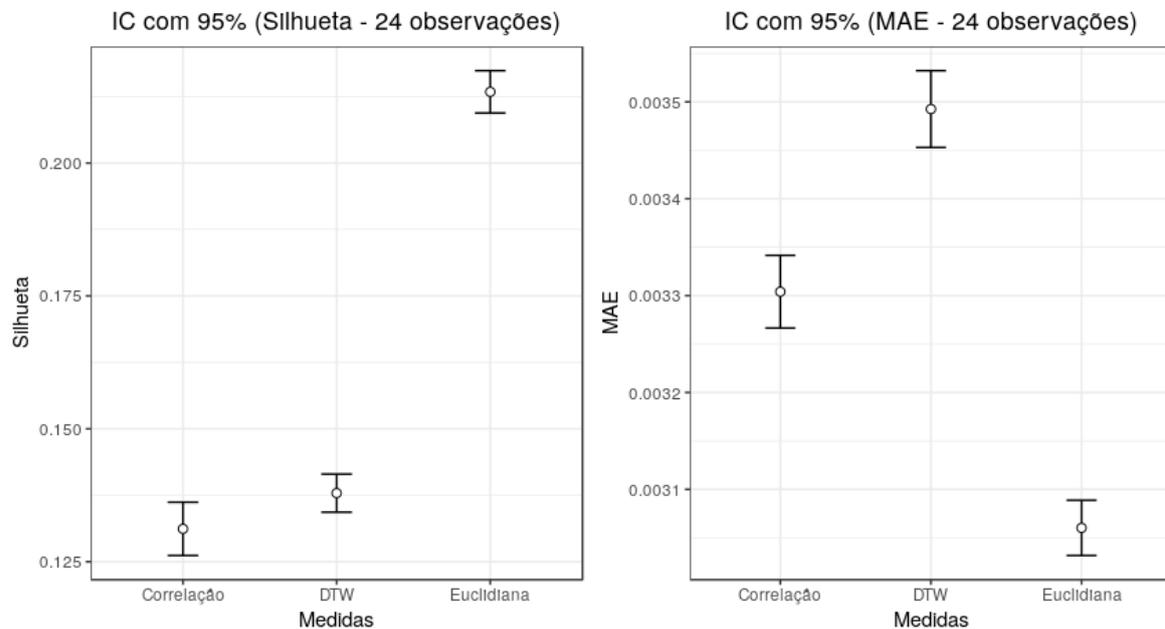


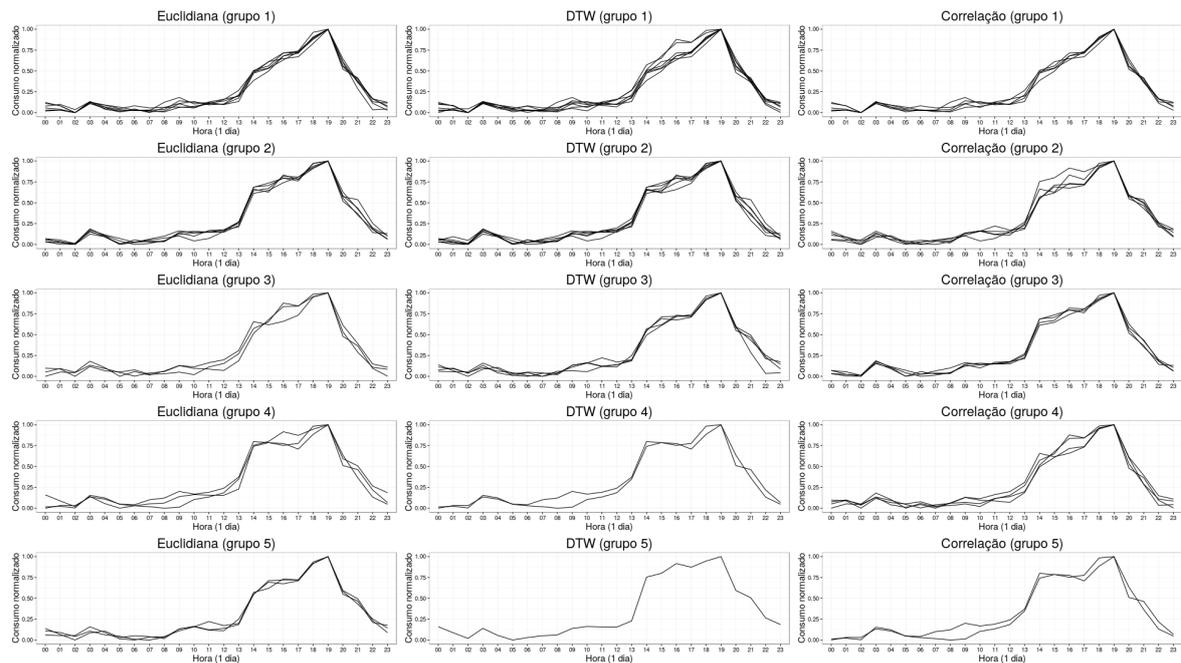
Tabela 5.3: Limites do intervalo de confiança para a silhueta e MAE (24 observações).

	Silhueta			MAE		
	IC inferior	Média	IC superior	IC inferior	Média	IC superior
<b>Correlação</b>	0,126	0,131	0,136	0,00326	0,0033	0,00334
<b>DTW</b>	0,134	0,138	0,142	0,00345	0,0035	0,00353
<b>Euclidiana</b>	0,209	0,213	0,217	0,00303	0,00306	0,00309

Para uma análise visual de curvas de carga agrupadas foram feitos agrupamentos com as 20 primeiras curvas de carga da base de dados. Dessa forma, foi possível verificar se os grupos formados tinham curvas parecidas. Na Figura 5.4 temos os agrupamentos com dados de 1 dia (24 observações) de cada perfil de consumo, no qual foram formados 5 grupos para cada medida de dissimilaridade. Por exemplo, na primeira coluna da figura temos 20 curvas de diferentes consumidores de um dia agrupadas com a utilização da distância euclidiana como medida de dissimilaridade.

Da coluna esquerda para a direita, da Figura 5.4, temos perfis agrupados utilizando distância euclidiana, DTW e correlação. A distância euclidiana obteve um melhor resultado nas métricas, que pode ser observado na Figura 5.4 e Tabela 5.4, no qual nos grupos criados usando a euclidiana temos perfis de consumo mais parecidos (menos variações) do que as demais (DTW e correlação). As diferenças entre os grupos formados são pequenas, isso

Figura 5.4: Agrupamento de 20 curvas de carga (1 dia - 24 observações).



porque os perfis de consumo da base de dados sintética são bem parecidos.

Tabela 5.4: Valor da silhueta e MAE para o agrupamento com 20 curvas de carga.

	Silhueta	MAE
<b>Correlação</b>	0,08	0,00105
<b>DTW</b>	0,15	0,00105
<b>Euclidiana</b>	0,22	0,00104

Como a base de dados sintética possui dados coletados a cada hora, foi utilizada outra base disponível publicamente que contém dados de 1 mês de consumo de 20 residências do Reino Unido com observações menores que 1 hora, chamada REFIT (MURRAY et al., 2015). Essa base foi escolhida para analisar se com uma quantidade menor de observações, por exemplo a cada 10 minutos, os resultados são semelhantes quando tenho observações a cada hora. Na Figura 5.5 temos as curvas de carga agrupadas de 1 dia com observações a cada 10 minutos, resultando em 144 observações.

Observando a Figura 5.5, da esquerda para a direita, temos perfis agrupados utilizando correlação, euclidiana e DTW. Semelhante aos resultados anteriores, a euclidiana obteve o maior valor da silhueta e o menor MAE entre as medidas de dissimilaridade (ver Tabela 5.5) e também possui grupos com curvas mais parecidas.

Figura 5.5: Agrupamento das curvas de carga da base REFIT (144 observações).

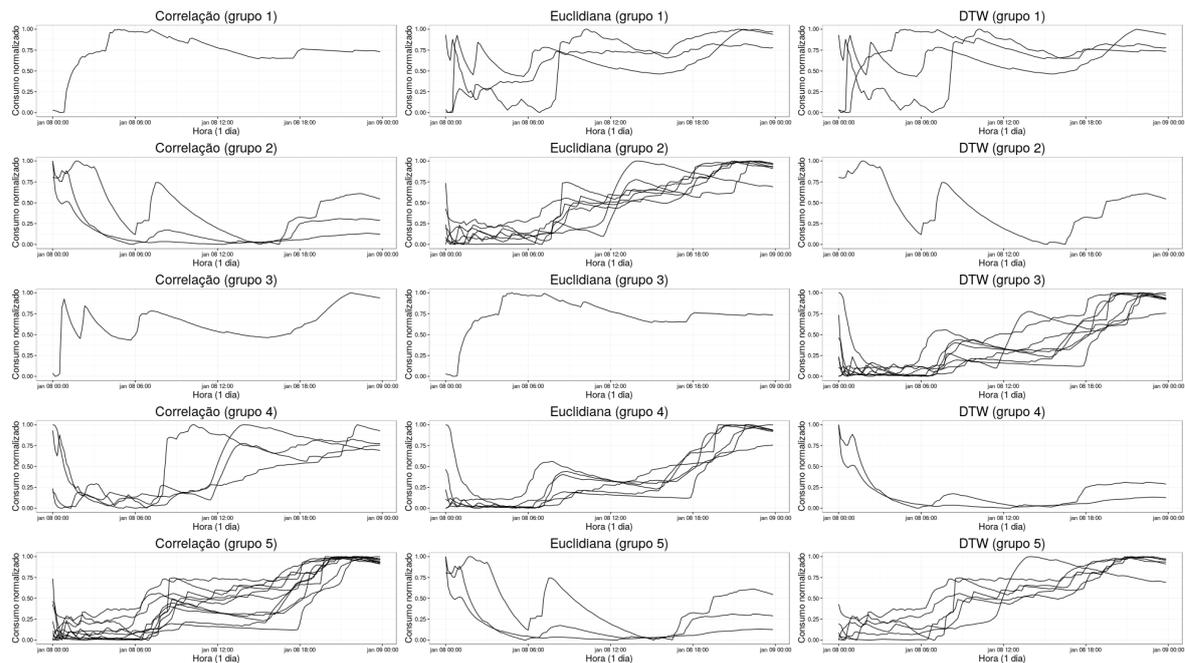


Tabela 5.5: Valor da silhueta e MAE (base REFIT - 144 observações).

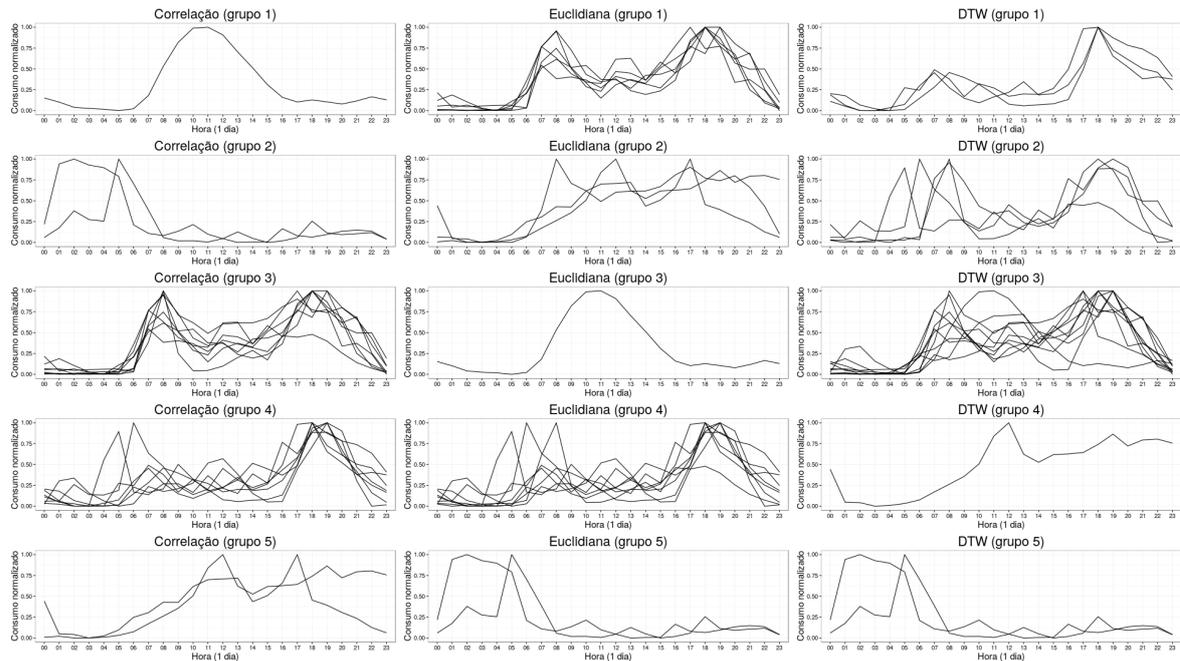
	<b>Silhueta</b>	<b>MAE</b>
<b>Correlação</b>	0,27	0,110
<b>DTW</b>	0,23	0,100
<b>Euclidiana</b>	0,30	0,099

Ainda com a base de dados REFIF foram feitos agrupamentos utilizando perfis de consumo com 24 observações, para simular um perfil no qual temos apenas dados de hora em hora e para verificar se os resultados são semelhantes ao agrupamento de perfis com 24 observações da base sintética. Os resultados estão ilustrados na Figura 5.6 e na Tabela 5.6 temos o valor da silhueta e MAE referente a esses agrupamentos. Os resultados foram semelhantes ao agrupamento dos perfis sintéticos, com a distância euclidiana obtendo os melhores resultados.

