

# Identificação de Perfis de Consumo Domiciliar de Energia Elétrica a partir de Algoritmos de Agrupamento

Alternative Title: Profiling Household Consumption with Clustering Algorithms

Eduardo de O. Andrade  
Departamento de Ciência da  
Computação  
Universidade Federal do Rio  
de Janeiro  
Rio de Janeiro, Brasil  
eduardoandrade@ufrj.br

Igor G. B. Sampaio  
Departamento de Eng. de  
Telecomunicações  
Universidade Federal  
Fluminense  
Niterói, Brasil  
igorgbs@gmail.com

José Viterbo  
Instituto de Computação  
Universidade Federal  
Fluminense  
Niterói, Brasil  
viterbo@ic.uff.br

João M. M. da Silva  
Departamento de Engenharia  
de Telecomunicações  
Universidade Federal  
Fluminense  
Niterói, Brasil  
jmarcos.ravel@gmail.com

Clodis Boscaroli  
Centro de C. Exatas e  
Tecnológicas  
Universidade Estadual do  
Oeste do Paraná  
Foz do Iguaçu, Brasil  
boscaroli@gmail.com

## RESUMO

A medição de consumo de energia por dispositivos inteligentes vem crescendo rapidamente, e estudos sobre estas medidas são do interesse de governos e empresas ao redor do mundo, uma vez que com a identificação de perfis de usuários dado o consumo de energia, emerge diferentes oportunidades, sejam no âmbito da prestação de serviços ou no de sustentabilidade. Este artigo apresenta uma abordagem de análise de agrupamentos para a identificação de padrões de consumo residencial de energia elétrica a partir de uma base de dados com diversos aparelhos que tiveram seus consumos de energias medidos regularmente. A partir da análise, procurou-se estabelecer perfis de consumo que colaborem para um melhor uso da energia, e foram obtidos resultados que identificaram padrões de consumo em diferentes condições de temperatura e tempo.

## Palavras-Chave

Base de dados de consumo de energia, perfil de consumo domiciliar, dispositivos inteligentes, agrupamento de dados

## ABSTRACT

The measurement of energy consumption by smart devices is growing rapidly. Studies of such measures is in the interest of governments and companies around the world. With the identification of user profiles by energy consumption, there

are plenty of opportunities that arise every time. This article presents an approach to the electric energy consumption in a home, and seeks to identify patterns. This is done on a database with many devices that have had their power consumption measured regularly. Through a clustering algorithm used on these data, we tried to establish consumer profiles that contribute to a better use of energy. They yielded results that identified energy consumption patterns on conditions such as temperature and time.

## Categories and Subject Descriptors

H.2.8 [Database Applications]: Data Mining. I.2.6 Learning

## General Terms

Datasets, Algorithms, Machine Learning

## Keywords

Energy consumption dataset, household consumption profile, smart devices, data clustering

## 1. INTRODUÇÃO

O interesse em tecnologias inteligentes que envolvem cognição, auto-reprodução, auto-correção e outras características, tem aumentado cada vez mais por parte das grandes empresas [23]. Google, Facebook e Apple são exemplos de empresas cujo interesse está pautado em buscar um aumento na capacidade de entendimento do usuário e de seus serviços. Tecnologias de automação doméstica, tendo como produtos principais: termostatos, câmeras, detectores de fumaça e outros sensores emergem nesse contexto gerando volumosas bases de dados para análise, surgindo a crescente demanda pela integração com algoritmos de aprendizagem de máquina.

Pensando no desenvolvimento de novos produtos como eletrodomésticos, eletroeletrônicos, dentre outros, e fazendo

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SBSI 2016, May 17th-20th, 2016, Florianópolis, Santa Catarina, Brazil  
Copyright SBC 2016.

uma conexão com a Internet das Coisas e tecnologias de automação doméstica, é possível observar a inserção de diversos protocolos de comunicação neste meio [34]. Com isso, percebe-se uma grande concentração por parte de pesquisadores, desenvolvedores, dentre outros, para o desenvolvimento de arquiteturas e novos protocolos que tornem tudo isso, uma realidade. Além desta diversidade de protocolos, pode-se verificar que por trás de todo este desenvolvimento, não existe apenas uma visão lucrativa no que diz respeito à venda destes dispositivos, mas sim, toda uma possibilidade de obtenção de informação [21]. A tendência é que todo este aparato tecnológico interconectado esteja presente em muitas residências ao redor do mundo em um futuro próximo, conforme afirma Want et al. [3].

Um exemplo de dispositivos como os supracitados são os medidores de energia, que podem ser utilizados para mensurar diversos parâmetros elétricos (corrente, tensão, potência, frequência, etc). Usualmente estes medidores monitoram e enviam informações regulares a um sistema central de uma rede elétrica qualquer [29]. Estes dispositivos podem ser utilizados com vistas a otimizar o consumo da rede, como já é feito em alguns países [32]. Além de monitorar uma dada rede elétrica, estes medidores acabam por enviar informações a respeito de seus usuários, tornando possível, por exemplo, a predição da quantidade de pessoas presentes em um determinado domicílio, o que torna seu uso extremamente delicado [29]. Inúmeras bases de dados deste tipo vindo sendo usadas para as mais variadas análises, porém, a necessidade da otimização do consumo energético é de interesse geral.

