

Thamires de Campos Luz

**Deteccção de Novidades em Aparelhos
Eletrônicos através do Monitoramento do
Consumo de Energia**

Sorocaba, SP

19 de Novembro de 2015

Thamires de Campos Luz

Detecção de Novidades em Aparelhos Eletrônicos através do Monitoramento do Consumo de Energia

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação (PPGCCS) da Universidade Federal de São Carlos como parte dos requisitos exigidos para a obtenção do título de Mestre em Ciência da Computação. Área de concentração: Engenharia de Software e Redes de Computadores.

Universidade Federal de São Carlos – UFSCar

Centro de Ciências em Gestão e Tecnologia – CCGT

Programa de Pós-Graduação em Ciência da Computação – PPGCCS

Orientador: Prof. Dr. Fábio Luciano Verdi

Coorientador: Prof. Dr. Tiago Agostinho de Almeida

Sorocaba, SP

19 de Novembro de 2015

de Campos Luz, Thamires

Detecção de Novidades em Aparelhos Eletrônicos através do
Monitoramento do Consumo de Energia / Thamires de Campos Luz. -- 2015.
71 f. : 30 cm.

Dissertação (mestrado)-Universidade Federal de São Carlos, campus
Sorocaba, Sorocaba

Orientador: Fábio Luciano Verdi

Banca examinadora: Katti Faceli, André Ponce de Leon F. de Carvalho

Bibliografia

1. detecção de novidade. 2. fluxo contínuo de dados. 3. energia elétrica.
I. Orientador. II. Sorocaba-Universidade Federal de São Carlos. III. Título.

Ficha catalográfica elaborada pelo Programa de Geração Automática da Biblioteca campus Sorocaba (B-So).

DADOS FORNECIDOS PELO(A) AUTOR(A)

Thamires de Campos Luz

Detecção de Novidades em Aparelhos Eletrônicos através do Monitoramento do Consumo de Energia

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação (PPGCCS) da Universidade Federal de São Carlos como parte dos requisitos exigidos para a obtenção do título de Mestre em Ciência da Computação. Área de concentração: Engenharia de Software e Redes de Computadores.

Prof. Dr. Fábio Luciano Verdi
Orientador - DComp/CCGT/UFSCar

Prof. Dra. Katti Faceli
DComp/CCGT/UFSCar

**Prof. Dr. André Ponce de Leon F. de
Carvalho**
ICMC/USP

Sorocaba, SP
19 de Novembro de 2015

Agradecimentos

Agradeço,

a Deus pelo dom da vida.

aos meus pais, Dirceu e Cristina, por todo amor, dedicação, confiança e encorajamento que sempre me proporcionaram.

ao meu orientador, Prof. Dr. Fábio, e co-orientador, Prof Dr. Tiago, por aceitarem me orientar neste estudo, pelos ensinamentos, conselhos e aprendizados passados.

aos professores da UFSCar pelas aulas e conhecimentos compartilhados.

a todos os colegas de laboratório pelas sugestões e conversas que contribuíram para a melhoria deste trabalho, em especial ao Johannes que me auxiliou inúmeras vezes ao longo do curso.

aos demais colegas de curso, pelas conversas e trabalhos em grupos concluídos.

a todos os meus amigos, em especial à Cássia, Daniel e Lucas por todos os momentos divertidos que compartilhamos ao longo do curso, e por sempre estarem dispostos a me ajudar em todas as ocasiões.

à UFSCar pela infraestrutura oferecida aos alunos.

à CAPES pelo auxílio financeiro.

e a todos aqueles outros que contribuíram diretamente ou indiretamente para a realização deste trabalho.

“Sonhos determinam o que você quer. Ação determina o que você conquista.”
(Aldo Novak)

Resumo

A energia elétrica no Brasil é gerada em sua maioria pelas usinas hidrelétricas. Essas hidrelétricas dependem do volume de água dos reservatórios para a geração de energia e, devido à diminuição dos índices pluviométricos, métodos alternativos de geração com custos mais altos são necessários. Esses custos são repassados para os usuários, encarecendo a conta de energia elétrica. Além disso, podemos atribuir o aumento da conta de energia elétrica também ao consumo exagerado e ao desperdício de energia, que muitas vezes pode não ser notado devido ao aumento ser gerado por falha elétrica ou humana, como o esquecimento da porta da geladeira aberta, por exemplo. Partindo da hipótese de que cada equipamento eletrônico apresenta um padrão no consumo energético, um alerta poderia ser enviado ao usuário caso o comportamento padrão fosse alterado. Sendo assim, neste trabalho é proposto um sistema que realiza detecções nas alterações no comportamento do consumo de energia elétrica dos aparelhos eletrônicos, de forma que alertas sejam enviados para o usuário. Para detectar a alteração de comportamento do consumo de energia elétrica dos aparelhos eletrônicos foram implementados sete métodos consolidados na literatura: Janela Deslizante, *Exponentially Weighted Moving Averages* (EWMA), Agrupamento, Média por Ciclo, Média por Estágio, Distribuição Gaussiana e *Self-Organizing Novelty Detection* (SONDE). Os experimentos demonstraram que os métodos foram eficientes na detecção de novidade em tempo real apresentando taxa de detecção das alterações comportamentais acima de 90% e falso-positivo abaixo de 10%, além de um tempo médio de resposta baixo.

Palavras-chaves: detecção de novidade; fluxo contínuo de dados; consumo energético.

Abstract

Electricity in Brazil is mostly generated by hydroelectric plants that depend on the volume of water in their reservoirs. Due to the fact that rainfall is decreasing, other methods with higher costs to generate energy are required. These costs are passed to users, increasing the energy bill. Furthermore, the wasting of energy and overconsumption also contribute to increase the energy bill. At the same time, the wasting of energy are not noticed by the user. To avoid such wasting, an alert could be sent as soon as an anomalous event is detected. In this way, we propose a system that sends an alert of any novelty detection in energy consumption through the analysis of the methods Sliding Window, Exponentially Weighted Moving Averages, Clustering, Average per Circle, Average per Stage, Gauss Distribution and Self-Organizing Novelty Detection. Results demonstrate that the methods evaluated are efficient in real time detection of novelties, presenting 90% of accuracy and 10% of recall, besides a low delay to send the alert.

Key-words: novelty detection; data stream; energy consumption.