Tabela 5.6: Valor da silhueta e MAE (base REFIT - 24 observações).

	<b>Silhueta</b>	<b>MAE</b>
<b>Correlação</b>	0,16	0,104
<b>DTW</b>	0,13	0,115
<b>Euclidiana</b>	0,18	0,102

Figura 5.6: Agrupamento das curvas de carga da base REFIT (24 observações).



Em todos os experimentos feitos agrupando as curvas de carga, seja variando a quantidade de observações ou utilizando outra base de dados, a distância euclidiana obteve os melhores resultados comparando com a correlação e DTW. Com base nisso, podemos inferir que a distância euclidiana é a melhor medida de dissimilaridade entre as três para representar a semelhança entre as curvas de carga.

Nos experimentos que envolvem agrupamentos combinando alguns fatores extraídos da base de dados (seção 5.2), a medida de dissimilaridade utilizada entre os perfis de consumo (perfis sintéticos combinados com os dados do questionário) foi a euclidiana. A escolha do uso da distância euclidiana foi devido aos melhores resultados nas métricas silhueta e MAE, por criar grupos com curvas de carga mais parecidas do que as outras medidas de dissimilaridade, e também pelo cálculo ser mais simples, exigindo menos tempo computacional, principalmente se a base de dados for grande.

## 5.2 Agrupamentos de consumidores de energia elétrica

Os resultados aqui são dos 15 cenários apresentados na seção 4.2 do capítulo 4, utilizando uma quantidade fixa de grupos, no qual esta quantidade foi alterada entre 3 e 8, gerando 90 (6

configurações de grupos x 15 cenários) experimentos. Os fatores extraídos da nossa base de dados que fazem parte desses cenários são: Características diversas (localização geográfica e valor de consumo), *Feedback* do consumidor, NIALM (posse de aparelhos) e o Perfil de consumo de energia elétrica. Os 15 cenários estão descritos a seguir e mais detalhes são encontrados no Capítulo 4, seção 4.2.

**Cenários:**

1. Características diversas;
2. Características diversas + *Feedback*;
3. Características diversas + *Feedback* + NIALM;
4. Características diversas + *Feedback* + Perfil;
5. Características diversas + *Feedback* + NIALM + Perfil;
6. Características diversas + NIALM;
7. Características diversas + Perfil;
8. Características diversas + NIALM + Perfil;
9. *Feedback*;
10. *Feedback* + NIALM;
11. *Feedback* + Perfil;
12. *Feedback* + NIALM + Perfil;
13. NIALM;
14. NIALM + Perfil;
15. Perfil.

Primeiro foram feitos agrupamentos com 3 grupos, para cada cenário. Dos agrupamentos obtidos, alguns cenários tiveram resultados iguais a de outros cenários, por exemplo: os grupos formados no cenário 2 são iguais aos grupos do cenário 9 (Ver Tabela 5.7). Com

os resultados utilizando a quantidade de grupo fixa em 3, podemos reduzir a quantidade de cenários propostos inicialmente de 15 para 7, facilitando a análise subjetiva dos resultados, pois não há necessidade de analisar 15 resultados se alguns são iguais. Na Tabela 5.7, cada número dentro da coluna cenário é o ID de cada residência da base de dados.

Tabela 5.7: Grupos formados nos cenários 2 e 9.

Grupo	Cenário 2	Cenário 9
1	1, 7, 14, 17, 20	1, 7, 14, 17, 20
2	3, 4, 6, 11, 12, 19, 22, 23, 26, 30, 33, 34, 36, 39, 40, 44	3, 4, 6, 11, 12, 19, 22, 23, 26, 30, 33, 34, 36, 39, 40, 44
3	2, 5, 8, 9, 10, 13, 15, 16, 18, 21, 24, 25, 27, 28, 29, 31, 32, 35, 37, 38, 41, 42, 43, 45, 46, 47	2, 5, 8, 9, 10, 13, 15, 16, 18, 21, 24, 25, 27, 28, 29, 31, 32, 35, 37, 38, 41, 42, 43, 45, 46, 47

Com os resultados dos agrupamentos com 3 grupos percebemos que alguns fatores não influenciam na formação dos grupos quando estão combinados com outros, por exemplo: As características diversas quando combinadas com o *Feedback* não têm influência nos resultados, pois o agrupamento com apenas o *Feedback* (Cenário 9) é o mesmo do agrupamento do cenário 2 (características + *Feedback*), assim podemos escolher utilizar o cenário que possui menos fatores (ou apenas 1 fator). Utilizar poucos fatores para criar grupos exige menos informações e, portanto, menos processamento e armazenamento, além de ser mais fácil obter uma informação do que várias.

Para as demais configurações de grupos (4 a 8) foi feita a mesma análise, ou seja, foi observado quais cenários apresentaram agrupamentos iguais e assim reduzimos a quantidade de cenários em cada configuração de grupo. Na Tabela 5.8 temos todos os cenários e os seus semelhantes em cada configuração de grupo, por exemplo, o cenário 3 é igual ao 10 em todas as configurações. O resultado da redução da quantidade de cenários é:

- 3 Grupos: 7 cenários (1, 9, 10, 11, 13, 14 e 15);
- 4 Grupos: 8 cenários (1, 6, 9, 10, 11, 13, 14 e 15);
- 5 Grupos: 8 cenários (1, 7, 9, 10, 11, 13, 14 e 15);
- 6 Grupos: 10 cenários (1, 6, 7, 9, 10, 11, 12, 13, 14 e 15);
- 7 Grupos: 9 cenários (1, 7, 9, 10, 11, 12, 13, 14 e 15);

- 8 Grupos: 10 cenários (1, 6, 7, 9, 10, 11, 12, 13, 14 e 15).

Tabela 5.8: Cenários semelhantes em cada configuração de grupo.

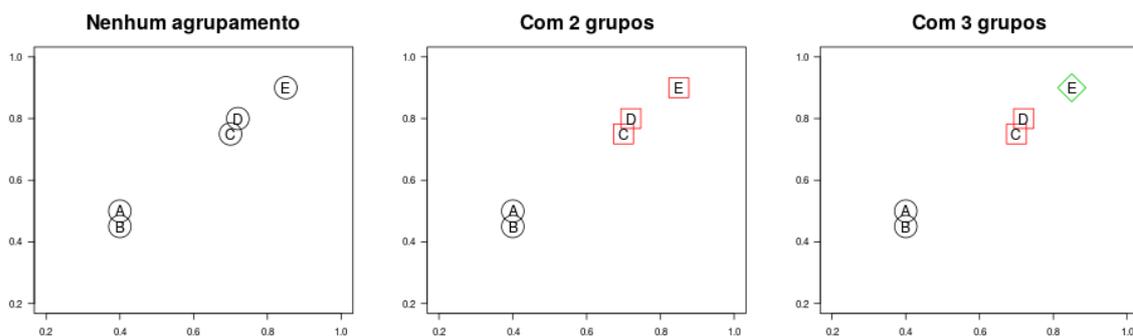
	Configurações de grupos					
	3	4	5	6	7	8
<b>Cenário 1</b>	-	-	-	-	-	-
<b>Cenário 2</b>	9	9	9	9	9	9
<b>Cenário 3</b>	10	10	10	10	10	10
<b>Cenário 4</b>	5, 11 e 12	5, 11 e 12	5, 11 e 12	11	11	11
<b>Cenário 5</b>	4, 11 e 12	4, 11 e 12	4, 11 e 12	12	12	12
<b>Cenário 6</b>	13	-	13	-	13	-
<b>Cenário 7</b>	15	15	-	-	-	-
<b>Cenário 8</b>	14	14	14	14	14	14
<b>Cenário 9</b>	2	2	2	2	2	2
<b>Cenário 10</b>	3	3	3	3	3	3
<b>Cenário 11</b>	4, 5 e 12	4, 5 e 12	4, 5 e 12	4	4	4
<b>Cenário 12</b>	4, 5 e 11	4, 5 e 11	4, 5 e 11	5	5	5
<b>Cenário 13</b>	6	-	6	-	6	-
<b>Cenário 14</b>	8	8	8	8	8	8
<b>Cenário 15</b>	7	7	-	-	-	-

Observando os resultados da redução da quantidade de cenários, apresentado anteriormente, percebemos que alguns cenários aparecem em todas as configurações de grupos e outros não, por exemplo: os cenários 1, 9, 10 e 11 estão presentes nas quantidades de 3 a 8 grupos, enquanto o cenário 12 está presente apenas nas quantidades de 6 a 8 grupos. Essa informação foi utilizada para ordenar os cenários por relevância (Tabela 5.9), de acordo com os cenários que mais aparecem em todas as configurações de grupos, no qual o exemplo anterior mostra que os cenários 1, 9, 10 e 11 são mais relevantes que o 12, pois aparecem em todas as configurações de grupos.

Foi possível diminuir a quantidade de cenários em cada experimento fixando a quantidade de grupos, mas em cada configuração de grupo a quantidade de cenários foi diferente, variando entre 7 e 10. Quando altera a quantidade de grupos alguns cenários deixam de ser semelhantes, como o que acontece com os cenários 6 e 13, que com 3 grupos eles são semelhantes mas quando aumento a quantidade de grupos para 4 eles ficam diferentes. Para um melhor entendimento do que acontece em cada agrupamento temos um exemplo na Figura 5.7, no qual:

- Temos 5 elementos, A, B, C, D e E, para serem agrupados (Figura 5.7);
- Fixando a quantidade de grupos em 2, o resultado é A e B em um grupo e C, D e E em outro;
- Porém, se alterar a quantidade de grupos para 3, o resultado é diferente do anterior, onde o elemento E forma um grupo sozinho.

Figura 5.7: Exemplo de um agrupamento com 2 e 3 grupos.



O elemento E ficou em um grupo separado pois ele não é muito semelhante aos elementos C e D (ver Figura 5.7). Antes ele ficou no mesmo grupo que C e D porque a quantidade de grupos eram 2, então ele teria que ficar em um grupo, no qual o mais próximo era com C e D. Esse é um exemplo do que acontece com os cenários, por isso tenho resultados diferentes quando é alterada a quantidade de grupos.

Observando os vários experimentos, podemos ver quais são os cenários que mais aparecem em cada configuração de grupo e assim ordená-los por relevância. A relevância além de ser em relação à quantidade de vezes que determinado cenário aparece nas diferentes configurações, também está relacionada a dificuldade de obter determinada informação (fatores). Por exemplo, as características diversas são informações fáceis de obter observando apenas a conta de energia, diferente do perfil de consumo que é necessário um equipamento para capturar as informações da rede elétrica. Os cenários estão ordenados por relevância na Tabela 5.9.

Temos apenas 10 cenários na Tabela 5.9 porque esse é o resultado da redução da quantidade de cenários, explicado anteriormente no início desta seção. A coluna “Quantidade

Tabela 5.9: Cenários e sua relevância.

Cenários	Fatores	Quantidade de grupos
1	Características diversas	3 a 8
9	<i>Feedback</i>	3 a 8
15	Perfil	3 a 8
11	<i>Feedback</i> + Perfil	3 a 8
13	Aparelhos	3 a 8
10	<i>Feedback</i> + Aparelhos	3 a 8
14	Aparelhos + Perfil	3 a 8
7	Características diversas + Perfil	5 a 8
6	Características diversas + Aparelhos	4, 6 e 8
12	<i>Feedback</i> + Aparelhos + Perfil	6 a 8

de grupos” informa em quais configurações determinado cenário aparece, por exemplo, o cenário 1 está presente nos experimentos de 3 a 8 grupos, enquanto o 12 aparece em apenas 3 experimentos (6 a 8 grupos).