Algumas pesquisas tem avançado nessa área, seja pela elaboração de algoritmos que obtiveram boas probabilidades no que se diz respeito à predição de pessoas que estão presentes em uma dada residência por meio do consumo de energia [8, 24], ou pela evolução no uso de hardware e software para coleta dos dados de consumo [10].

Este artigo apresenta uma análise realizada sobre a base de dados UK-DALE [2] a partir de técnicas de agrupamento de dados considerando como características principais consumo e tempo. Os resultados gerados demonstram potencial para agregar valor à tomada de decisão a partir dos padrões identificados no consumo de aparelhos comuns situados em uma determinada residência da base de dados UK-DALE. O consumo de energia não se restringe apenas ao caso de residências, e as pesquisas e análises como as aqui apresentadas podem se expandir para outros setores comerciais.

Os medidores inteligentes que são instalados têm a capacidade de monitorar o consumo em intervalos de tempo pequenos, menores que um minuto, por exemplo. No Reino Unido, onde foi elaborada a base de dados utilizada neste trabalho, estima-se que os medidores inteligentes serão capazes de poupar um gasto na casa de bilhões de libras.

Acerca disto, os serviços públicos de alguns países, como na Irlanda, já tomam medidas para incentivar financeiramente os usuários que façam utilização da energia em momentos que não sejam os considerados horários de pico. Isto é útil para equilibrar a curva de demanda de energia elétrica. Esta prática no Brasil há alguns anos poderia ter evitado o que ficou conhecido como "apagão", crise nacional do setor ocorrida em 2001 pela falta de planejamento e investimentos, tanto na geração de energia quanto em formas de otimizar o seu uso. Para Monacchi et al. [1], a coordenação dos dispositivos é capaz de otimizar os custos de funcionamento e promover a conscientização de modo a aumentar a conser-

vação de energia e eficiência.

Outra questão a considerar são os impactos ambientais. O governo alemão, por exemplo, pretende reduzir a emissão de carbono e GHG (*greenhouse gas*, em inglês) em cerca de 80% até 2050 [4]. Para que isto possa ocorrer, é preciso alterações na matriz energética alemã e um consumo de energia mais eficiente. Esta é uma preocupação global atual e crescente, e que foi assunto central da conferência Rio+20 ocorrida no ano de 2012.

Diversos algoritmos foram utilizados na tentativa de aglomerar e analisar os dados relativos ao consumo de energia elétrica [9]. Alguns apresentaram resultados melhores, mas são um pouco mais custosos ou complexos de implementar. Todos os trabalhos mencionados neste artigo e referenciados são úteis para o crescimento nesta área de perfis de usuários, padrões e gasto de energia. É possível perceber que ainda não foi utilizado nenhum algoritmo para uma análise específica desta base de dados (UK-DALE), visto que a mesma foi elaborada recentemente (2015). Beckel et al. [33] testaram um algoritmo (NILM - *Non-Intrusive Load Monitoring*) para estimar o nível de consumo elétrico de um determinado aparelho (eletrodomésticos e afins) a partir de dados de consumo agregado de residências na base de dados (ECO (*dataset*)), o que também poderia ser feito com a base de dados ora utilizada.

A abordagem feita neste trabalho é um pouco diferente da descrita e testada por Beckel et al. [33], uma vez que a tentativa desta pesquisa foi de traçar um padrão de usuário a partir de seu consumo, em um intervalo de tempo específico, utilizando algoritmos de agrupamento. Este trabalho pode ser relevante, devido ao fato de aparelhos inteligentes estarem cada vez mais inseridos em nosso cotidiano, e com isso problemas com privacidade e segurança surgem com a vinda destes dispositivos [12].

## 2. DADOS DE CONSUMO DOMICILIAR

A base de dados UK-DALE é uma base de dados de acesso público que contém informações sobre as potências de energia elétricas consumidas de cada *channel* e os intervalos de tempo estão disponíveis em diversos arquivos que estão muito bem separados em pastas relativas à cada residência do *dataset*. Para visualizar os dados basta utilizar um software qualquer que possa lidar com arquivos de texto, principalmente aqueles capazes de trabalhar bem com uma grande quantidade de dados e valores separados por vírgula, como os populares arquivos de formato CSV.

A escolha desta base de dados se deu não somente pelo fato da integridade e confiabilidade dos dados, mas também pela quantidade de dados contidos na mesma. Outros conjuntos de dados como o GREEND, tal como UK-DALE já foram utilizados para inferência de ocupantes em uma determinada residência e otimização de recursos energéticos baseados em consumo [1], bem como serviu de base para estudos no que se diz respeito a formulação de uma nova representação de dados (aparelhos, medidores, etc), [31].

Esta pesquisa, como dito, buscou a identificação de perfis de usuários a partir do consumo em suas casas, e a utilização de uma base real (não simulada), pouco explorada e de uma grande dimensão de valores, foi útil para uma verdadeira percepção desse propósito.