Lista de ilustrações

Figura 1 – Exemplo de funcionamento do método Média por Ciclo.	31
Figura 2 – Exemplo de funcionamento do método Média por Estágio.	32
Figura 3 – Exemplo de funcionamento do método de Distribuição Gaussiana.	33
Figura 4 – Exemplo de funcionamento do método EWMA.	33
Figura 5 – Exemplo de funcionamento da Janela Deslizante.	34
Figura 6 – Exemplo de funcionamento do SONDE.	35
Figura 7 – Exemplo de funcionamento do método Agrupamento por Densidade.	36
Figura 8 – Comportamento padrão do consumo de energia da geladeira.	41
Figura 9 – Comportamento padrão do consumo de energia da cafeteira.	42
Figura 10 – Comportamento padrão do consumo de energia do computador.	43
Figura 11 – Comportamento padrão do consumo de energia elétrica do Eletroforese Capilar.	44
Figura 12 – Média por Ciclo - Ilustração de um ciclo com novidade detectada para a Geladeira.	47
Figura 13 – Média por Ciclo - Ilustração de ciclos com novidades detectadas para a Cafeteira.	48
Figura 14 – Média por Estágio - Ilustração de um estágio com novidades detectadas para a Geladeira.	48
Figura 15 – Média por Estágio - Ilustração do estágio padrão e demais estágios com novidades detectadas para a Cafeteira.	49
Figura 16 – Agrupamento por Densidade - Ilustração de novidade identificada nos grupos definidos para o consumo da Geladeira.	50
Figura 17 – Agrupamento por Densidade - Ilustração das novidades detectadas nos grupos para o consumo da Cafeteira.	50
Figura 18 – Distribuição Gaussiana - Ilustração das novidades detectadas para o consumo da Computador.	51
Figura 19 – Distribuição Gaussiana - Ilustração das novidades detectadas para o consumo do Eletroforese Capilar.	52
Figura 20 – EWMA - Ilustração das novidades detectadas para o consumo do Computador.	52
Figura 21 – EWMA - Ilustração das novidades detectadas para o consumo do Eletroforese Capilar.	53
Figura 22 – Janela Deslizante - Ilustração das novidades detectadas para o consumo do Computador.	54
Figura 23 – Janela Deslizante - Ilustração das novidades detectadas para o consumo do Eletroforese Capilar.	54

Figura 24 – SONDE - Ilustração das novidades detectadas para o consumo do Computador.	55
Figura 25 – SONDE - Ilustração das novidades detectadas para o consumo do Eletroforese Capilar.	56
Figura 26 – Agrupamento por Densidade - Ilustração dos grupos identificado para o consumo do Computador.	56
Figura 27 – Agrupamento por Densidade - Ilustração dos grupos identificado para o consumo do Eletroforese Capilar.	57
Figura 28 – Arquitetura do sistema criado para a detecção das novidades em tempo real.	61
Figura 29 – Tela de cadastro de dispositivos.	62
Figura 30 – Tela de controle de treinamento e monitoramento de dispositivos.	63
Figura 31 – Tela de visualização de consumo médio no período de 10 dias.	63

Lista de tabelas

Tabela 1 – Exemplo da estrutura da tabela de armazenamento da coleta do consumo de energia elétrica.	40
Tabela 2 – Dados sobre as coletas realizadas para cada equipamento eletrônico. . .	43
Tabela 3 – Resultados obtidos pelos métodos adaptados para aparelhos com padrão de consumo cíclico.	58
Tabela 4 – Resultados sumarizados dos métodos analisados para aparelhos com padrão de consumo não cíclico.	58

Lista de abreviaturas e siglas

AM	Aprendizado de Máquina
ARIMA	Auto-Regressive Integrated Moving Average
DN	Detecção de Novidade
DANTE	Detection of Anomalies and Novelities in Time sERies with self-organizing networks
DETECTNOD	DiscrETE Cosine Transform based NOvelty and Drift detection
EWMA	Exponentially Weighted Moving Averages
FA	Falso Alarme
FCD	Fluxo Contínuo de Dados
GAUSS	Distribuição Gaussiana
JD	Janela Deslizante
kWh	Quilowatt-hora
MC	Média por Ciclo
ME	Média por Estágio
NNARIMA	Neural Network Auto-Regressive Integrated Moving Average
ND	Novidade Detectada
NodePM	Novelty Detection Power Meter
OLINDDA	OnLIne Novelty and Drift Detection Algorithm
RNA	Redes Neurais Artificiais
SONDE	Self-Organizing Novelty Detection
TMA	Tempo Médio de Alarme

Lista de símbolos

Δ	Variação
ω	Peso de amostras passadas
ζ	Taxa de aceite
β	Margem de aceite
μ	Média
σ	Desvio padrão
α	Similaridade mínima
D	Deslocamento
I	Intervalo
J	Janela
rad	Raio

Sumário

	Introdução	23
1	FUNDAMENTAÇÃO TEÓRICA	25
1.1	Detecção de novidade	25
1.2	Fluxo contínuo de dados	26
1.3	Aprendizado de máquina	27
1.4	Trabalhos relacionados	28
1.5	Métodos	29
1.5.1	Métodos indicados para aparelhos com consumo cíclico	30
1.5.1.1	Média por Ciclo	30
1.5.1.2	Média por Estágio	30
1.5.2	Métodos indicados para aparelhos com consumo não cíclico	31
1.5.2.1	Distribuição Gaussiana	31
1.5.2.2	<i>Exponentially Weighted Moving Averages</i>	32
1.5.2.3	Janela Deslizante	33
1.5.2.4	<i>Self-Organizing Novelty Detection</i>	34
1.5.3	Método genérico	35
1.5.3.1	Agrupamento por Densidade	35
1.5.4	Parâmetros	36
2	COLETA DE DADOS	39
2.1	Geladeira	40
2.2	Cafeteira	41
2.3	Computador	41
2.4	Eletroforese Capilar	42
3	EXPERIMENTAÇÃO E RESULTADOS	45
3.1	Medidas de desempenho	45
3.2	Experimentação	46
3.2.1	Aparelhos com padrão de consumo cíclico	46
3.2.1.1	Média por Ciclo	46
3.2.1.2	Média por Estágio	46
3.2.1.3	Agrupamento por Densidade	47
3.2.2	Aparelhos com padrão de consumo não cíclico	49
3.2.2.1	Distribuição Gaussiana	49
3.2.2.2	<i>Exponentially Weighted Moving Averages</i>	51