Foram escolhidos os 5 primeiros cenários mais relevantes para analisar os grupos formados em cada experimento, pois os demais cenários utilizam os mesmos fatores. Em cada cenário foi escolhido 1 grupo para uma análise mais detalhada dos resultados, com o objetivo de fazer uma validação subjetiva. A escolha dos grupos foi feita aleatoriamente para não influenciar nos resultados da análise. Na subseção 5.2.1 temos essa avaliação subjetiva.

### 5.2.1 Analisando 1 grupo de cada cenário relevante

Devido a grande quantidade de resultados (grupos formados) foram escolhidos aleatoriamente quais grupos seriam analisados, verificando se os grupos formados fazem sentido, ou seja, é composto por residências (consumidores) semelhantes. Foi escolhido 1 grupo de cada cenário relevante (1, 9, 11, 13 e 15) utilizando a configuração com 6 grupos, no qual essa configuração também foi escolhida aleatoriamente entre as 6 disponíveis.

#### Cenário 1: Características diversas

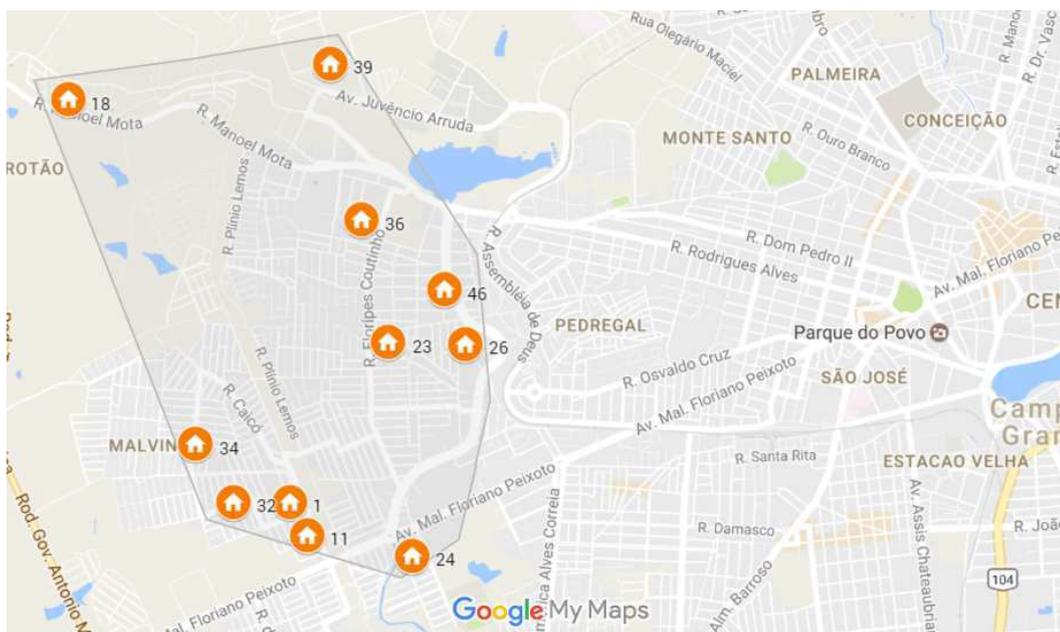
As características diversas são a localização geográfica e o valor de consumo mensal de energia elétrica de cada residência. As análises foram feitas utilizando essas informações. O grupo escolhido é formado por 11 residências: 1, 11, 18, 23, 24, 26, 32, 34, 36, 39 e 46. Esses números representam cada residência (um identificador) das 47 da nossa base de

dados, no qual na Tabela 5.10 temos algumas informações desse grupo e na Figura 5.8 temos uma noção da localização geográfica de cada residência.

Tabela 5.10: Informações sobre as residências de um grupo do cenário 1.

ID	Bairro	Consumo (kWh / mês)
1	Malvinas	170
11	Malvinas	123
18	Serrotão	250
23	Bodocongó	85
24	Dinamérica	187
26	Bodocongó	178
32	Malvinas	230
34	Malvinas	130
36	Bodocongó	116
39	Novo Bodocongó	163
46	Bodocongó	134

Figura 5.8: Localização geográfica de cada residência de um grupo do cenário 1.



Nesse agrupamento podemos observar que as residências estão próximas geograficamente (Figura 5.8), com exceção da 18, mas ela ficou nesse grupo por ter o valor de consumo próximo da 32, que pode ser observado na Tabela 5.10. A residência 23 possui um valor de consumo bem abaixo da média do seu grupo, mas está próxima geograficamente de várias residências, por isso é esperado ela estar nesse grupo. Além disso, a maioria das residências

estão na mesma faixa de consumo, no qual apenas a 18 e 32 estão na faixa entre 201-300 kWh. Percebemos que o grupo formado faz sentido, pois não há residências com características muito diferentes das outras.

Observando o valor da métrica silhueta para esse grupo, temos um valor de 0,23, que de acordo com a faixa de valores da métrica não podemos considerar esse grupo como bom, pois o ideal seria um valor próximo a 1. Esse valor não condiz com a análise subjetiva, que mostra que o grupo é formado por residências semelhantes. Mais detalhes sobre a avaliação subjetiva e a silhueta podem ser encontrados na seção 5.2.2.

### Cenário 9: *Feedback*

O *feedback* está relacionado as recomendações passadas para os consumidores, no qual são criados grupos de consumidores com interesses comuns, ou seja, consumidores que estão dispostos a realizar as mesmas ações para economizar energia. O grupo escolhido é formado por 14 residências: 2, 8, 13, 16, 18, 24, 25, 28, 29, 31, 38, 41, 42 e 47. Na Tabela 5.11 temos informações desse grupo relacionadas às classes de recomendações.

Tabela 5.11: Informações sobre as residências de um grupo do cenário 9.

ID	TA	AF	TL	MK	MC	AL	MR	MH
2	1	1	1	0	0	0	1	0
8	0	1	0	0	0	0	1	1
13	0	1	1	0	0	0	0	0
16	0	1	1	0	1	1	1	0
18	0	1	1	0	0	0	0	1
24	1	1	1	0	1	0	0	0
25	0	1	1	0	0	0	1	0
28	1	1	1	0	0	0	1	1
29	1	1	1	0	0	0	1	0
31	1	1	0	0	0	0	0	0
38	1	1	1	0	1	0	0	1
41	1	1	1	0	1	0	0	1
42	1	1	1	0	1	1	1	1
47	1	1	1	0	0	0	1	1

Observando as classes de recomendações, todos os consumidores das residências desse grupo estão dispostos a aprender funções e funcionamentos dos seus aparelhos, porém não estão dispostos a realizar mudanças na casa. Em relação às outras classes, temos algumas variações em relação às ações dos consumidores, como pode ser observado na Tabela 5.11:

- Em relação à “adicionar atividade de limpeza”, 2 consumidores dos 14 desse grupo não querem praticar essa ação;
- O mesmo acontece com a “troca de lâmpadas”, onde 2 consumidores não estão dispostos a fazer isso;
- Para “troca de aparelhos” e “mudanças de conforto”, 5 consumidores têm opiniões diferentes dos demais;
- Para “mudança na rotina” e “mudança no horário de atividade” metade dos consumidores do grupo estão dispostos a praticar essa ação e a outra metade não.

Há variações em relação às ações dos consumidores, mas no geral eles são parecidos observando as classes de recomendações que estão dispostos a seguir, ou seja, nesse grupo há pessoas que estão dispostas a praticar as mesmas ações para economizar energia. Assim podemos dizer que o grupo formado faz sentido. O valor da silhueta foi de 0,44, ou seja, um valor baixo indicando um grupo ruim, resultado diferente da avaliação subjetiva.

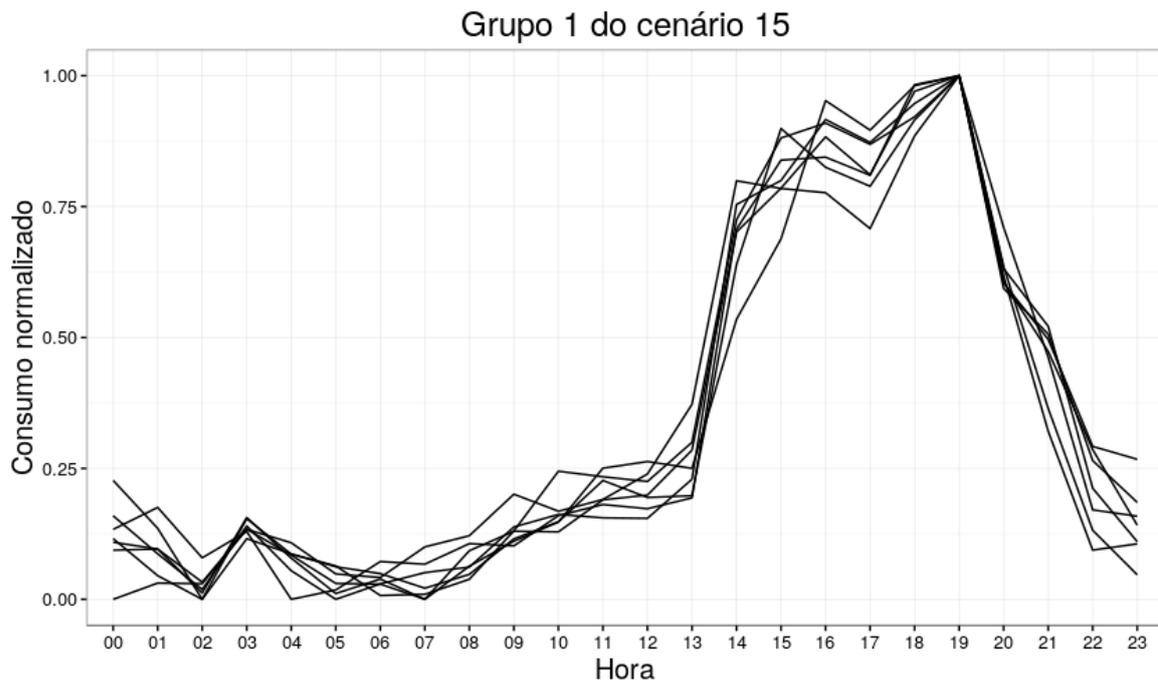
### **Cenário 15: Perfil**

Para o agrupamento dos perfis de consumo (curvas de carga) a medida de dissimilaridade utilizada foi a distância euclidiana, como discutido na seção 5.1. Um dos grupos criados foi com as seguintes residências: 2, 9, 27, 37, 40, 42 e 45. Foram utilizados os perfis de consumo de cada residência, resultando no agrupamento ilustrado na Figura 5.9, no qual podemos observar que os perfis são parecidos, maior consumo da metade do dia até o final da noite. O erro médio absoluto (MAE) é de 0,034 e o valor da silhueta é igual a 0,35.

Os perfis de consumo nos ajudam a entender como as pessoas utilizam a energia ao longo do dia, como é o caso desse grupo que tem um consumo maior no horário de pico (das 18h às 21h) e nesse horário o valor da tarifa é mais cara. Isso geralmente acontece pois as pessoas estão fora de casa durante o dia, devido a escola ou trabalho, e a noite estão todos reunidos utilizando a iluminação da residência, televisão, chuveiro elétrico, etc.

Informações apenas de perfil de consumo geram grupos que fazem sentido, como é o caso do grupo da Figura 5.9, pois torna possível identificar como aquele grupo utiliza a energia elétrica ao longo do tempo e com isso podemos criar programas para ajudar as

Figura 5.9: Perfis de consumo de um grupo do cenário 15.



peças a utilizar a energia de forma eficiente, como por exemplo: mudar sua rotina para não utilizar muitos aparelhos em horário de pico.

### **Cenário 11: *Feedback* + Perfil**

Neste cenário temos grupos formados por meio da utilização de dois fatores, feedback do consumidor e perfil de consumo da residência. O grupo escolhido é formado por 12 residências: 3, 4, 11, 12, 19, 23, 26, 33, 34, 36, 39 e 44. Na Tabela 5.12 temos informações desse grupo relacionadas às classes de recomendações e na Figura 5.10 temos os perfis de consumo agrupados.

Todos os consumidores das residências desse grupo não estão dispostos a aprender funções e funcionamento dos seus aparelhos e em relação às outras classes temos:

- Em relação à TL, MK e MR, 2 consumidores têm opiniões diferentes dos demais do grupo, ou seja, estão ou não dispostos a praticar essas ações;
- Para as classes TA, MC e AL há 4 consumidores dos 12 que compõem o grupo com opiniões diferentes;

Tabela 5.12: Informações sobre as residências de um grupo do cenário 11.