Os dados utilizados neste trabalho são provenientes de uma residência na região Sul da Inglaterra. Nossa proposta central consistiu-se em elaborar perfis de consumo com base

nestes dados. Análises foram realizadas diante do uso de energia elétrica na residência por meio de seus moradores, ou seja, qualquer pessoa que tenha acionado um aparelho no período considerado para esta pesquisa e também aqueles que apresentavam um gasto contínuo por ficarem ligados praticamente todo tempo. Para exemplificar o uso perene de energia por parte destes dispositivos, podemos citar a geladeira e painéis solares.

Os parâmetros utilizados são baseados na potência elétrica (*Watts*) retornada pelos aparelhos e seus respectivos intervalos de tempo, descritos através de uma instância de tempo (*timestamp*). O *timestamp* representa um instante de tempo, conforme o padrão UNIX<sup>1</sup>. Posteriormente, utilizou-se um algoritmo para agrupar os dispositivos presentes na base, de acordo com a semelhança de consumo entre eles. Além disto, também foi considerado o período de medição do consumo dos aparelhos, a fim de conseguir melhores resultados para a análise dos perfis de consumo.

Os grupos ou *clusters* representam aparelhos que foram agrupados em um mesmo conjunto, devido à aproximação (similaridade) de seus respectivos gastos de energia elétrica. Cada aparelho ou conjunto de aparelhos iguais com os consumos mensurados foi chamado de *channel*. Cada *channel* possui seu gasto avaliado constantemente em certos intervalos de tempo. Logo, os grupos são constituídos por vários grupos de *channels* com características similares de consumo de energia.

Ainda, cada instância é definida por um intervalo de tempo. Em cada intervalo de tempo, estão presentes diversos atributos. Um atributo apresenta o consumo de energia elétrica específico de um aparelho qualquer, ou seja, uma instância possui vários atributos que correspondem ao consumo de energia de cada aparelho naquele intervalo de tempo.

A base de dados apresenta o consumo de energia elétrica de cinco residências distintas, cada uma delas com seus respectivos aparelhos, bem como eletrodomésticos, eletroeletrônicos, etc. Cada um desses dispositivos teve registrado sua potência (*Watts*) em um determinado instante de tempo (*timestamp*) e para o caso de aparelhos não ligados continuamente, o status do interruptor (0 para os que estavam desligados e 1 para os que estavam ligados naquele determinado instante de tempo). Neste trabalho dados de uma residência foram analisados ao longo de três anos, compreendidos entre 2012 e 2015.

Os números de aparelhos diferem para cada casa. A primeira residência rotulada na base de dados apresenta 53 *channels* diferentes. A segunda, 52 *channels*. A terceira residência apenas 5 *channels* e a quarta, 6 *channels*. Por último, a quinta residência apresenta 25 *channels*. A Tabela 1 traz a associação entre a primeira residência utilizada nessa pesquisa e os aparelhos. Lembrando que um *channel* pode ser um aparelho como *fridge* ou um conjunto, como *kitchen\_lights*. Foram utilizados 32 dos 53 *channels* disponibilizados na base de dados, uma vez que foram identificados alguns *channels* com intervalos de medição de tempo longos, que pouco contribuíam para a análise e dificultavam as relações entre os dados de consumo de energia dos aparelhos.

Alguns aparelhos são mais difíceis de identificar pelo nome, como *htpc* (*home theater*). Para estes, existe um arquivo *readme.txt* que os descrevem. O *channel* rotulado por *aggregate*, como o nome já diz, é um agregado dos consumos de

Aparelhos
<i>boiler</i>
<i>solar_thermal_pump</i>
<i>washing_machine</i>
<i>dishwasher</i>
<i>tv</i>
<i>kitchen_lights</i>
<i>htpc</i>
<i>kettle</i>
<i>toaster</i>
<i>fridge</i>
<i>microwave</i>
<i>breadmaker</i>
<i>amp_livingroom</i>
<i>adsl_router</i>
<i>kitchen_dt_lamp</i>
<i>bedroom_ds_lamp</i>
<i>lighting_circuit</i>
<i>livingroom_s_lamp2</i>
<i>subwoofer_livingroom</i>
<i>livingroom_lamp_tv</i>
<i>kitchen_lamp2</i>
<i>kitchen_phone&amp;stereo</i>
<i>samsung_charger</i>
<i>bedroom_d_lamp</i>
<i>coffee_machine</i>
<i>kitchen_radio</i>
<i>bedroom_chargers</i>
<i>gas_oven</i>
<i>data_logger_pc</i>
<i>childs_table_lamp</i>
<i>childs_ds_lamp</i>
<i>baby_monitor_tx</i>

Tabela 1: Residência 1 e seus respectivos aparelhos

energia elétrica dos aparelhos na casa.