3.2.2.3	Janela Deslizante	53
3.2.2.4	<i>Self-Organizing Novelty Detection</i>	54
3.2.2.5	Agrupamento por Densidade	55
3.3	Avaliação dos experimentos	57
4	SISTEMA PARA DETECÇÃO DE NOVIDADE	61
4.1	Cadastros	61
4.2	Treinamento e monitoramento de dispositivos	62
4.3	Gráfico	62
	Conclusão	65
	Referências	67

Introdução

A energia elétrica no Brasil é gerada predominantemente por usinas hidrelétricas através da água acumulada nos reservatórios. Conseqüentemente, a capacidade de geração de energia está diretamente relacionada com o volume de chuva necessário para abastecer as represas. Em virtude do aquecimento global e da recente diminuição dos índices pluviométricos, a capacidade de geração das usinas hidrelétricas vem reduzindo significativamente. Para evitar a escassez de energia, o governo precisa recorrer a métodos alternativos e mais caros de geração de energia elétrica, como as usinas termoeletricas. Dessa forma, para suprir o crescente aumento nos custos, foi implantado em 2013, o sistema tarifário diferenciado para ajustar a cobrança do consumo de energia elétrica de acordo com a capacidade dos reservatórios. Com isso, a tarifa padrão é acrescida nos períodos em que há escassez de chuva. O sistema tarifário, conhecido como *Bandeiras Tarifárias*, entrou em vigor no ano de 2015 e apresenta três tipos de bandeiras: (a) bandeira verde: quando as condições estão favoráveis para a geração de energia e, portanto, não há nenhuma tarifa adicional; (b) bandeira amarela: quando as condições não estão favoráveis, porém o reservatório não está próximo do seu limite, então uma tarifa adicional de R\$ 0,025 é cobrada para cada quilowatt-hora (kWh); e (c) bandeira vermelha: quando o reservatório está próximo do seu limite a tarifa adicional cobrada para cada quilowatt-hora (kWh) é de R\$ 0,055 (ANEEL... , 2015).

Recentemente, a conta de energia elétrica é uma grande preocupação para a maioria da população, pois além do consumo de energia ter aumentado significativamente nos últimos anos, o valor da energia também vem sofrendo grande reajuste, impactando severamente o orçamento de muitas famílias.

Relatórios recentes demonstram que o consumo de energia elétrica *per capita* no Brasil aumentou cerca de 25% somente nos últimos dez anos (DATA... , 2015), principalmente devido à aquisição e dependência de novos aparelhos eletrônicos. Segundo dados de pesquisas, o consumo energético médio residencial é maior nos períodos de verão (EPE... , 2015), em que pode-se atribuir o maior gasto ao uso do ar condicionado. Contudo, uma parcela desse aumento pode ser creditada a desperdícios resultantes do mau funcionamento dos aparelhos e de falhas humanas, como o simples ato de esquecer a porta da geladeira aberta. Geralmente, esses eventos demoram ou não são notados pelos usuários, contribuindo para o aumento desnecessário e inconsciente da conta de energia elétrica.

Partindo da hipótese de que cada aparelho possui um padrão de consumo, qualquer alteração nesse padrão pode ser entendida como uma *novidade* no comportamento esperado para tal aparelho. Assim sendo, através do monitoramento e análise contínua do

consumo de energia de cada dispositivo, é possível detectar alterações comportamentais e alertar os usuários para que alguma providência possa ser tomada rapidamente.

Objetivos

Este trabalho propõe detectar alterações de padrões no consumo energético dos aparelhos eletrônicos ao longo do tempo. Através da implementação e avaliação de técnicas estatísticas e de aprendizado de máquina aplicados ao monitoramento, e da análise temporal do fluxo contínuo de consumo de energia elétrica de cada aparelho é possível realizar essas detecções, bem como enviar alertas da alteração detectada em tempo real para auxiliar o usuário no processo de economia de energia.

A partir das técnicas implementadas, foi projetado um sistema que realiza duas atividades principais: (a) detecção da novidade; e (b) envio de alarme em tempo real, que consiste no disparo de um e-mail para o usuário.

Estrutura do trabalho

Para facilitar a leitura e compreensão, o restante deste trabalho está organizado da seguinte forma:

- Capítulo 1: apresenta os principais assuntos para o entendimento deste trabalho. Assim sendo, são conceituados brevemente os termos detecção de novidade, fluxo contínuo de dados e aprendizado de máquina. Os métodos avaliados neste trabalho também são detalhados neste capítulo;
- Capítulo 2: é detalhado todo o processo de coletas de dados realizadas;
- Capítulo 3: são apresentados os experimentos realizados, a parametrização utilizada e os resultados obtidos com os métodos;
- Capítulo 4: apresenta o sistema desenvolvido para auxiliar no monitoramento dos dispositivos eletrônicos; e
- Conclusão: as discussões dos resultados, as principais conclusões e possíveis linhas para trabalhos futuros são apresentados.

1 Fundamentação teórica

Neste capítulo é fornecida a fundamentação teórica básica para melhor compreensão do trabalho. Nas Seções 1.1 e 1.2 são apresentados os conceitos básicos e as terminologias presentes na literatura, bem como a justificativa da escolha dos termos utilizados ao longo deste trabalho. Na Seção 1.3 é conceituado brevemente o aprendizado de máquina (AM) e como sua evolução proporcionou a realização deste trabalho. Os principais trabalhos relacionados ao tema são apresentados na Seção 1.4. Os métodos avaliados e como funcionam são apresentados na Seção 1.5.

1.1 Detecção de novidade

De maneira geral, a tarefa de detectar alterações de padrões é estudada desde o século XIV. Inicialmente, soluções genéricas foram propostas com base no uso de métodos estatísticos e, no decorrer dos anos, outras técnicas foram propostas para resolver problemas cada vez mais específicos. De fato, atualmente, grande parte dessas técnicas são empregadas em sistemas especialistas, como por exemplo, na tarefa de detectar invasores em sistemas de cartões de crédito, invasores nas redes de computadores e divergências em diagnósticos médicos (CHANDOLA; BANERJEE; KUMAR, 2009).