ID	TA	AF	TL	MK	MC	AL	MR	MH
3	0	0	1	0	1	0	1	0
4	1	0	1	0	1	0	1	0
11	1	0	1	0	1	0	1	0
12	1	0	1	1	0	0	1	1
19	0	0	1	0	1	0	0	1
23	1	0	1	0	1	0	1	0
26	1	0	1	0	1	1	0	0
33	1	0	0	0	0	0	1	0
34	0	0	1	0	1	1	1	1
36	1	0	1	1	0	1	1	1
39	1	0	1	0	1	0	1	0
44	0	0	0	0	0	1	1	1

- A classe que envolve mudança no horário de atividades (MH) é a mais dividida, pois 5 consumidores estão dispostos a seguir essa classe de recomendação e 7 não querem fazer isso.

Observando o *feedback* temos um grupo que faz sentido ter essa formação, no qual as ações dos consumidores são semelhantes. Mas o grupo também foi formado utilizando o perfil de consumo de cada residência e podemos observar na Figura 5.10 que eles são parecidos, com um pico de consumo próximo às 20h. O erro médio absoluto (MAE) é de 0,033 e o valor da silhueta para o grupo é de 0,35.

Podemos dizer que esse grupo faz sentido, pois temos residências que possuem perfis de consumo semelhantes e também pessoas que estão dispostas a praticar as mesmas ações para economizar energia elétrica.

### Cenário 13: NIALM

O NIALM é uma técnica de monitoramento de carga não intrusivo que fornece informações a nível de aparelho, que no nosso caso são quais aparelhos existem em cada residência. O grupo escolhido aleatoriamente é formado por 8 residências: 1, 14, 17, 20, 23, 30, 36 e 40. O grupo é formado de acordo com os aparelhos de cada residência, no qual as que possuem os mesmos aparelhos são mais semelhantes. Na Tabela 5.13 temos os aparelhos existentes em cada residência do grupo e o valor da métrica silhueta para esse grupo é de 0,25.

Observando esse grupo percebemos que todas as residências não possuem chuveiro elé-

Figura 5.10: Perfis de consumo de um grupo do cenário 11.

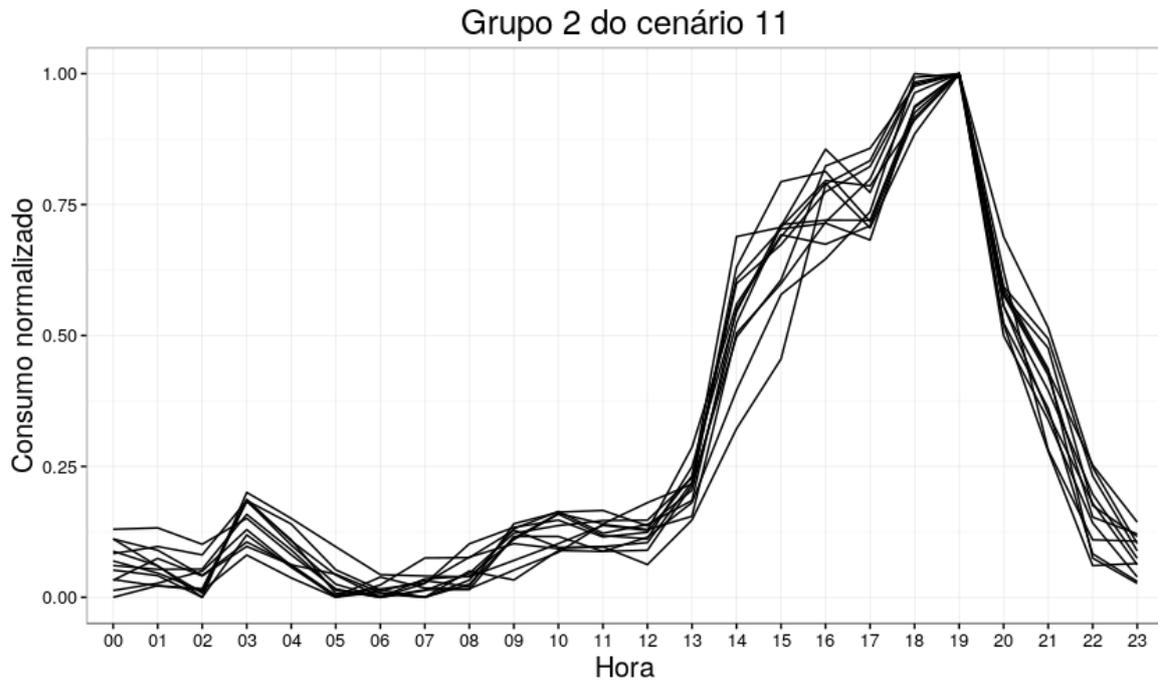


Tabela 5.13: Aparelhos em cada residência de um grupo do cenário 13.

id	Chuv. elétrico	Ar-condicionado	Lav. roupas	Micro-ondas	Ventilador	Forno elétrico	Lav. louças	Ferro elétrico	Geladeira
1	0	0	0	1	2	0	0	1	2
14	0	0	1	1	2	0	0	1	1
17	0	0	1	1	3	0	1	1	1
20	0	0	1	1	2	0	0	1	1
23	0	0	0	0	1	0	0	1	1
30	0	0	1	1	1	0	0	1	2
36	0	0	0	1	0	0	0	1	1
40	0	0	0	1	1	0	0	1	1

trico, ar-condicionado e forno elétrico, assim como há ferro elétrico e geladeira em todas, mas em relação à posse dos outros aparelhos temos:

- Metade das residências possuem máquina de lavar roupas e a outra metade não;
- Apenas 1 residência não possui microondas, no qual essa é a residência que tem o menor valor de consumo em relação às demais de seu grupo;
- Em relação ao ventilador, as residências que possuem mais de 1 tem o valor de consumo maior;

- A 17 é a única residência que possui lava louças, o que a deixa um pouco diferente das demais em relação à posse de aparelhos.

Criar grupos baseados apenas em informações do NIALM é interessante, pois ele tem potencial de gerar informações boas, por exemplo: recomendações específicas de aparelhos podem ser passadas para o grupo, como é o caso desse grupo que não adiantaria passar recomendações sobre chuveiro elétrico, sabendo que em nenhuma residência há este equipamento.

Além das recomendações, também poderíamos descobrir se alguma residência tem um valor de consumo maior por utilizar mais tal aparelho, o aparelho é antigo e por isso consome mais energia ou alguma coisa do tipo, por exemplo: as residências 14 e 20 possuem os mesmos aparelhos, tanto no tipo quanto na quantidade, e também a mesma quantidade de pessoas que habitam essas residências, mas o valor de consumo da 20 é 120 kWh enquanto o consumo da 14 é de 145 kWh mensal; Então, por que a 14 consome mais do que a 20? Pode ser porque as pessoas da residência 14 usam mais os aparelhos em relação à 20, ou algum aparelho é muito antigo, ou usam mais o ventilador em horário de pico.

São vários os motivos que podem gerar essa diferença no valor de consumo das residências que possuem os mesmos aparelhos. Isso é algo interessante que pode ser investigado em trabalhos futuros.

### **5.2.2 Análise subjetiva e a métrica silhueta**

Os valores observados da métrica silhueta em cada cenário discutido anteriormente não são bons em relação ao que é esperado pela métrica, ou seja, valores próximos a 1 indicando um agrupamento com grupos densos e separados. Para entender o motivo desses valores baixos e se a silhueta é uma boa métrica para validar os agrupamentos dentro do contexto estudado, foi feita uma avaliação subjetiva de um cenário, comparando essa avaliação com os valores resultantes da silhueta. O critério utilizado para essa avaliação subjetiva foi observar se os grupos são formados por residências com características semelhantes, ou seja, se estão próximas geograficamente e com valores de consumos semelhantes.

Para esta análise utilizamos o cenário 1 na configuração de 6 grupos, que tem 0,3 como valor absoluto da silhueta, no caso, esse é o valor da silhueta para o agrupamento geral

(média das silhuetas dos grupos). O valor relativo, ou seja, o valor da silhueta de cada grupo formado e as residências que compõem cada grupo são:

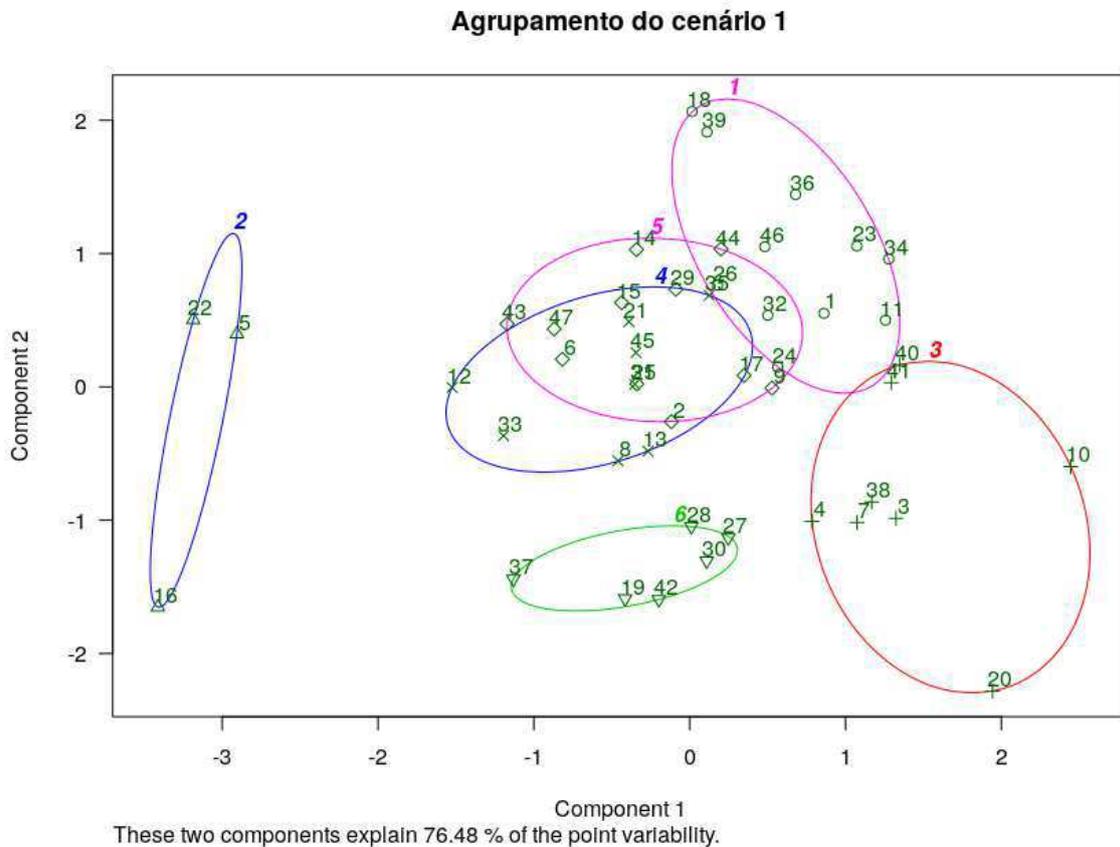
- Grupo 1: 1, 11, 18, 23, 24, 26, 32, 34, 36, 39 e 46 (silhueta = 0,23);
- Grupo 2: 5, 16 e 22 (silhueta = 0,40);
- Grupo 3: 3, 4, 7, 10, 20, 38, 40 e 41 (silhueta = 0,28);
- Grupo 4: 8, 12, 13, 21, 31, 33, 35 e 45 (silhueta = 0,43);
- Grupo 5: 2, 6, 9, 14, 15, 17, 25, 29, 43, 44 e 47 (silhueta = 0,23);
- Grupo 6: 19, 27, 28, 30, 37 e 42 (silhueta = 0,35).

O valor da métrica silhueta para o agrupamento foi de 0,3 e observando a Figura 5.11 (elaborada usando a função *clusplot* do R) percebemos que os grupos estão próximos uns dos outros e isso influencia no valor da métrica, pois um agrupamento ideal é com grupos densos e separados. Mesmo que os grupos sejam compactos, eles estão próximos e isso diminui o valor da silhueta. Isso acontece pois a população utilizada contém consumidores parecidos, por exemplo: 33 das 47 residências estão na mesma faixa de consumo, entre 0-200 kWh, enquanto 10 estão entre 201-300 kWh e apenas 4 na faixa acima de 300 kWh.