Existem dois *datasets* com as informações de potência e intervalos de tempo. A diferença entre eles, é dada pela frequência de coleta dos dados. Na primeira, as informações foram coletadas de 1 em 1 segundo, já na segunda, de 6 em 6 segundos. Isto mostra a grande diferença entre as duas opções de download dos (*datasets*), uma vez que a primeira possui uma quantidade muito superior de informações, porém, para a proposta deste artigo, a utilização da segunda opção foi bastante satisfatória, haja vista tenha sido possível gerar resultados satisfatórios.

Dado que os dados foram obtidos entre os anos compreendidos entre 2012 e 2015, o uso dos aparelhos e consumo de energia elétrica ficaram sujeitos a certas circunstâncias que precisam ser consideradas:

- A temperatura é fator fundamental. Os usuários modificam o uso que fazem de uma boa quantidade de aparelhos baseado nas condições climáticas;
- Há uma perceptível diferença no perfil dos usuários em relação a dias de semana, fins de semana, feriados e férias. Logo, a quantidade destes dias diferentes no período em que foi criada a base de dados, é relevante.
- Cada residência tem sua particularidade e números de

<sup>1</sup><http://www.unixtimestamp.com/>

frequentadores. Além dos moradores, provavelmente podem ter ocorrido inúmeras visitas nas casas durante os anos considerados, alterando o resultado de consumo.

Estas são algumas das circunstâncias avaliadas como principais e que poderiam mudar os dados coletados nas residências. Alguns nomes de aparelhos sugerem a presença de certos moradores, como *childs\_table\_lamp* na *House 1*. Fora isto, não é possível identificar precisamente quantos moradores se fazem presentes em cada casa. É preciso mencionar que este trabalho levou em consideração apenas uma residência e dois dias do *dataset* relativos ao mês de Abril de 2013. Avaliou-se um dia convencional da semana (segunda-feira) e um dia de fim de semana (sábado). Temperatura, perfil dos usuários e outros detalhes são resumidos apenas para estes dois dias.

### 3. IDENTIFICAÇÃO DE PERFIS DE CONSUMO

Esta seção descreve o processo de identificação de perfis de consumo de energia a partir do algoritmo *Simple K-means* disponível na ferramenta Weka [22], uma ferramenta *open source* para o processo de mineração de dados. O *Simple K-means* é melhor caracterizado na seção seguinte.

Para obter-se uma maior precisão nos resultados e previsões, as médias de temperatura durante os dias de coleta dos dados foram obtidas a partir da consulta a repositórios de dados climáticos e adicionadas ao *dataset*. Como as residências presentes na base de dados se localizam no sul da Inglaterra, obtivemos a média de temperatura de cada dia do período selecionado, em graus Celsius, para a capital inglesa, Londres. Inclusive, o *dataset* climático foi obtido do próprio Imperial College, situado em Londres. A Tabela 2 apresenta estas medições:

Tabela 2: Temperatura média diária nos dias 1 e 6 de Abril de 2013 na cidade de Londres

Dia	Data	Temperatura Média
D1	1 de Abril de 2013	2°C
D2	6 de Abril de 2013	3°C

O dia 1 de Abril de 2013 corresponde a uma segunda-feira e o dia 6 de Abril de 2013, a um sábado. Como pode-se ver na Tabela 2, foram dias de baixas temperaturas. Inclusive, o fim do mês de Março e início de Abril, foi o período mais rigoroso do ano de 2013 em termos de baixas temperaturas. Como dito, a questão de um dia da semana, dia 1, ser um dia do meio da semana e o dia 6 ser um dia de fim de semana, apresenta uma diferença visível na mudança do padrão de consumo de energia.

Neste trabalho, logo após a geração dos resultados dos agrupamentos pelo algoritmo, ou seja, depois da separação das instâncias (faixas de tempo) e seus atributos (dados de consumo por aparelho) em grupos, estes foram situados em relação às fatias de tempo correspondentes. Em específico, foram utilizados dois dias distintos (um dia da semana e um dia do fim de semana), buscando uma clara visualização de padrões de consumo em faixas de tempo, ao invés de apenas identificar o conjunto de aparelhos com consumos

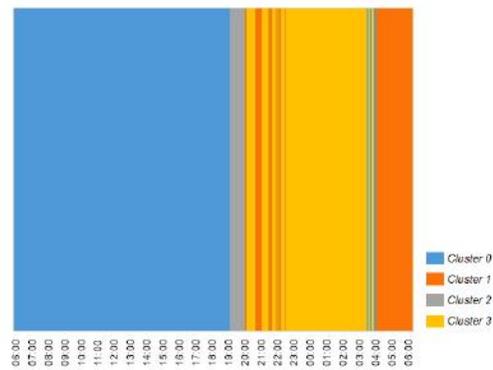


Figura 1: Distribuição dos grupos no dia 1, de acordo com o tempo

semelhantes, uma vez que o algoritmo não trabalha com análises temporais.