Geralmente, a tarefa de detectar alterações em padrões de comportamento pode ser dividida em três categorias: detecção de novidades, anomalias e *outliers* (CHANDOLA; BANERJEE; KUMAR, 2009). A diferença entre elas está relacionada ao fato de que a anomalia e *outliers* são considerados alterações indesejadas, sendo que *outlier* pode, ainda, ser um ruído ou falha de leitura. Por outro lado, a novidade é considerada como um comportamento que pode ser ou não incorporado ao padrão em determinada situação, ou ainda, ser ou não considerado como algo indesejado dependendo da aplicação. Os termos são relacionados e dependendo do contexto podem ser empregados mais de um deles (CHANDOLA; BANERJEE; KUMAR, 2009).

De acordo com Spinosa (SPINOSA, 2008), a detecção de novidade pode ser interpretada como a tarefa de descobrir situações que diferem do padrão esperado, ou como a alteração abrupta de um comportamento determinado, sendo que essa diferença, na maioria dos casos, é algo indesejado. Recentemente, esse tipo de problema vem ganhando destaque em pesquisas científicas, principalmente para evitar defeitos de funcionamento que podem ser originados de pequenas alterações comportamentais.

Neste trabalho, a detecção de alteração no padrão de consumo elétrico foi tratada como um problema de detecção de novidades (DN), pois apenas em algumas situações a

novidade está associada a um evento indesejado, pois mesmo eventos com novos padrões comportamentais do usuário serão detectados da mesma maneira.

Segundo Chandola *et al.* (CHANDOLA; BANERJEE; KUMAR, 2009), os seguintes fatores devem ser levados em conta na detecção de novidades (ou anomalias):

- As definições do padrão e da novidade podem ser imprecisas. Algumas situações próximas do limiar podem ser consideradas como novidades ou não, dependendo da aplicação;
- Em caso de anomalia, quando tratar-se de ataque malicioso, o intruso pode adaptar-se para parecer com um processo padrão, tornando a sua detecção bem mais complexa;
- Em alguns cenários, o padrão pode ser alterado constantemente e, portanto, o conceito do que é novidade também pode ser modificado;
- A ocorrência da novidade pode variar em cada cenário. Por exemplo, a alteração constante em uma análise de exames médicos pode ser considerada como novidade, e em outro cenário pode ser o comportamento padrão esperado; e
- *Outliers* podem ser confundidos com novidades. Contudo, *outliers* geralmente são situações decorrentes de falhas ou erros de leitura.

Além dos fatores citados, deve-se ainda levar em consideração que para os aparelhos eletrônicos a detecção de alteração de comportamento é ainda mais complexa, pois cada aparelho eletrônico possui um padrão próprio de consumo de energia e a análise dos dados deve ser realizada continuamente (FILHO *et al.*, 2014).

1.2 Fluxo contínuo de dados

Detectar novidades com base no consumo de energia elétrica de cada aparelho envolve realizar a análise individualizada e contínua do consumo no decorrer do tempo. Em outras palavras, isso significa que as técnicas propostas para tal propósito necessitam manipular um fluxo contínuo de dados (FCD) de consumo elétrico para cada aparelho envolvido. Segundo Aggarwal (AGGARWAL, 2006), um FCD é uma sequência contínua e infinita de dados que apresenta ou não dependência em relação ao tempo.

Os FCDs podem ser considerados como séries temporais quando apresentam dependência direta com o tempo (PAIVA, 2014). Neste trabalho, utilizaremos a nomenclatura fluxo contínuo de dados, apesar deles também serem séries temporais. De acordo com Babcock *et al.* (BABCOCK *et al.*, 2002), os FCDs possuem as seguintes características:

- Os dados chegam de forma contínua;
- Possuem tamanho ilimitado;
- Uma vez que um exemplo (sinal) é processado, ele não é mais utilizado; e
- Os dados não ficam armazenados em memória.

Recentemente, grande parte dos métodos especialistas empregados para detectar novidades e anomalias são propostos baseados nas técnicas consolidadas de aprendizado de máquina. Ou seja, com esses métodos, atualmente, é possível realizar a detecção de novidades em FCDs provenientes da coleta de consumo de energia de dispositivos eletrônicos, pois no decorrer dos anos as novas técnicas propostas passaram a não mais tratar os dados como discretos e limitados, mas sim, como um fluxo contínuo e com a possibilidade de serem ilimitados (GAMA; GABER, 2007).

1.3 Aprendizado de máquina

Para realizar a detecção de novidade foram analisados os principais métodos estatísticos e de aprendizado de máquina (AM) presentes na literatura. De modo geral, o AM possui o objetivo de estudar algoritmos que encontrem e desempenhem atividades de acordo com alguma experiência aprendida para uma determinada tarefa (MITCHELL, 1997). Métodos de AM geralmente podem ser divididos em três categorias com base no tipo de aprendizado desempenhado (RUSSELL; NORVIG, 2010):

- Supervisionados: um especialista apresenta ao método exemplos com rótulos pré-determinados. Baseado nesses rótulos, o método deve gerar uma hipótese que mapeie novos exemplos para os rótulos já conhecidos;
- Não-supervisionados: o método explora um conjunto de dados a fim de descobrir padrões comuns sem a rotulagem de um especialista; e
- Semi-supervisionados: quando o especialista apresenta ao método alguns exemplos rotulados.

Neste trabalho, somente os métodos não-supervisionados foram utilizados. Para realizar a detecção de novidade com base no consumo de energia elétrica é preciso analisar um conjunto de amostras relacionadas ao tempo, e não cada amostra de forma isolada. Portanto, as amostras não apresentam rótulos.

Antigamente, o AM se concentrava principalmente em problemas com dados estáticos, no qual uma boa parte do conjunto de dados está disponível para a fase de aprendizado (GAMA, 2010). Recentes avanços tecnológicos permitiram que a aquisição de novos

dados fossem feitos em larga escala (AGGARWAL, 2006). Contudo, com o surgimento de novos cenários que demandam análise de dados de forma contínua e em tempo real, novos desafios surgiram devido às limitações computacionais ainda existentes e estudos ainda são necessários (AGGARWAL, 2006).