O grupo 1 desse cenário foi discutido na subseção 5.2.1, no qual esse é um grupo que fez sentido a sua formação, pois é formado por residências com características semelhantes. O grupo 2 é esparso, principalmente em relação à residência 16, como pode ser observado na Figura 5.12, ela está mais distante geograficamente. Apesar disso, esse é um grupo que faz sentido, pois estão na mesma faixa de consumo e não há outras residências próximas geograficamente e no valor de consumo para pertencerem a esse mesmo grupo. As residências mais próximas desse grupo são as do grupo 4, porém elas estão em uma faixa de consumo bastante diferente, de 0-200 kWh.

Em relação ao grupo 3, ele também é esparso no que diz respeito a localização geográfica, no qual as residências 10 e 20 são as mais distantes, porém todas estão na mesma faixa de consumo, entre 0 e 200 kWh. Mesmo sendo esparso na localização, não existem outras residências com características mais semelhantes, por exemplo: a residência 7 está próxima geograficamente da 28 (Figura 5.12), porém o consumo das duas são bastante diferentes, no

Figura 5.11: Agrupamento do cenário 1.

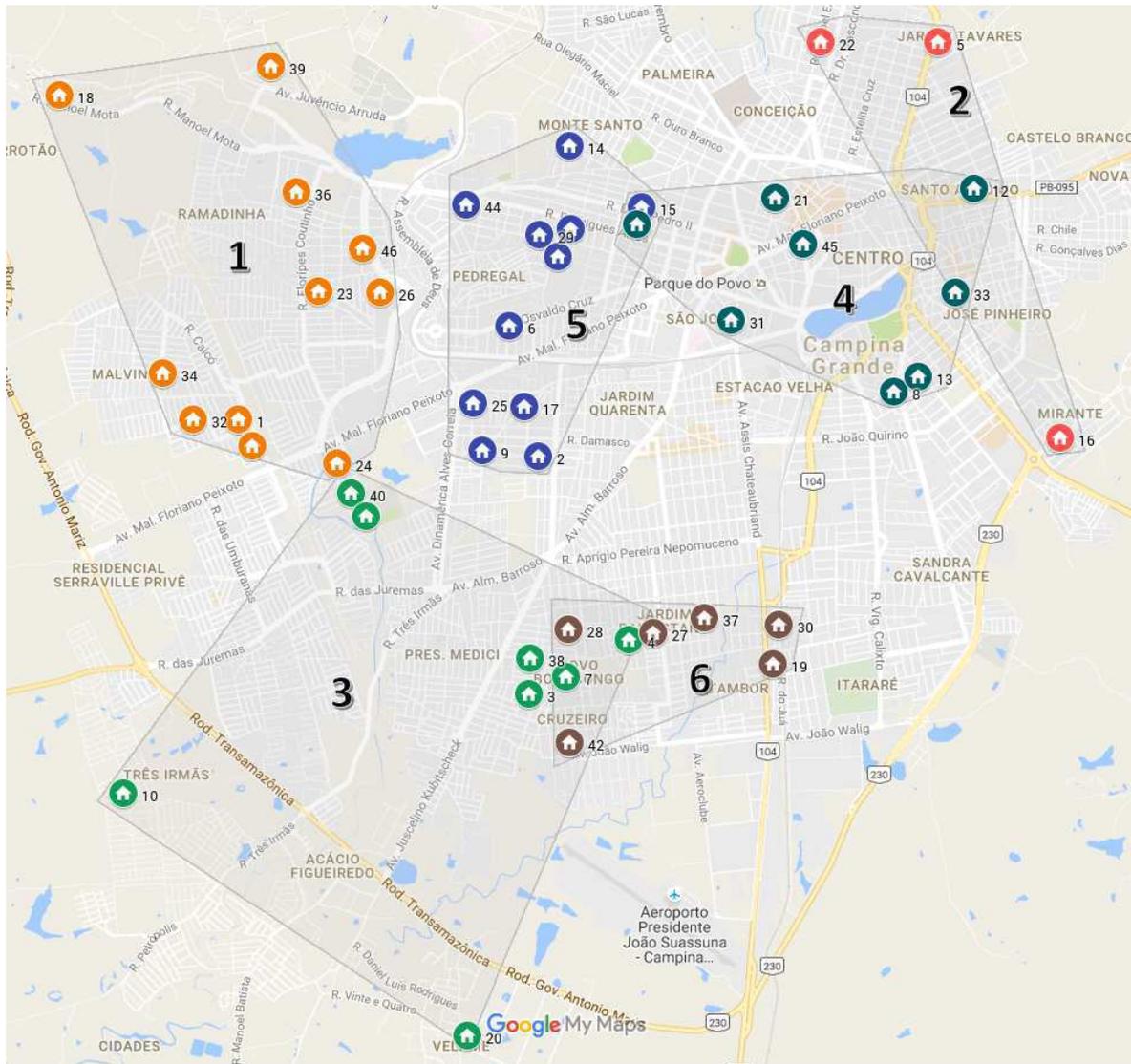


qual a 7 tem um consumo de 105 kWh e a 28 um consumo de 226 kWh. O grupo 4 também é formado por residências dentro dessa faixa de consumo, mas, diferente do 3, as residências estão próximas geograficamente, fazendo sentido a formação desse grupo.

Já no grupo 5 as residências estão próximas geograficamente, mas em relação ao valor de consumo temos 55% das residências na faixa 1 e 45% na faixa 2. Esse não é um dos melhores grupos formados, devido a sua variação nos valores de consumo, mas a localização geográfica das residências favorecem essa formação, pois não há outras residências próximas geograficamente e no valor de consumo.

O último grupo é formado por 6 residências e os valores de consumo variam entre as três faixas, onde: 2 residências pertencem a faixa 0-200 kWh, 3 a faixa 201-300 kWh e apenas 1 está na faixa acima de 300 kWh. Em relação à localização das residências, elas estão próximas. Podemos considerar que faz sentido o grupo formado, mesmo com consumos diferentes, pois não há outras residências mais parecidas que possam fazer parte desse grupo

Figura 5.12: Agrupamento do cenário 1 - localização geográfica.



ou essas residências irem para outros grupos já formados.

O valor absoluto da silhueta não condiz com a avaliação subjetiva, que mostra grupos com residências semelhantes, e por isso ela não é uma boa métrica para avaliar de forma absoluta os grupos. Em relação ao valor da silhueta de cada grupo, temos a Tabela 5.14, no qual os grupos foram ordenados do melhor para o pior<sup>1</sup> conforme a análise subjetiva e o valor da métrica silhueta, para verificar se a silhueta é uma boa métrica relativa para avaliar os grupos. Na análise subjetiva, o melhor grupo é aquele que possui residências próximas geograficamente e também em seu valor de consumo.

<sup>1</sup>"pior" não quer dizer que o grupo formado é ruim, apenas é um *ranking* dos resultados.

Tabela 5.14: Classificação dos grupos do cenário 1.

Classificação	Avaliação subjetiva	Silhueta
1º	Grupo 1	Grupo 4
2º	Grupo 2	Grupo 2
3º	Grupo 4	Grupo 6
4º	Grupo 5	Grupo 3
5º	Grupo 6	Grupo 1
6º	Grupo 3	Grupo 5

Se a classificação dada pela silhueta fosse igual à classificação da avaliação subjetiva, a silhueta seria uma boa métrica para avaliar os grupos individualmente. Mas, observando a classificação dos grupos na Tabela 5.14, temos diferenças entre a análise subjetiva e a silhueta, o que mostra que a métrica silhueta não é boa para avaliar de forma relativa os grupos. Mesmo a silhueta sendo bastante utilizada na literatura em agrupamentos de um modo geral, para o contexto desta dissertação ela não se mostrou uma boa métrica para validação dos grupos.

### 5.3 Considerações finais

Neste capítulo foram apresentados os resultados dos experimentos para encontrar a melhor medida de dissimilaridade entre as curvas de carga e dos agrupamentos de consumidores residenciais utilizando 4 fatores extraídos da base de dados. Os experimentos feitos com as medidas de dissimilaridade mostram que utilizar a distância euclidiana para representar a semelhança entre as curvas de carga é melhor do que utilizar a correlação ou DTW e isso motivou o seu uso nos agrupamentos de consumidores quando foram utilizados as curvas de carga.

Os resultados dos agrupamentos mostram que alguns fatores não influenciam na formação dos grupos quando estão combinados com outros e isso possibilitou a redução da quantidade de cenários propostos, no qual reduzimos de 15 para 10 cenários (explicado na seção 5.2). Esses 10 cenários foram ordenados por relevância, ou seja, em relação à quantidade de vezes que determinado cenário aparece nas diferentes configurações de grupos e também a dificuldade de obter determinada informação (fatores).

A partir desses cenários relevantes, avaliamos de forma subjetiva alguns grupos forma-

dos e comparamos essa avaliação com os resultados da métrica silhueta, no qual a silhueta não se mostrou uma boa métrica para a validação dos grupos, deixando assim uma brecha para trabalhos futuros em encontrar uma métrica que avalie os grupos dentro do contexto abordado nesta dissertação.

# Capítulo 6

## Conclusão

Devido ao crescimento da demanda de energia elétrica no Brasil e o impacto negativo que esse crescimento pode causar ao meio ambiente e para a sociedade, este trabalho de dissertação apresenta a análise de algumas estratégias de agrupamento de consumidores residenciais de energia elétrica, com o objetivo de entender e avaliar o processo de agrupamento quando aplicado ao contexto de eficiência energética e isso é importante pois quando feito na prática, com dados reais, pode ajudar na criação de grupos representativos de uma população e na análise dos dados dos consumidores residenciais. O setor residencial, alvo deste trabalho, é o segundo maior consumidor de energia elétrica no Brasil – cerca de 21,3% do consumo total no ano de 2015 (ENERGIA, 2016).

Neste estudo foi proposto o uso de algumas características que representam os consumidores residenciais e suas residências para a criação de grupos. Até onde estudamos, esse tipo de agrupamento utilizando características demográficas, de hábitos e posse de equipamentos elétricos não foi explorado na literatura, como é feito no presente trabalho, o que mostra a importância do desenvolvimento deste trabalho voltado para a análise de estratégias de agrupamento dentro do contexto de eficiência energética.

Inicialmente foi feita uma investigação de medidas de dissimilaridade para representar a semelhança entre perfis de consumo de energia elétrica, que é uma das características extraídas das residências. As medidas utilizadas foram a correlação, distância euclidiana e o DTW (*Dynamic Time Warping*), pois são medidas de distância bastante utilizadas na literatura quando se quer comparar séries temporais, que no nosso caso são os perfis de consumo. Entre as três, a distância euclidiana se destacou com os melhores resultados nos

experimentos realizados, seja variando a quantidade de observações das séries ou a base de dados.

Como a distância euclidiana representou melhor a semelhança entre os perfis de consumo, ela foi utilizada para o agrupamento dos perfis na segunda etapa deste trabalho. Nessa segunda etapa foram feitos agrupamentos utilizando os 4 fatores extraídos da base de dados: Características diversas (localização geográfica e valor de consumo), *Feedback* do consumidor, NIALM (posse de aparelhos) e o Perfil de consumo de energia elétrica.

Foram criados 15 cenários a partir da combinação desses fatores e vários agrupamentos foram feitos variando a quantidade de grupos. Por meio dos resultados desses agrupamentos foi possível reduzir a quantidade de cenários por serem semelhantes e também escolher os cenários (fatores) mais relevantes a serem considerados quando se quer criar grupos de consumidores residenciais. Os cenários mais relevantes são:

- Cenário 1 que utiliza a localização geográfica e o valor de consumo;
- Cenário 9 com o uso do *feedback* do consumidor;
- Cenário 15 que usa informações de perfis de consumo;
- Cenário 11 com a combinação de dois fatores, *feedback* e perfil de consumo;
- Cenário 13 que utiliza informações a nível de aparelhos (NIALM).

A relevância de cada cenário é em relação à quantidade de vezes que determinado cenário aparece nas diferentes configurações de grupos e também a dificuldade de obter determinada informação (fatores), como explicado no Capítulo 5.

Cada cenário gerou grupos diferentes, pois cada um deles tem um uso diferente, ou seja, o uso de um fator ou a combinação deles vai depender da aplicação. Por exemplo: se for de interesse criar grupos de residências com valor de consumo semelhante, não faz sentido usar informações de NIALM ou *feedback*; porém, se for de interesse agrupar residências com aparelhos semelhantes, então o uso de “apenas” NIALM faz mais sentido. Cada cenário é bom para uma aplicação específica.