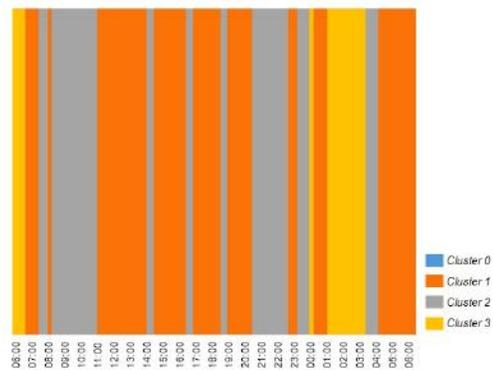


Figura 2: Distribuição dos grupos no dia 6, de acordo com o tempo

#### 3.1 Análise de Agrupamentos

O algoritmo de agrupamento de dados utilizado foi o *Simple K-means* [35]. O *Simple K-means* procura desenvolver a formação de grupos com base em valores médios similares (distância euclidiana) entre os atributos das diversas instâncias presentes em um *dataset* qualquer. É preciso escolher um  $K$  inicial (número de grupos). Após análise de testes empíricos, o número de grupos escolhidos foi quatro ( $K = 4$ ), uma vez que resultados mais expressivos foram obtidos com esta configuração, além da intenção de separar os dias em madrugada, manhã, tarde e noite, para que pudesse tornar a análise mais realista. Entre os diversos algoritmos de agrupamento presentes no Weka, o *Simple K-means* foi o que apresentou a melhor distribuição das amostras entre os grupos, sendo este motivo determinante para sua escolha. No caso deste trabalho, os atributos avaliados foram relativos ao consumo de energia elétrica de cada aparelho presente na primeira residência da base de dados UK-DALE, sendo feita uma relação com os intervalos de tempo dos dias considerados. Estes intervalos foram separados de hora em hora durante 24 horas para cada dia, mais precisamente, de 6 ho-

ras da manhã do dia 1 ou dia 2, até 6 horas da manhã dos dias seguintes (teça-feira e domingo).

Nas Figuras 1 e 2, pode-se observar a coloração com os quatro grupos (de  $K = 4$ ). Cada *cluster* teve uma cor atribuída. O Grupo 0 ficou com a cor azul, o Grupo 1 com a cor laranja, o Grupo 2 com a cor cinza e o Grupo 3 com a cor amarela. A Figura 1 é relativa ao dia 1, segunda-feira de abril do ano de 2013 e a Figura 2 relativa ao dia 6, sábado de abril do ano de 2013. No eixo horizontal tem-se os intervalos de tempo medidos em horas a partir destes dias.

Como o intervalo foi avaliado em horas, algumas faixas de presença dos grupos não aparecem na Figura 1. O Grupo 0, cor azul, está presente mas não se consegue observá-lo visualmente. Ocorre que o grupo 0 está distribuído em pequenos intervalos de tempo durante várias partes do dia. Caso fosse feito o intervalo em poucos minutos, seria possível ver estreitas faixas azuis na Figura 2. A Tabela 3 e a Tabela 4, demonstram a percentagem da distribuição dos grupos nos dias 1 e 6, respectivamente, e também, a quantidade de instâncias.

Número do cluster	Quantidade de instâncias	Percentagem
0	8148	57%
1	3188	22%
2	509	4%
3	2625	18%

Tabela 3: Percentagem da distribuição dos grupos no dia 1

Número do cluster	Quantidade de instâncias	Percentagem
0	3042	21%
1	7787	54%
2	1680	12%
3	1891	13%

Tabela 4: Percentagem da distribuição dos grupos no dia 6

Totalizando o número de instâncias em cada dia, temos 14400 instâncias dadas por intervalos de medição de consumo de energia elétrica de 6 em 6 segundos.

Os valores médios gerais de consumo de cada aparelho e depois em cada grupo foram observados, e notou-se no dia 6 aparelhos apresentando um consumo elevado em determinado grupo, indicando que ficaram ligados no mesmo período de tempo. Por exemplo, a luz da cozinha (*kitchen\_lights*) com o aumento do consumo da torradeira (*toaster*). O consumo elevado no grupo 0 é característico de ambos os atributos. Esta relação pode ser observada na Tabela 5.

Atributos	Média geral	Média do cluster 0	Média do cluster 1	Média do cluster 2	Média do cluster 3
<i>kitchen_lights</i>	7,4246	13,8245	5,3515	13,8024	0
<i>toaster</i>	14,8678	21,5874	15,9878	13,9167	0,2909

Tabela 5: Atributos *kitchen\_lights* e *toaster* relacionados pelas suas médias nos grupos do dia 6

Associando este resultado com os intervalos de tempo, observa-se que não há um intervalo longo de tempo presente na Figura 2 que caracterize o Grupo 0. A predição pode ser um cenário que envolva curtos período de uso dos aparelhos durante vários horários do dia.

Outro exemplo notado foi o da Tabela 6, na qual o Grupo 3 aparece como aquele que apresenta o maior relacionamento entre as médias elevadas de consumo. A parte da manhã e madrugada é a mais presente neste exemplo. Uma situação que englobe algumas pessoas dormindo mais tarde em um dia de fim de semana e assistindo televisão ou acordando, no caso da manhã, pode ser uma situação que explique esse comportamento. O Grupo 2 na Tabela 5 também apresenta uma média um pouco elevada de consumo. Há um intervalo de tempo associado com o Grupo 2 na Figura 2 que ocorre logo após o intervalo de tempo associado com o Grupo 3, entre as 3 e 4 horas da madrugada. Uma explicação seria uma situação de um lanche antes de algumas pessoas dormirem. O intervalo de tempo entre as 4 e 6 horas da madrugada é relativo ao Grupo 1 em ambos os dias, 2 e 7.