1.4 Trabalhos relacionados

Os estudos para realizar a detecção de novidades vêm crescendo cada vez mais devido à importância do assunto e dos desafios encontrados na área. Apesar de diversos trabalhos apresentarem estudos relacionados à detecção de novidades e anomalias (CHANDOLA; BANERJEE; KUMAR, 2009; DING et al., 2014; PIMENTEL et al., 2014), esta seção apresenta de forma sucinta apenas os principais trabalhos relacionados à detecção de novidades em FCDs.

Um dos métodos mais simples e tradicionais para realizar previsões em séries temporais é a técnica estatística chamada Janela Deslizante (JD), ou Média Deslizante. Mesmo com a diversidade de métodos disponíveis para realizar tal tarefa, a Janela Deslizante continua sendo uma das mais utilizadas (HANSUN, 2013). Neste contexto, Hansun (HANSUN, 2013) e Wang et al. (WANG; ZHAO; CHANG, 2014) apresentam comparações dos métodos derivados da JD, tais como o *Exponentially Weighted Moving Averages* (EWMA), demonstrando que a JD e suas variações podem apresentar bom desempenho em diversos cenários.

Vários trabalhos presentes na literatura destacam os métodos baseados em Redes Neurais Artificiais (RNA) como sendo os principais métodos atualmente disponíveis para a detecção de anomalias (MARKOU M.; SINGH, 2003). Nesse sentido, Albertini e de Mello (ALBERTINI; MELLO, 2007) propuseram o método *Self-Organizing Novelty Detection* (SONDE), que é um método não-supervisionado baseado em RNA e *Self-Organizing Map*, no qual a detecção da novidade é realizada quando um novo comportamento é exposto ao método e nenhum neurônio é capaz de representá-lo. Em termos de avaliação, os autores utilizaram as bases de dados Biomed, Leukemia e DLBCL. Os resultados obtidos foram comparados aos resultados de outros algoritmos baseados em RNA e demonstraram que o SONDE apresenta melhores resultados. Aguayo e Barreto (AGUAYO; BARRETO, 2008) também utilizaram RNAs para analisar detecções de novidades e anomalias em séries temporais. Os autores apresentaram o projeto *Detection of Anomalies and Novelties in Time sEries with self-organizing networks* (DANTE), no qual são conduzidas comparações de diferentes métodos baseados em RNAs na tarefa de detecção de novidades.

Outra estratégia bastante empregada na literatura é agrupamento dos dados. Estudos demonstram que os métodos de agrupamento vêm apresentando bons resultados na detecção de padrões em séries temporais (LIAO, 2005; NGUYEN; WOON; NG, 2014;

SILVA et al., 2013). Com isso, métodos baseados em agrupamento foram propostos na detecção de novidades, como é o caso do método *OnLine Novelty and Drift Detection Algorithm* (OLINDDA) (SPINOSA, 2008) e *DiscrETE Cosine Transform based NOvelty and Drift detection* (DETECTNOD) (HAYAT; HASHEMI, 2010).

De modo geral, os métodos atualmente existentes e presentes na literatura foram propostos para serem empregados em cenários genéricos. Contudo, alguns trabalhos específicos sobre detecção de novidades com base no consumo elétrico também são encontrados. Chou e Telaga (CHOU; TELAGA, 2014) partem do pressuposto de que a detecção da novidade pode ser realizada caso o valor previsto para o consumo elétrico apresente uma grande diferença do valor real. Os autores realizaram comparações entre os métodos *Auto-Regressive Integrated Moving Average* (ARIMA) e uma versão adaptada e implementada como uma rede neural artificial (NNARIMA).

Filho et al. (FILHO et al., 2014) propuseram o método *Novelty Detection Power Meter* (NodePM) para detectar as novidades presentes no consumo de energia coletado dos aparelhos eletrônicos por sensores de rede sem fio em uma *smart grid*. Os autores compararam os resultados do método proposto com o SONDE (ALBERTINI; MELLO, 2007). Embora o NodePM tenha apresentado melhores resultados com uma base de dados pequena (apenas uma semana de coleta), o SONDE apresentou melhor desempenho em bases de dados maiores (duas ou mais semanas de coleta) (FILHO et al., 2014).

Apesar de existir uma grande variedade de métodos propostos para detectar novidades e anomalias, ainda não há um consenso de qual técnica é a mais adequada, principalmente em cenários que envolvem fluxos contínuos de dados. Este trabalho se diferencia dos demais trabalhos citados anteriormente em três aspectos principais: (a) realizar a detecção de novidades com base no monitoramento do consumo de energia elétrica; (b) aplicação e avaliação de sete métodos distintos para o cenário de energia elétrica; e (c) proposta e implementação de um sistema para enviar alertas para o usuário quando uma novidade é detectada. Foram implementados neste trabalho os principais métodos presentes na literatura, que foram avaliados com base no monitoramento em tempo real do consumo de energia elétrica de dispositivos eletroeletrônicos.

1.5 Métodos

Para detectar automaticamente as novidades no consumo de energia dos aparelhos foram implementados e analisados neste trabalho os principais métodos presentes na literatura aplicados na detecção de novidades em FCDs. Também foram analisadas técnicas estatísticas para a detecção de novidade. Para isso, foram consideradas ambas as categorias de consumo:

- *Métodos indicados para aparelhos com consumo cíclico* – Média por Ciclo e Média por Estágio, sendo que ambos são métodos estatísticos (FELLER, 1968);
- *Métodos indicados para aparelhos com consumo não cíclico* – Distribuição Gaussiana (FELLER, 1968), *Exponentially Weighted Moving Averages* (TAI; LIN; CHEN, 2009), Janela Deslizante (HANSUN, 2013) e *Self-Organizing Novelty Detection* (ALBERTINI; MELLO, 2007); e
- *Método indicados para aparelhos com consumo cíclicos e não cíclicos (genérico)* – Agrupamento por Densidade (SILVA et al., 2013).

Cada um desses métodos é apresentado brevemente nas subsecções a seguir.

1.5.1 Métodos indicados para aparelhos com consumo cíclico

Nesta subsecção são expostos os métodos Média por Ciclo e Média por Estágio que são recomendados para aparelhos com padrão de consumo cíclico.

1.5.1.1 Média por Ciclo

O método nomeado Média por Ciclo (MC) baseia-se na frequência de tempo em que ocorre um ciclo completo de consumo elétrico. Na fase de treinamento, o método identifica o ciclo de acordo com alterações de estágios de consumo. Um ciclo é definido quando esse consumo repete-se com comportamento parecido a algum já apresentado.