Com os grupos formados surge a necessidade de validar a qualidade dos mesmos e uma das formas de fazer isso é utilizando métricas de validação de agrupamentos. Existem várias

métricas com esse propósito, mas neste trabalho foi utilizada a métrica silhueta por ser intuitiva e bastante utilizada na literatura. Apesar dela ser uma boa métrica para agrupamentos em geral, a silhueta não se mostrou uma boa métrica para a validação dos grupos dentro do contexto desta dissertação, deixando assim uma brecha para trabalhos futuros em encontrar uma métrica que avalie os grupos de forma *offline*, ou seja, uma métrica matemática que retorne um valor informando a qualidade dos grupos.

O agrupamento de consumidores residenciais ajuda a entender como os consumidores pensam e agem a partir do momento em que temos conhecimento dos hábitos de consumo e posse de aparelhos dos diversos tipos de consumidores. Além disso, os grupos também ajudam a gerar campanhas e recomendações específicas para grupos específicos, com o objetivo de incentivar o uso eficiente de energia elétrica por parte dos consumidores. Desta forma podemos perceber que o uso de agrupamento de consumidores residenciais é viável.

## 6.1 Como o sistema poderia ser usado na prática?

Mostramos como usar diferentes fatores para criar agrupamentos, no uso prático, isso poderia ser avaliado com base em métricas realistas para os grupos, como: redução do consumo, achatamento dos picos de consumo e se os grupos formados aumentam a participação das pessoas no sistema iterativo (por exemplo, uma aplicação móvel). Para entender melhor essas métricas, vejamos alguns exemplos. Se, observando todos os consumidores de um grupo, existir uma redução no consumo de eletricidade de cada um ou da maioria, avaliamos esse grupo como bom, pois a formação dele contribuiu para a redução no consumo de energia elétrica.

Quando um grupo é formado, podemos passar recomendações específicas para o grupo de acordo com as características dele e também podemos observar qual daquelas residências do grupo teve uma maior redução no seu consumo de energia e assim extrair informações dessa residência que possa ajudar as demais do seu grupo a economizar energia. Assim podemos perceber que a formação de grupos pode contribuir para a redução no consumo de energia.

Em relação ao achatamento dos picos de consumo, podemos analisar isso observando os perfis de consumo (curvas de carga) do grupo, seguindo a mesma ideia anterior, no qual

o achatamento indica uma redução do consumo na residência e se isso ocorrer na maioria delas, avaliamos esse grupo como bom. E em um sistema onde é possível obter o *feedback* do consumidor em relação às recomendações passadas para o grupo, podemos avaliar esse grupo observando a frequência de cada consumidor no sistema e verificar se o grupo contribuiu para o aumento da participação das pessoas no sistema.

Um problema que podemos encontrar é a entrada ou saída de consumidores dos grupos, pois isso pode alterar a formação dos grupos, deixando eles esparsos ou pequenos, resultando em um agrupamento ruim. Podemos resolver isso sem alterar o agrupamento já existente, calculando a distância entre o novo elemento (consumidor) e os centróides dos grupos já existentes, no qual esse novo elemento será alocado ao centróide mais próximo a ele.

Porém, pode ainda ocorrer dos grupos ficarem dispersos ao longo do tempo, devido a entrada de vários elementos e isso poderá tornar o agrupamento ruim. Uma forma de tratar isso é utilizando uma métrica para avaliar os grupos sempre que novos elementos forem adicionados, caso o valor da métrica indique que o agrupamento está ruim, refaz o agrupamento gerando novos grupos. Essa ideia também serve para a saída de consumidores.

O agrupamento pode ser utilizado na prática de acordo com a viabilidade de coletar determinadas informações, como discutido anteriormente, a utilização de determinados fatores depende da aplicação.

## 6.2 Limitações e Trabalhos futuros

A inexistência de uma base de dados reais com consumidores residenciais foi uma das dificuldades encontradas neste trabalho, limitando os experimentos aqui feitos em uma pequena população com alguns dados sintéticos. A nossa base foi construída utilizando informações coletadas de um questionário e com curvas de carga sintéticas, resultando em apenas 47 residências. Mesmo com um pequeno conjunto de dados, essa pesquisa gerou um conhecimento inicial sobre agrupamento de consumidores residenciais.

Em trabalhos futuros sugere-se repetir os experimentos em uma base de dados maior e com informações reais de residências. Além disso, é necessário encontrar uma métrica que avalie a qualidade dos grupos formados, pois a silhueta não é uma boa métrica para avaliar de forma absoluta e relativa os grupos. Um outro ponto a ser considerado em trabalhos

futuros é analisar outros fatores que podem influenciar no consumo de energia elétrica, como por exemplo: grau de escolaridade, renda familiar, quantidade de cômodos e de pessoas na residência, entre outros.

# Referências Bibliográficas

AGENCY, I. E. *Energy Efficiency*. 2016. Disponível em: <<http://www.iea.org/topics/energyefficiency/>>. Acesso em: Acesso em: 11 de Outubro de 2016.

ALENCAR, A. B. *Mineração e visualização de coleções de séries temporais*. 2007. <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-17012008-103928/pt-br.php>>. Dissertação de mestrado, Instituto de Ciências Matemáticas e de Computação - ICMC/USP. Acesso em: 11 de Novembro de 2016.

ANDRADE, F. F. et al. Metodologia para descoberta de clusters mínimos e significativos dos perfis de carga típicos de consumidores industriais e comerciais de uma distribuidora de energia elétrica. *Décimo Quinto Encontro Regional Ibero-americano do CIGRÉ*. Foz do Iguaçu-PR, Brasil, 5 2013.

ARBELAITZ, O. et al. An extensive comparative study of cluster validity indices. *Pattern Recogn.*, Elsevier Science Inc., New York, NY, USA, v. 46, n. 1, p. 243–256, jan. 2013. ISSN 0031-3203. Disponível em: <<http://dx.doi.org/10.1016/j.patcog.2012.07.021>>.

ARMEL, K. C. et al. Is disaggregation the holy grail of energy efficiency? the case of electricity. *Energy Policy*, v. 52, p. 213 – 234, 2013. ISSN 0301-4215. Special Section: Transition Pathways to a Low Carbon Economy. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0301421512007446>>.

BATRA, N. Non intrusive load monitoring: Systems, metrics and use cases. In: *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*. New York, NY, USA: ACM, 2015. (SenSys '15), p. 501–502. ISBN 978-1-4503-3631-4. Disponível em: <<http://doi.acm.org/10.1145/2809695.2822523>>.

BATRA, N. et al. NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring. In: *Fifth International Conference on Future Energy Systems (ACM e-Energy)*. Cambridge, UK: [s.n.], 2014.

BATRA, N.; SINGH, A.; WHITEHOUSE, K. If you measure it, can you improve it? exploring the value of energy disaggregation. In: *Proceedings of the 2Nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*. New York, NY, USA: ACM, 2015. (BuildSys '15), p. 191–200. ISBN 978-1-4503-3981-0. Disponível em: <<http://doi.acm.org/10.1145/2821650.2821660>>.

BEKKER, M. J. et al. Encouraging electricity savings in a university residential hall through a combination of feedback, visual prompts, and incentives. *Journal of Applied Behavior Analysis*, v. 43, n. 2, p. 327–331, 2010. Disponível em: <<http://eric.ed.gov/?id=EJ889225>>.

BENÍTEZ, I. et al. Dynamic clustering segmentation applied to load profiles of energy consumption from spanish customers. *International Journal of Electrical Power & Energy Systems*, v. 55, p. 437 – 448, 2014. ISSN 0142-0615. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0142061513004043>>.

CARVALHO, J. P. J.; NETO, M. K. *Agrupamento de padrões de curva de carga utilizando algoritmos e técnicas de agrupamento como alternativa tarifária*. 2010. <[http://nupet.daelt.ct.utfpr.edu.br/tcc/engenharia/doc-equipe/2010\\_2\\_32/2010\\_2\\_32\\_artigo.pdf](http://nupet.daelt.ct.utfpr.edu.br/tcc/engenharia/doc-equipe/2010_2_32/2010_2_32_artigo.pdf)>. Universidade Tecnologia Federal do Pará - UTFPR. Acesso em: 07 de Novembro de 2016.

CASTRO, D. F. *Eficiência energética aplicada a instalações elétricas residenciais*. 2015. <<http://monografias.poli.ufrj.br/monografias/monopoli10013941.pdf>>. Projeto de graduação, Universidade Federal do Rio de Janeiro - UFRJ. Acesso em: 17 de Outubro de 2016.

CHICCO, G. Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy*, v. 42, n. 1, p. 68 – 80, 2012. ISSN 0360-5442. 8th World Energy System Conference, {WESC} 2010. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0360544211008565>>.

EFFICIENCY, Q. R. E. *NIALM et desagregation courbe de charge*. 2015. <<http://qualisteo.com/ws/nialm-et-desagregation-courbe-de-charge/>>. Acesso em: 02 de Novembro de 2016.

ELETROBRAS. *Avaliação do mercado de eficiência energética no Brasil - Pesquisa de Posse de equipamentos e Hábitos de uso (PPH): Ano base 2005/2006*. 2007. <<http://www.procelinfo.com.br/main.asp?View=%7B4A5E324F-A3B0-482A-B1CD-F75A2A150480%7D>>. Acesso em: 01 de Novembro de 2016.

ENERGÉTICA, E. de P. *Avaliação da Eficiência energética na indústria e nas residências no horizonte decenal (2010-2019)*. 2010. <[http://www.epe.gov.br/mercado/Documents/S%C3%A9rie%20Estudos%20de%20Energia/20100809\\_4.pdf](http://www.epe.gov.br/mercado/Documents/S%C3%A9rie%20Estudos%20de%20Energia/20100809_4.pdf)>. Série: Estudos da demanda. Ministério de Minas e Energia. Acesso em: 25 de Outubro de 2016.

ENERGÉTICA, E. de P. *Anuário estatístico de energia elétrica 2016: ano base 2015*. 2016. <<http://www.epe.gov.br/AnuarioEstatisticodeEnergiaEletrica/Anu%C3%A1rio%20Estat%C3%ADstico%20de%20Energia%20El%C3%A9trica%202016.xls>>. Acesso em: 02 de Novembro de 2016.

ENERGIA, M. de Minas e. *Balanço Energético Nacional 2016: Ano Base 2015*. [S.l.], 2016. Disponível em: <[https://ben.epe.gov.br/downloads/Relatorio\\_Final\\_BEN\\_2016.pdf](https://ben.epe.gov.br/downloads/Relatorio_Final_BEN_2016.pdf)>. Acesso em: 04 de Outubro de 2016.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. Advances in knowledge discovery and data mining. In: FAYYAD, U. M. et al. (Ed.). Menlo Park, CA, USA: American Association for Artificial Intelligence, 1996. cap. From Data Mining to Knowledge Discovery: An Overview, p. 1–34. ISBN 0-262-56097-6. Disponível em: <<http://dl.acm.org/citation.cfm?id=257938.257942>>.

FERREIRA, A. M. et al. A new method for pattern recognition in load profiles to support decision-making in the management of the electric sector. *International Journal of Electrical Power & Energy Systems*, v. 53, p. 824 – 831, 2013. ISSN 0142-0615. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0142061513002603>>.

FISCHER, C. Feedback on household electricity consumption: a tool for saving energy? *Energy Efficiency*, v. 1, n. 1, p. 79–104, 2008. ISSN 1570-6478. Disponível em: <<http://dx.doi.org/10.1007/s12053-008-9009-7>>.

FOURNIER, A. C. P. *Energia Elétrica no setor residencial à luz do consumo consciente: Resultados do estudo de caso no município de Santo André (SP)*. 2010. <<http://www.anppas.org.br/encontro5/cd/artigos/GT17-266-239-20100903221411.pdf>>. V Encontro Nacional da Anppas - Florianópolis, Santa Catarina. Acesso em: 18 de Outubro de 2016.

HART, G. W. *Prototype Nonintrusive Appliance Load Monitor*. [S.l.], 1985.

HOUDE, S. et al. Real-time feedback and electricity consumption: A field experiment assessing the potential for savings and persistence. *The Energy Journal*, v. 34, n. 1, 2013. Disponível em: <<http://www.iaee.org/en/publications/ejarticle.aspx?id=2506>>.

IGLESIAS, F.; KASTNER, W. Analysis of similarity measures in times series clustering for the discovery of building energy patterns. *Energies*, MDPI AG, v. 6, n. 2, p. 579–597, Jan 2013. ISSN 1996-1073. Disponível em: <<http://dx.doi.org/10.3390/en6020579>>.

KANG, J.; LEE, J.-H. Electricity customer clustering following experts' principle for demand response applications. *Energies*, MDPI AG, v. 8, n. 10, p. 12242–12265, Oct 2015. ISSN 1996-1073. Disponível em: <<http://dx.doi.org/10.3390/en81012242>>.