Atributos	Média geral	Média do cluster 0	Média do cluster 1	Média do cluster 2	Média do cluster 3
<i>tv</i>	13,5671	5,2127	1,0885	7,5214	83,7636
<i>htpc</i>	19,1875	11,9027	13,5347	13,1333	59,5627
<i>amp_livingroom</i>	2,8749	0,9859	0,0376	1,5268	18,7953
<i>subwoofer_livingroom</i>	2,3287	1,0743	0,3769	1,4458	13,1682
<i>livingroom_lamp_tv</i>	1,6	0	0,0001	0	12,1835

Tabela 6: Atributos *tv*, *htpc*, *amp\_livingroom*, *subwoofer\_livingroom* e *livingroom\_lamp\_tv* relacionados pelas suas médias nos grupos do dia 6

Na Tabela 7 referente ao dia 1, segunda-feira, particularidades entre os atributos também foram encontradas, com um caso de associação diferente: o atributo *boiler* apresenta uma média alta de consumo de energia, enquanto que o atributo *solar\_thermal\_pump* apresenta uma média baixa de consumo. Pode-se concluir que enquanto um dos aparelhos é utilizado, o outro não é utilizado ou bem pouco. Relacionando essa informação com os intervalos de tempo, o *boiler* tem uma média de consumo baixa no Grupo 0 que envolve principalmente antes do horário das 19 horas. O *solar\_thermal\_pump* ao contrário, apresenta uma média de consumo elevada de energia antes desse horário.

Atributos	Média geral	Média do cluster 0	Média do cluster 1	Média do cluster 2	Média do cluster 3
<i>boiler</i>	24,2264	0,3595	68,4179	59,6916	38,9322
<i>solar_thermal_pump</i>	9,5599	16,6448	0,6568	0	0

Tabela 7: Atributos *boiler* e *solar\_thermal\_pump* relacionados pelas suas médias nos grupos do dia 1

Mediante a análise dos dias 1 e 6, nota-se uma diferença no perfil de consumo a partir da comparação entre a Tabela 6 e a Tabela 8.

No dia 6, observa-se uma média geral mais elevada, determinando um consumo de energia muito maior dos atributos avaliados no fim de semana. Além disso, os dias diferem nos grupos também com médias mais elevadas e nos intervalos de tempo. O dia 1 apresenta maior similaridade de consumo no Grupo 2, predominando o horário entre as 19 e 20 horas. Uma situação de assistir televisão na sala de estar próximo

Atributo	Média geral	Média do cluster 0	Média do cluster 1	Média do cluster 2	Média do cluster 3
<i>tv</i>	4,7353	0,9413	0,9615	108,0589	0,9581
<i>htpc</i>	8,9958	6,8703	10,2158	71,5972	2,0046
<i>amp_livingroom</i>	0,8248	0,0064	0	23,2318	0
<i>subwoofer_livingroom</i>	0,833	0,2193	0,3775	15,9804	0,3417
<i>livingroom_lamp_tv</i>	0,327	0	0	9,2495	0

Tabela 8: Atributos *tv*, *htpc*, *amp\_livingroom*, *subwoofer\_livingroom* e *livingroom\_lamp\_tv* relacionados pelas suas médias nos grupos do dia 1

do horário de jantar ou após o horário de trabalho, revela-se mais uma possível inferência.

Verifica-se, por fim, que os dias 1 e 6 fizeram parte dos dias mais frios do mês de Abril de 2013. Quando a média do atributo *boiler* é verificada para o mês inteiro, tem-se o valor da média em torno de 19,47, abaixo do dia 1, o que confirma a relação da temperatura com os dados de consumo, e faz o atributo *boiler* no geral apresentar uma média mais elevada de consumo de energia elétrica nos dias mais frios. Esta relação é verificada em alguns outros atributos do *dataset*.

A análise acerca da identificação destes perfis de consumo nos permite conceber algumas aplicações. Os dois dias verificados apresentam perfis distintos, sendo possível separá-los pelo padrão de consumo, ou seja, pode-se rotular como um dia de fim de semana ou não. Fora isto, também há a questão dos aparelhos que ficaram ligados em determinadas faixas de tempo. Neste caso, um exemplo pode ser a alta probabilidade de alguém na residência estar presente em um determinado cômodo em algum momento do dia. Uma possibilidade seria a presença na sala de estar, pelos resultados obtidos no Grupo 3 da Tabela 6. Estendendo a análise deste trabalho, é viável/factível a elaboração de muitas outras aplicações.