Após a identificação do ciclo, o método calcula qual foi a duração padrão, além da média dos valores de consumo. A detecção da novidade ocorre quando um ciclo completo possui um tempo de duração inferior ao tempo de duração padrão acrescido de uma margem de aceite, ou quando a média dos valores do consumo do ciclo for divergente do padrão da média encontrada na fase de treinamento.

A Figura 1 ilustra o funcionamento desse método. Considerando que um ciclo padrão apresenta tempo padrão t . Um novo ciclo é identificado como novidade quando a sua variação de tempo Δt for inferior a n , sendo que n corresponde a t acrescido do parâmetro de margem de aceite β definido pelo usuário. É importante observar que o ciclo identificado como novidade apresentou duração Δt menor que n . A regra pode ser expressa por

$$n = t + \beta, \text{ caso } \Delta t < n, \text{ então novidade detectada.} \quad (1.1)$$

1.5.1.2 Média por Estágio

O método Média por Estágio (ME) é similar ao Média por Ciclo, porém ele verifica a duração e a média de consumo de energia para cada estágio de consumo, ao invés de

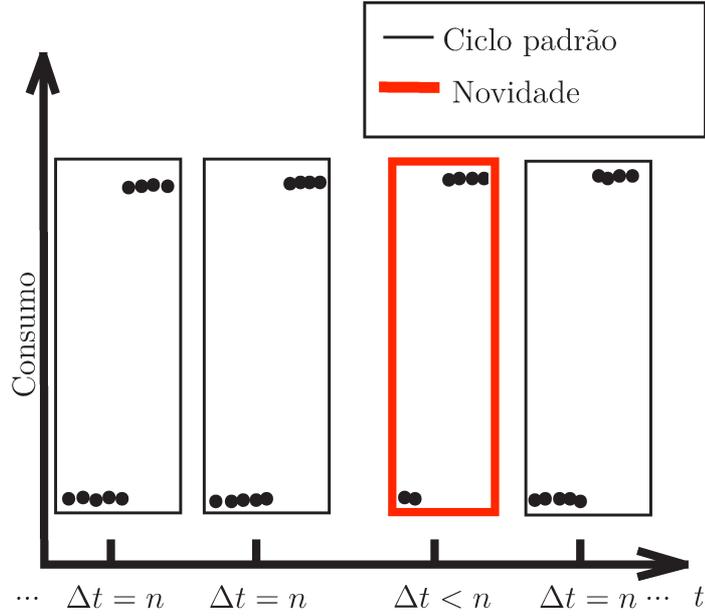


Figura 1 – Exemplo de funcionamento do método Média por Ciclo.

um ciclo completo. Na fase de treinamento é calculada a duração de tempo e a média de consumo de cada estágio.

Para detectar novidades, cada vez que o consumo de energia é alterado para um estágio diferente, o tempo de permanência do estágio é comparado ao tempo padrão de permanência encontrado no treinamento acrescido de uma margem de aceite, para estágios que tenham a mesma média de consumo.

A Figura 2 ilustra o funcionamento do método. Considerando que o estágio padrão de consumo “baixo” apresenta tempo padrão t , o estágio destacado foi identificado como novidade, pois a sua duração Δt foi inferior a n , que corresponde ao t acrescido de uma margem de aceite β , portanto

$$n = t + \beta, \text{ caso } \Delta t < n, \text{ então novidade detectada.} \quad (1.2)$$

1.5.2 Métodos indicados para aparelhos com consumo não cíclico

Nesta subseção são explicados os métodos Distribuição Gaussiana, *Exponentially Weighted Moving Averages*, Janela Deslizante e *Self-Organizing Novelty Detection*, que são recomendados para aparelhos com padrão de consumo não cíclico.

1.5.2.1 Distribuição Gaussiana

A Distribuição Gaussiana (GAUSS), também conhecida como Distribuição Normal, é muito utilizada para detectar anomalias (LIMTHONG et al., 2014). A Distribuição Gaussiana considera que os dados se distribuem levando em conta a média (μ) e o desvio

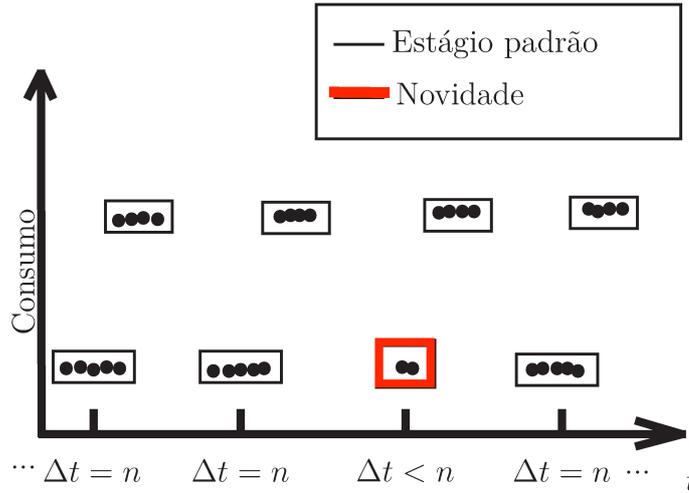


Figura 2 – Exemplo de funcionamento do método Média por Estágio.

padrão (σ), de tal forma que a média se refere ao centro da distribuição e o desvio ao espalhamento dos dados (FELLER, 1968). Através do cálculo de probabilidade da distribuição $f(x)$, é possível identificar uma novidade caso esta probabilidade seja inferior ao parâmetro taxa de aceite ζ , definido pelo usuário. A probabilidade de distribuição é calculada por

$$f(x | \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1.3)$$

A Figura 3 ilustra o funcionamento desse método. Através dos dados de treinamento, o algoritmo calcula a distribuição de probabilidade. Com base nela, os pontos de leitura que ficam fora dessa margem são considerados novidades.