KARLIN, B.; ZINGER, J. F.; FORD, R. The effects of feedback on energy conservation: A meta-analysis. *Psychological Bulletin*, v. 141, n. 6, p. 1205–1227, 2015. Disponível em: <<https://www.apa.org/pubs/journals/releases/bul-a0039650.pdf>>.

KIM, Y.-I.; KO, J.-M.; CHOI, S.-H. Methods for generating ttps (typical load profiles) for smart grid-based energy programs. In: *2011 IEEE Symposium on Computational Intelligence Applications In Smart Grid (CIASG)*. [S.l.: s.n.], 2011. p. 1–6. ISSN 2326-7682.

LAVIN, A.; KLABJAN, D. Clustering time-series energy data from smart meters. *Energy Efficiency*, v. 8, n. 4, p. 681–689, 2015. ISSN 1570-6478. Disponível em: <<http://dx.doi.org/10.1007/s12053-014-9316-0>>.

LI, X.; BOWERS, C. P.; SCHNIER, T. Classification of energy consumption in buildings with outlier detection. *IEEE Transactions on Industrial Electronics*, v. 57, n. 11, p. 3639–3644, Nov 2010. ISSN 0278-0046.

LIAO, T. W. Clustering of time series data-a survey. *Pattern Recogn.*, Elsevier Science Inc., New York, NY, USA, v. 38, n. 11, p. 1857–1874, nov. 2005. ISSN 0031-3203. Disponível em: <<http://dx.doi.org/10.1016/j.patcog.2005.01.025>>.

MÜLLER, M. R.; FRANCO, E. M. C. Clusterização de curvas de carga para o método de dias similares. V *Simpósio Brasileiro de Sistemas Elétricos - SBSE*, v. 001, p. 1–6, 2014. ISSN 2177-6164. Disponível em: <<http://www4.unioeste.br/portalpos/media/File/fabiana.silva/2013%20-%20Marcos%20Ricardo%20Muller.pdf>>.

MURRAY, D. et al. A data management platform for personalised real-time energy feedback. In: \_\_\_\_\_. *Proceedings of the 8th International Conference on Energy Efficiency in Domestic Appliances and Lighting*. [S.l.: s.n.], 2015.

OLIVEIRA, L. A. A. de. *Tratamento de Dados de Curvas de Carga Via Análise de Agrupamentos e Transformada Wavelets*. 2013. <<http://www.cos.ufrj.br/index.php/pt-BR/publicacoes-pesquisa/details/15/2439>>. Tese de Doutorado, Universidade Federal do Rio de Janeiro - UFRJ/COPPE. Acesso em: 07 de Novembro de 2016.

PAGANS, F. G. *Clustering and Other Unsupervised Learning Methods*. 2015. <<https://www.packtpub.com/books/content/clustering-and-other-unsupervised-learning-methods>>. Acesso em: 22 de Novembro de 2016.

RÄSÄNEN, T.; RUUSKANEN, J.; KOLEHMAINEN, M. Reducing energy consumption by using self-organizing maps to create more personalized electricity use information. *Applied Energy*, v. 85, n. 9, p. 830 – 840, 2008. ISSN 0306-2619. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S030626190800007X>>.

REZENDE, S. O. *Sistemas Inteligentes - Fundamentos e Aplicações*. Barueri, SP, Brasil: Manole, 2003. ISBN 85-204-1683-7. Disponível em: <[http://www.manole.com.br/cgi/supercart.exe/idSearch?ok=detalhes.htm&nothing=nada.htm&b=221&prod\\_id=197&store\\_friend=&depto=94](http://www.manole.com.br/cgi/supercart.exe/idSearch?ok=detalhes.htm&nothing=nada.htm&b=221&prod_id=197&store_friend=&depto=94)>.

RIBEIRO, I. P. *Sistemas de Recomendação aplicados à Eficiência Energética*. 2016. Dissertação de mestrado, Coordenação de Pós-Graduação em Ciência da Computação - COPIN/UFCG.

RODRIGUES, T. A. de O. et al. Predição de função de proteínas através da extração de características físico-químicas. *Revista de Informática Teórica e Aplicada - RITA*, v. 22, n. 1, 2015. Disponível em: <<http://www.seer.ufrgs.br/rita/article/view/RITA-VOL22-NR1-29>>.

ROUSSEEUW, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, v. 20, p. 53 – 65, 1987. ISSN 0377-0427. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0377042787901257>>.

RUSSELL, S. J.; NORVIG, P. *Artificial Intelligence - A Modern Approach*. 3. ed. [S.l.]: Pearson, 2009. ISBN 0-13-604259-7.

SMITH, B. A.; WONG, J.; RAJAGOPAL, R. *A Simple Way to Use Interval Data to Segment Residential Customers for Energy Efficiency and Demand Response Program Targeting*. 2012. <<http://aceee.org/files/proceedings/2012/data/papers/0193-000182.pdf>>. ACEEE Summer Study on Energy Efficiency in Buildings. Acesso em: 25 de Outubro de 2016.

TEIVE, R. C. G. et al. Novo método de caracterização das curvas de carga típicas de consumidores industriais integrado à realidade das redes inteligentes. *VIII Congresso de Inovação Tecnológica em Energia Elétrica (VIII CITENEL)*, 2015.

TRUNG, K. N. et al. Event detection and disaggregation algorithms for nialm system. *2nd International Workshop on Non-Intrusive Load Monitoring*, 5 2014.

TSIPORKOVA, E. *Dynamic Time Warping Algorithm for Gene Expression Time Series*. 2016. <<http://www.psb.ugent.be/cbd/papers/gentxwarper/DTWalgorithm.htm>>. Acesso em: 11 de Novembro de 2016.

VINE, D.; BUYS, L.; MORRIS, P. The effectiveness of energy feedback for conservation and peak demand : a literature review. *Open Journal of Energy Efficiency*, Research Publishing, Inc., v. 2, n. 1, p. 7–15, 2013. Disponível em: <<http://eprints.qut.edu.au/58017/>>.

WILLIAMS, J. Clustering household electricity use profiles. In: *Proceedings of Workshop on Machine Learning for Sensory Data Analysis*. New York, NY, USA: ACM, 2013. (MLSDA '13), p. 19:19–19:26. ISBN 978-1-4503-2513-4. Disponível em: <<http://doi.acm.org/10.1145/2542652.2542656>>.

XU, R.; WUNSCH II, D. Survey of clustering algorithms. *Trans. Neur. Netw.*, IEEE Press, Piscataway, NJ, USA, v. 16, n. 3, p. 645–678, maio 2005. ISSN 1045-9227. Disponível em: <<http://dx.doi.org/10.1109/TNN.2005.845141>>.

ZEDAN, F. M. et al. A nonzero sum approach to interactive electricity consumption. *IEEE Transactions on Power Delivery*, v. 25, n. 1, p. 66–71, Jan 2010. ISSN 0885-8977.

ZVINGILAITÉ, E.; TOGEBY, M. *Impact of Feedback about energy consumption*. 2015. <[http://www.ea-energianalyse.dk/reports/1517\\_impact\\_of\\_feedback\\_about\\_energy\\_consumption.pdf](http://www.ea-energianalyse.dk/reports/1517_impact_of_feedback_about_energy_consumption.pdf)>. Acesso em: 17 de Outubro de 2016.

# Apêndice A

## Posse de aparelhos e Hábitos de consumo de energia elétrica

Este questionário faz parte de uma pesquisa de Mestrado em Ciência da Computação da Universidade Federal de Campina Grande (UFCG). O objetivo desta pesquisa é obter características para a construção de uma população sintética que será utilizada em experimentos de um algoritmo na área de eficiência energética.

A sua participação é anônima, não incluindo nenhum tipo de informação pessoal que possa lhe identificar.

Agradecemos desde já pela sua colaboração.

**\* Obrigatório**

1. Em que bairro a sua casa está localizada? \*

\_\_\_\_\_

2. Quantas pessoas moram na sua casa (contando com você)? \*

\_\_\_\_\_

3. Você possui algum empregado(a) da lista abaixo?

Empregada Doméstica

Diarista

Babá

Motorista

Jardineiro

Outro: \_\_\_\_\_

4. Qual o nível de escolaridade do responsável pela casa?

- Analfabeto
- Ensino Fundamental
- Ensino Médio incompleto
- Ensino Médio completo
- Ensino Superior incompleto
- Ensino Superior completo

5. Qual o valor médio da conta de energia nos últimos meses? \*

- Menos de R\$ 100
- Entre R\$ 100 e R\$ 200
- Entre R\$ 200 e R\$ 300
- Entre R\$ 300 e R\$ 400
- Entre R\$ 400 e R\$ 500
- Mais que R\$ 500

6. Qual o consumo médio dos últimos meses? \*

Consumo em KWh. Você pode verificar na sua conta de energia (Histórico de Consumo). Imagem ilustrativa ao lado.

\_\_\_\_\_

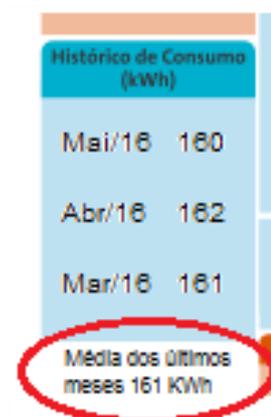
7. Qual o tipo de residência? \*

- Casa     Apartamento
- Other: \_\_\_\_\_

8. Quais desses aparelhos você possui e usa em sua residência? \*

(Pode marcar mais de uma opção.)

- Chuveiro elétrico
- Ar-condicionado
- Máquina de lavar roupas
- Microondas
- Ventilador
- Forno elétrico
- Lava louça



Ferro elétrico

Geladeira

9. Com base na resposta anterior, qual a quantidade de aparelhos? \*

	1	2	3	4 ou mais
Chuveiro elétrico	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ar-condicionado	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Máquina de lavar roupas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Microondas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ventilador	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Forno elétrico	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lava louça	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ferro elétrico	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Geladeira	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

10. Qual a renda média total da sua casa?

Menos de R\$ 1.000

Entre R\$ 1.000 e R\$ 3.000

Entre R\$ 3.000 e R\$ 5.000

Entre R\$ 5.000 e R\$ 8.000

Mais de R\$ 8.000

Não quero informar.

11. Você estaria disposto a realizar algumas ações desse tipo para reduzir o consumo de energia da sua casa? \* (Pode marcar mais de uma opção.)

Trocar aparelhos elétricos antigos por mais novos.

Aprender funções e funcionamento de seus aparelhos elétricos.

Trocar lâmpadas por modelos mais eficientes.

Realizar mudanças no interior da casa (cor da parede, posições de aparelhos, ...).

Mudanças de hábitos que alterem o seu conforto (como diminuir temperatura do chuveiro elétrico).

Adicionar atividades de limpeza específicas de alguns aparelhos elétricos.

Mudar rotina de uso de alguns aparelhos (como acumular roupas sujas para lavar uma única vez).

Mudar horários de algumas atividades da sua rotina.

Other: \_\_\_\_\_

12. Na sua casa são adotadas práticas para economizar energia? \*

Sim     Não     Não sei

13 . Se SIM, para a resposta anterior, as práticas adotadas para economizar energia foram:

- insuficientes  
 Suficientes  
 Mais que suficientes  
 Não sei  
 Other: \_\_\_\_\_

14. Na sua opinião a necessidade da redução de consumo/aumento da conta de energia alterou sua qualidade de vida/conforto?

- Não causou desconforto  
 Causou desconforto  
 Causou muito desconforto  
 Aprendi a viver com o mesmo conforto economizando energia  
 Other: \_\_\_\_\_

15 . Como você identifica o consumo elétrico dos aparelhos domésticos? \*

- Etiqueta  
 Selo Procel  
 Pesquisa em outros lugares  
 Não me interessa  
 Other: \_\_\_\_\_

16 . Você recebe ou pesquisa por informações/dicas sobre economizar energia? Se SIM, quais suas fontes de informações?

(Pode marcar mais de uma opção.)

- Internet  
 Televisão  
 Jornais  
 Revistas  
 Conta de energia  
 Amigos  
 Loja/vendedores  
 Other: \_\_\_\_\_

## Apêndice B

# Avaliação de grupos feita por especialistas

Para avaliar o uso da métrica silhueta dentro do contexto abordado solicitamos a 5 especialistas voluntários da área de eficiência energética que avaliassem, baseado na opinião pessoal/subjetiva deles, alguns grupos de residências, informando quão bom são os grupos. Essa avaliação teve como resultado a classificação de 6 grupos feita pelos especialistas, no qual o valor 1 foi atribuído ao melhor grupo e 6 ao pior. Com isso comparamos essa classificação com a da silhueta, para verificar se os grupos com melhor silhueta foram normalmente melhor avaliados pelos especialistas.