## 4. CONCLUSÕES

O investimento em soluções envolvendo dispositivos inteligentes é uma tendência. A área é muito ampla e aberta a diferentes ramos de pesquisa. A compreensão das medições retornadas pelos dispositivos, envolve a percepção de padrões e melhorias. Dentre as melhorias possíveis, o aumento da eficiência e diminuição dos custos em torno do consumo de energia são exemplos relevantes. O processo de análise de agrupamentos é capaz de retornar resultados importantes quando da identificação de perfis de consumo. Apesar deste artigo se resumir a base de dados UK-DALE, o processo pode servir para qualquer base de dados de consumo de energia.

Diante da realidade brasileira, é necessário pensar em soluções geralmente de baixo custo. Monacchi et al. [1] apresentam uma proposta que pode envolver um *Raspberry Pi* como plataforma de *hardware*, conectado juntamente com alguns sensores e bateria. Além disto, em termos de *software*, utiliza-se uma biblioteca *open source* com *scripts* para a coleta dos dados. Basicamente desta forma e com algum tipo de armazenamento dos dados, podendo ser local ou até mesmo em nuvem, seriam resguardados os arquivos .CSV gerados com as informações de consumo de energia elétrica. Assim, uma aplicação envolvendo residências no Brasil é viável, pois é um país continental que pode oferecer diversas medições e padrões diferentes, dependendo das regiões onde

seriam realizados os estudos sobre os dados de consumo coletados.

Este trabalho consistiu da análise de agrupamento de dados a partir de um algoritmo clássico na literatura que permitiu identificar alguns padrões de consumo e relacioná-los a outras informações, fortalecendo a compreensão dos resultados, de forma a ser possível traçar um perfil de consumo, principalmente relacionado a intervalos de tempo. No caso, para exemplificar a abordagem proposta de cruzamento dos dados, verificou-se um perfil de consumo domiciliar que envolveu a utilização de energia elétrica em dois dias distintos por diversos aparelhos em uma única residência.

Resultados sobre os perfis foram obtidos e descritos, englobando o consumo dos aparelhos, horários durante os dias e temperatura. A partir daí, medidas poderiam ser tomadas em relação a horários de consumo elevado, por exemplo, com tributação diferenciada, como já ocorrem em alguns países.

O tema abrangendo os perfis de consumo, pode ser ampliado, focalizando em outros pontos, com a inclusão de outras residências e a consideração da quantidade de pessoas presentes nas casas. Ainda, é preciso observar que com o desenvolvimento contínuo dos dispositivos inteligentes. A expectativa é a de que estes tipos de dados sejam cada vez mais comuns e acessados. A transparência destas informações também é importante, sendo então privacidade e segurança outras áreas passíveis de novas pesquisas.

A análise de dias distintos ou até mesmo, diferentes estações do ano dessa base de dados, ou o uso de outros *datasets* e algoritmos de agrupamento de dados são também possibilidades de trabalhos futuros, bem como o uso de intervalos de tempos maiores que envolvam semanas, meses ou anos e que retornem médias mais gerais que possam demonstrar resultados mais conclusivos.

## 5. REFERÊNCIAS

- [1] Monacchi, A., Egarter, D., Elmenreich, W., D'Alessandro, S., Tonello, A.M. GREEND: An energy consumption dataset of households in Italy and Austria. In Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on, pages 511–516. IEEE, 2014.
- [2] Kelly, J., and Knottenbelt, W. UK-DALE: A dataset recording UK domestic appliance-level electricity demand and whole-house demand. CoRR, abs/1404.0284, 2014.
- [3] Want, R., Schilit, B.N., and Jenson, S. Enabling the Internet of Things. *Computer*, (1):28–35, 2015.
- [4] Paetz, A.-G., Dütschke, E., and Fichtner, W. Smart Homes as a means to sustainable energy consumption: a study of consumer perceptions. *Journal of Consumer Policy*, 35(1):23–41, 2012.
- [5] Armel, K.C., Gupta, A., Shrimali, G., and Albert, A. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy*, vol. 52, no. 0, pp. 213 – 234, 2013.
- [6] Hink, R.C.B., Beaver, J.M., Buckner, M.A., Morris, T., Adhikari, U., and Pan, S. Machine learning for power system disturbance and cyber-attack discrimination. In Resilient Control Systems (ISRCs), 2014 7th International Symposium on, pages 1–8. IEEE, 2014.
- [7] Bonfigli, R., Squartini, S., Fagiani, M., and Piazza, F. Unsupervised algorithms for non-intrusive load