Para o cálculo da probabilidade de distribuição utilizou-se a biblioteca *Scipy* (SCIPY, 2015). Essa biblioteca implementa o método *norm* que retorna a probabilidade de distribuição do dado. Dessa forma, é comparado o retorno do método com ζ . Caso o valor calculado seja menor que a ζ , este será considerado como sendo uma novidade.

$$\text{Caso } f(x | \mu, \sigma) < \zeta, \text{ então novidade detectada.} \quad (1.4)$$

1.5.2.2 Exponentially Weighted Moving Averages

O método *Exponentially Weighted Moving Averages* (EWMA) é uma variação do método *Simple Moving Average* e realiza a previsão do valor seguinte da série temporal Y aplicando pesos (ω) às amostras passadas no intervalo de tempo I , sendo que os dados mais recentes exercem maior influência na previsão (HANSUN, 2013). A detecção de novidade ocorre quando o ponto de leitura diverge do valor previsto S acrescido de uma margem de aceite β . O EWMA pode ser calculado por

$$S_t = (2/w + 1) \cdot Y_t + (1 - (2/w + 1)) \cdot S_{t-1} \quad (1.5)$$

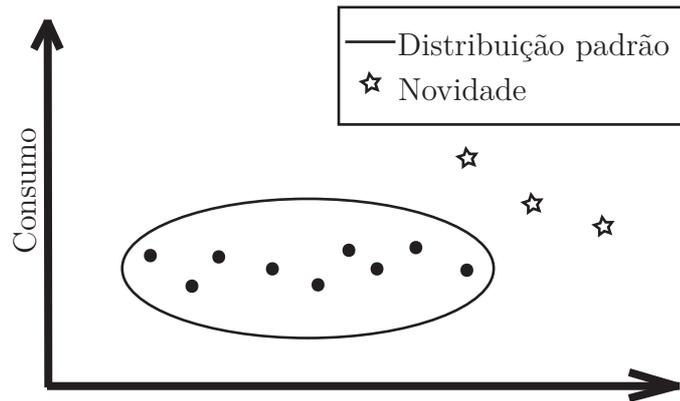


Figura 3 – Exemplo de funcionamento do método de Distribuição Gaussiana.

A Figura 4 apresenta o funcionamento do método. Os pontos de leitura definem uma margem de tolerância que adapta-se conforme os valores de leitura variam.

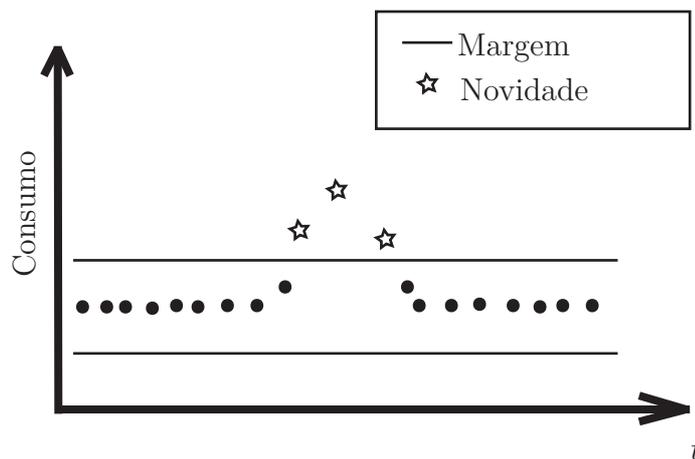


Figura 4 – Exemplo de funcionamento do método EWMA.

1.5.2.3 Janela Deslizante

O método Janela Deslizante (JD) é muito popular na tarefa de previsão de dados que variam ao longo do tempo (HANSUN, 2013). Ele também é conhecido como *Simple Moving Average* ou apenas *Moving Average* (JUN; YING, 2011). Seja uma série temporal S e uma janela de tamanho J que se desloca em k unidades de tempo, a média dos valores da janela é calculada. Assim, a novidade é detectada quando os pontos de uma janela JD qualquer exceder a média padrão (μ) acrescida do parâmetro margem de aceite β . A média é definida pela equação

$$JD = \frac{S_t + S_{t-1} + \dots + S_{t-(k-1)}}{n} \quad (1.6)$$

$$, \text{ caso } JD > \mu + \beta, \text{ então novidade detectada} \quad (1.7)$$

A Figura 5 apresenta o funcionamento do método. A novidade é identificada quando a média dos consumos da janela ultrapassa a margem definida pela média obtida na base de treinamento, acrescida da margem de aceite.

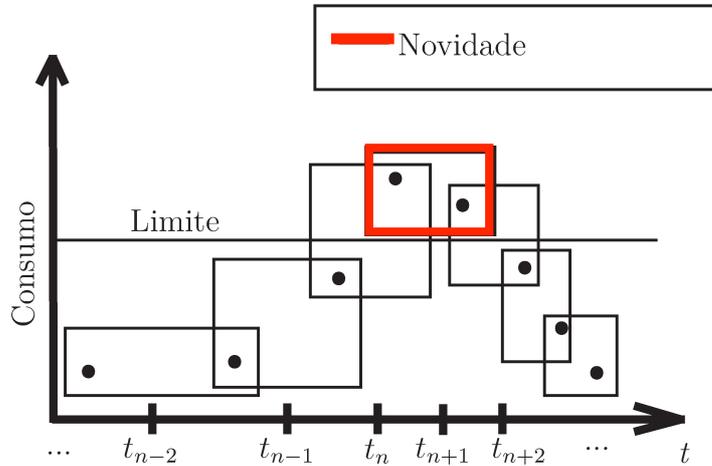


Figura 5 – Exemplo de funcionamento da Janela Deslizante.

1.5.2.4 Self-Organizing Novelty Detection

Na Seção 1.4 são apresentados os principais trabalhos relacionados. O SONDE é comparado com o método NodePM por Filho *et al.* (FILHO *et al.*, 2014). E dentre esses dois métodos o SONDE foi selecionado para implementação neste trabalho, pois Filho *et al.* (FILHO *et al.*, 2014) relata que o SONDE apresentou melhores resultados no decorrer do tempo.

O método *Self-Organizing Novelty Detection* (SONDE) trata-se de uma rede neural artificial baseado em *Self-Organizing Map* para representar os padrões obtidos. A arquitetura do SONDE é composta por três camadas: (a) camada de entrada de dados, (b) camada de ativação do neurônios, em que se calcula a similaridade do dado com os neurônios e (c) camada de saída, em que o neurônio de maior representatividade para o dado de entrada é otimizado para melhor representar os dados. Um neurônio é representado pela média w_c de centróides de seus padrões, a média do raio rad_c ao redor desses padrões e o grau de similaridade mínima α_c .