Foram analisados grupos de 6 cenários diferentes, nos quais 5 são dos cenários mais relevantes mostrados na seção 5.2 do capítulo 5. Em cada cenário desses foram escolhidos aleatoriamente 2 grupos para serem analisados pelos especialistas. O outro cenário foi escolhido de um agrupamento formado aleatoriamente, sem uso de nenhuma informação (característica) das residências.

Esse agrupamento aleatório foi feito com as 47 residências da nossa base de dados e foram criados 8 grupos aleatórios, por exemplo: Grupo A = (16,18,25,36,43,47), Grupo B = (2,7,10,28,35,47), ..., Grupo H = (1,5,6,9,12). Os grupos foram criados apenas usando esses IDs de cada residência. Isso foi feito para saber se um agrupamento puramente aleatório seria bem avaliado ou não pelos especialistas. O esperado é que os grupos desse cenário não tenham uma boa avaliação, pois não foram formados a partir de características que representassem cada residência e isso diminui as chances de ter residências semelhantes no

grupo.

A configuração de 8 grupos foi escolhida para que a quantidade de residências em cada grupo fosse pequena, pois a quantidade de informações a respeito de cada residência é grande (essas informações estão no repositório *online*<sup>1</sup>) e isso torna a análise demorada e cansativa, correndo o risco do especialista desistir da análise. Essa configuração também foi utilizada para os cenários relevantes.

Os grupos utilizados foram dos 5 cenários mais relevantes e 1 do agrupamento aleatório, totalizando 12 grupos analisados (2 de cada cenário), porém eles não foram analisados ao mesmo tempo, pois os grupos foram divididos em 2 conjuntos de 6 grupos. Os especialistas avaliaram os grupos sem saber a qual cenário cada grupo pertencia, isso para evitar qualquer viés nos resultados. Os grupos foram rotulados do 1 ao 6 da seguinte forma em cada conjunto de grupos:

- Grupo 1: formado utilizando o feedback do consumidor;
- Grupo 2: formado utilizando o feedback e o perfil de consumo residencial;
- Grupo 3: formado aleatoriamente;
- Grupo 4: formado utilizando as informações de quais aparelhos existem em cada residência (NIALM);
- Grupo 5: formado utilizando o perfil de consumo de cada residência;
- Grupo 6: formado utilizando a localização geográfica e o valor de consumo mensal (características diversas).

As residências que compõem cada grupo de cada cenário analisado estão no Apêndice C (formulário aplicado para cada especialista), ordenadas do grupo 1 ao 6.

O resultado ilustrado na Figura B.1 é da classificação dos 6 grupos (do primeiro conjunto) feita por 5 especialistas, no qual cada ponto preto no gráfico é a resposta de cada avaliador. Lembrando que o valor 1 foi atribuído ao melhor cenário e 6 ao pior. Na Tabela B.1 temos a comparação entre essa classificação e a da silhueta para o mesmo conjunto de grupos.

<sup>1</sup>[https://github.com/HarllanAndrye/EnergyEfficiency-clustering-/tree/master/Scenarios\\_clusters](https://github.com/HarllanAndrye/EnergyEfficiency-clustering-/tree/master/Scenarios_clusters)

Figura B.1: Classificação dos grupos do 1º conjunto de grupos avaliado.

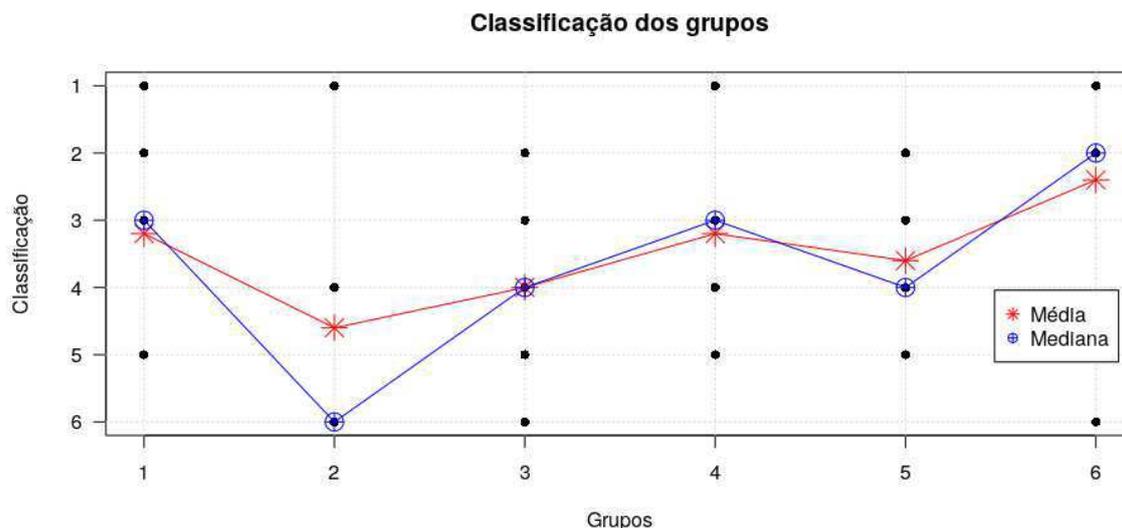


Tabela B.1: Comparação entre a classificação feita por especialistas e pela Silhueta (1º conjunto de grupos).

Classificação	Especialistas	Silhueta
1º	Grupo 6 (Carac. Diversas)	Grupo 4 (NIALM)
2º	Grupo 1 (Feedback)	Grupo 1 (Feedback)
3º	Grupo 4 (NIALM)	Grupo 6 (Carac. Diversas)
4º	Grupo 5 (Perfil)	Grupo 5 (Perfil)
5º	Grupo 3	Grupo 2 (Feedback + Perfil)
6º	Grupo 2 (Feedback + Perfil)	-

Na Figura B.2 também temos o resultado da classificação dos grupos do segundo conjunto, porém esses dados são da avaliação de 4 especialistas. Devido ao tempo para a análise dos grupos ser longo, um deles não quis avaliar o segundo conjunto de grupos. Na Tabela B.2 temos a comparação desse resultado com o da silhueta.

Nas tabelas B.1 e B.2 temos a comparação entre a classificação feita por especialistas e pela silhueta, no qual o cenário 3 não aparece na coluna Silhueta devido não ter como calcular a silhueta para esse cenário, pois foi formado aleatoriamente. Um agrupamento é feito a partir de algumas informações (características) extraídas dos dados e para o cálculo da silhueta é utilizada uma matriz de distância, que é feita a partir dessas características. Como não tenho essas características no agrupamento aleatório, não dá para obter esse cálculo.

Os critérios utilizados pelos especialistas para essa avaliação foram: o grupo com menor

Figura B.2: Classificação dos grupos do 2º conjunto de grupos avaliado.

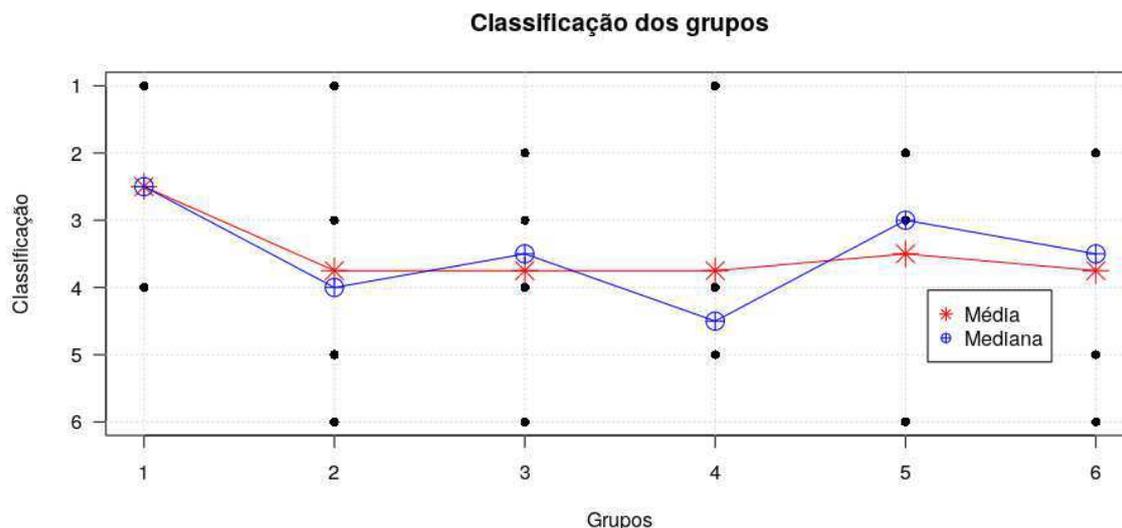


Tabela B.2: Comparação entre a classificação feita por especialistas e pela Silhueta (2º conjunto de grupos).

Classificação	Especialistas	Silhueta
1º	Grupo 1 (Feedback)	Grupo 1 (Feedback)
2º	Grupo 5 (Perfil)	Grupo 4 (NIALM)
3º	Grupo 6 (Carac. Diversas)	Grupo 5 (Perfil)
4º	Grupo 3	Grupo 6 (Carac. Diversas)
5º	Grupo 2 (Feedback + Perfil)	Grupo 2 (Feedback + Perfil)
6º	Grupo 4 (NIALM)	-

variação em relação às características que representam cada residência é considerado o melhor e o grupo que contém a maioria dos consumidores dispostos a aceitar recomendações foi melhor classificado. A localização geográfica não foi uma característica considerada importante para os avaliadores, sendo utilizada com “critério de desempate” em casos onde os grupos eram bastante parecidos nas outras características.

Comparando a classificação feita por especialistas e a classificação usando os valores da métrica silhueta, percebemos que os cenários com melhor silhueta não foram melhor avaliados pelos especialistas. Porém, os resultados foram inconclusivos devido a aleatoriedade da classificação atribuída por cada especialista e é inviável fazer muitos experimentos desse tipo pelo grande tempo necessário dos especialistas.

# Apêndice C

## Avaliação de grupos

Essa avaliação faz parte de uma pesquisa de Mestrado em Ciência da Computação da Universidade Federal de Campina Grande (UFCG) e tem como objetivo verificar a qualidade dos grupos de residências formados por meio de informações/características das mesmas. Esse agrupamento é importante para entender o comportamento dos consumidores residenciais, passo necessário para programas de eficiência energética.

Essa avaliação é voltada para especialistas da área de eficiência energética e é baseada na opinião pessoal/subjetiva do avaliador de quão bom são os grupos.

Ordene os grupos abaixo, onde 1 é o melhor grupo e 6 o pior, sendo um grupo considerado bom quando os consumidores neste grupo estariam potencialmente interessados nas mesmas recomendações para redução do consumo de energia elétrica ou teriam características comuns que ajudassem na definição de grandes perfis de consumo.

Use os documentos dos links abaixo para avaliar os grupos formados. Esses documentos contêm informações sobre as residências dos grupos formados e sobre os interesses dos consumidores dessas residências em praticar ações para redução do consumo de energia elétrica.

Localização geográfica e consumo mensal: <https://goo.gl/KxnsLP>

Posse de aparelhos: <https://goo.gl/3j5Bhf>

Ações para redução do consumo de energia elétrica: <https://goo.gl/PGs6Fb>

**\* Obrigatório**

1. Ordene os grupos abaixo do melhor para o pior, onde 1 é o melhor grupo e 6 o pior. \*

<b>Grupos</b>	<b>Residências (id)</b>
Grupo 1	5, 10, 21, 27, 32, 35, 37 e 45
Grupo 2	22
Grupo 3	3, 44, 13, 31, 9 e 47
Grupo 4	14, 17, 20 e 30
Grupo 5	3, 13, 23, 28, 31, 35, 41, 43 e 47
Grupo 6	19, 28, 37 e 42

2. Ordene os grupos abaixo do melhor para o pior, onde 1 é o melhor grupo e 6 o pior. \*

<b>Grupos</b>	<b>Residências (id)</b>
Grupo 1	8, 16, 18, 25, 28, 41, 42 e 47
Grupo 2	2, 5, 8, 10, 13, 18, 24, 25, 28, 29, 31, 38, 41 e 47
Grupo 3	17, 20, 21, 24 e 32
Grupo 4	3, 21 e 34
Grupo 5	2, 9, 27, 37, 40 e 42
Grupo 6	3, 4, 7, 20, 27, 30 e 38