- monitoring: An up-to-date overview. *Environment and Electrical Engineering (EEEIC)*, 2015 IEEE 15th International Conference on. IEEE, 2015.
- [8] Kleiminger, W., Staake, T., and Santini, S. Occupancy detection from electricity consumption data. *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. ACM, 2013.
- [9] McLoughlin, F., Duffy, A. and Conlon, M. A clustering approach to domestic electricity load profile characterization using smart metering data. *Applied Energy*, vol. 141, pp. 190-199, 2015.
- [10] Barker, S., Mishra, A., Irwin, D., Cecchet, E., Shenoy, P., and Albrecht, J. Smart\*: An Open Data Set and Tools for Enabling Research in Sustainable Homes. *Proceedings of the 2012 Workshop on Data Mining Applications in Sustainability*, Beijing, China, August 2012.
- [11] Zhang, H., Yu, N., and Hu, H. The optimal noise distribution for privacy preserving in mobile aggregation applications. *International Journal of Distributed Sensor Networks*, 2014.
- [12] Axelrod, C.W. Enforcing security, safety and privacy for the Internet of Things. *Systems, Applications and Technology Conference (LISAT)*, 2015 IEEE Long Island. IEEE, 2015.
- [13] Peng, H., Zhang, X., Chen, H., Wu, Y., Zeng, J., and Li, D. Enable privacy preservation for k-NN query in two-tiered wireless sensor networks. *Communications (ICC)*, 2015 IEEE International Conference on. IEEE, 2015.
- [14] Oliveira, S.R.M., and Zaiane, O.R. Privacy preserving clustering by data transformation. *Journal of Information and Data Management*, 1(1):37, 2010.
- [15] Alotaibi, K., Rayward-Smith, V.J., Wang, W., and de la Iglesia, B. Non-linear dimensionality reduction for privacy-preserving data classification. *Privacy, Security, Risk and Trust (PASSAT)*, 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom). IEEE, 2012.
- [16] Chunhua, S., Bao, F., Zhou, J., Takagi, T., and Sakurai, K. A new scheme for distributed density estimation based privacy-preserving clustering. *Availability, Reliability and Security*, 2008. ARES 08. Third International Conference on. IEEE, 2008.
- [17] Gokulnath, C., Pryan, M.K., Balan, E.V., Rama, P.K.P., and Jeyanthi, R. Preservation of privacy in data mining by using PCA based perturbation technique. *Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)*, 2015 International Conference on. IEEE, 2015.
- [18] Kaur, R., and Bansal, M. Transformation approach for boolean attributes in privacy preserving data mining. *Next Generation Computing Technologies (NGCT)*, 2015 1st International Conference on. IEEE, 2015.
- [19] Bertino, E., Lin, D., and Jiang, W. A survey of quantification of privacy preserving data mining algorithms. In *Privacy-preserving data mining*, pages 183–205. Springer, 2008.
- [20] Jalla, H.R., and Girija, P.N. An efficient algorithm for privacy preserving data mining using hybrid transformation. *International Journal of Data Mining & Knowledge Management Process*, 4(4):45, 2014.
- [21] Weber, R.H. Internet of things—new security and privacy challenges. *Computer Law & Security Review*, 26(1):23–30, 2010.
- [22] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I.H. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.
- [23] Gadakari, T., Mushatata, S., and Newman, R. Intelligent buildings: Key to achieving total sustainability in the built environment. *Journal of Engineering, Project, and Production Management*, 4(1):2–16, 2014.
- [24] Chen, D., Barker, S., Subbaswamy, A., Irwin, D., and Shenoy, P. Non-intrusive occupancy monitoring using smart meters, in: *Proc. BuildSys'13*, ACM, Rome, Italy, 2013.
- [25] Kolter, J.Z., and Johnson, M.J. REDD: A public data set for energy disaggregation research. In *Workshop on Data Mining Applications in Sustainability (SIGKDD)*, San Diego, CA, volume 25, pages 59–62. Citeseer, 2011.
- [26] Irwin, D., Wu, A., Barker, S., Mishra, A., Shenoy, P., and Albrecht, J. Exploiting Home Automation Protocols for Load Monitoring in Smart Buildings. In *BuildSys*, 2011.
- [27] Witten, I.H., Frank, E., and Hall, M.A. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.
- [28] Pang-Ning, T., Steinbach, M., and Kumar, V. *Introduction to data mining*. Vol. 1. Boston: Pearson Addison Wesley, 2006.
- [29] Dong, C., Kalra, S., Irwin, D., Shenoy, P., and Albrecht, J. Preventing Occupancy Detection From Smart Meters. *IEEE Transactions on Smart Grid*, 6(5):2426–2434, 2015.
- [30] Benzi, F., Anglani, N., Bassi, E., and Frosini, L. Electricity Smart Meters Interfacing the Households. *IEEE Transactions on Industrial Electronics*, 58(10):4487–4494, 2011.
- [31] Kelly, J., and Knottenbelt, W. Metadata for energy disaggregation. In *Proc. CDS*. IEEE, 2014.
- [32] Beckel, C., Sadamori, L., Staake, T., and Santini, S. Revealing household characteristics from smart meter data. *Energy* 78:397–410, 2014.
- [33] Beckel, C., Kleiminger, W., Cicchetti, R., Staake, T., Santini, S. The ECO data set and the performance of non-intrusive load monitoring algorithms. *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*. ACM, 2014.
- [34] Castellani A.P., Bui, N., Casari, P., Rossi, M., Shelby, Z., and Zorzi, M. Architecture and protocols for the internet of things: A case study. In *Proc. IEEE PerCom*, pages 678–683, 2010.
- [35] Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., and Wu, A.Y. The analysis of a simple k-means clustering algorithm. *ACM Symposium on Computational Geometry*. ACM Press, pages 100–109, 2000.