A ativação do neurônio é definida com base na similaridade do dado de entrada e dos neurônios existentes. Caso a similaridade calculada seja menor que a similaridade mínima α_0 definida pelo usuário, o neurônio passa a representar este dado e o valor da similaridade mínima passa a ser a do próprio neurônio. Esse processo é repetido até que todos os neurônios sejam comparados e o de menor similaridade seja escolhido para representar este dado de entrada. Com isso a média w_c e o raio rad_c são atualizados com a incorporação do valor do dado de entrada.

Dessa forma, os dados que apresentam comportamentos similares são representados por um mesmo neurônio. Quando um novo padrão é apresentado e o método não consegue associá-lo a nenhum neurônio existente, um novo neurônio é criado, caracterizando a novidade (ALBERTINI; MELLO, 2007). A Figura 6 ilustra o funcionamento desse método.

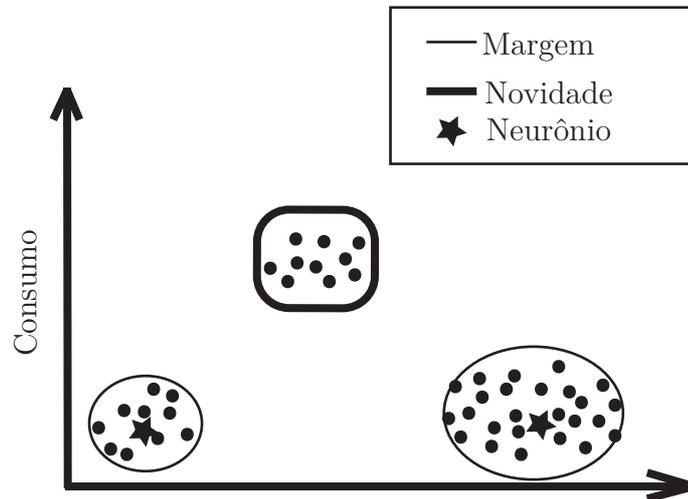


Figura 6 – Exemplo de funcionamento do SONDE.

1.5.3 Método genérico

Nesta subseção é exposto o método Agrupamento por Densidade que pode ser utilizado para detectar novidades em aparelhos que apresentam padrão de consumo cíclico ou não cíclico.

1.5.3.1 Agrupamento por Densidade

Neste trabalho, foi avaliado um método de agrupamento baseado em densidade (LIAO, 2005). Basicamente, os dados de treinamento são agrupados de acordo com o valor da média de consumo, e se diferencia das técnicas tradicionais de agrupamento, pois também leva em conta o tempo no cálculo da centróide do grupo.

Na fase de treinamento, os grupos são definidos de acordo com a similaridade de consumo e a proximidade no tempo. Quando uma nova amostra apresenta a mesma média de consumo das amostras passadas e o tempo passado de leitura é inferior a um limiar, esta amostra é inserida no mesmo grupo das amostras imediatamente passadas.

Um novo *cluster* é criado quando um ponto de leitura apresentar consumo superior ou inferior a média dos consumos do grupo imediatamente anterior, ou se sua proximidade com relação ao tempo for maior que o limiar estipulado. Após definidos os grupos, é

calculada a densidade e o centro de cada *cluster*, sendo que grupos cuja centróide apresenta valor de consumo próximo, também devem possuir densidade similar.

Uma novidade é detectada quando a densidade calculada para um novo grupo apresentar uma diferença com relação as densidades dos grupos com centro de consumo similares, ou se não houver um grupo com a mesma média de consumo. A Figura 7 ilustra o funcionamento desse método.

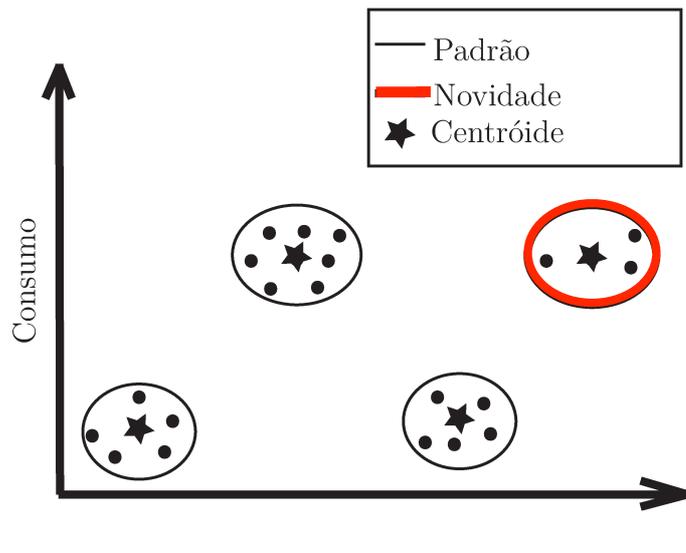


Figura 7 – Exemplo de funcionamento do método Agrupamento por Densidade.

1.5.4 Parâmetros

Cada método implementado possui um conjunto de parâmetros que podem ser ajustados pelo usuário. Na maioria dos casos, esses parâmetros estão relacionados à tolerância dos métodos com relação a alteração de padrão que deverá ser considerada como novidade. Nesta seção é apresentado um resumo dos parâmetros previamente citados, para facilitar a leitura.

- **Métodos indicados para aparelhos com consumo cíclico**

1. Média por Ciclo

- *Margem de aceite* (β): margem que acrescida ao valor padrão de duração de ciclo determina o limiar para se considerar como novidade.

2. Média por Estágio

- *Margem de aceite* (β): margem que acrescida ao valor padrão de duração de um estágio determina o limiar para se considerar como novidade.

- **Métodos indicados para aparelhos com consumo não cíclico**

1. Distribuição Gaussiana

- *Taxa de aceite* (ζ): limiar aceite para determinar se o dado é novidade ou padrão.

2. EWMA

- *Margem de aceite* (β): valor de margem que determina se o dado é uma novidade.
- *Intervalo* (I): intervalo de leitura de amostras passadas que devem ser consideradas.
- *Peso* (ω): quantidade de amostras anteriores que exercem influência na predição do próximo valor.

3. Janela Deslizante

- *Tamanho da janela* (J): valor que determina o tamanho da janela.
- *Deslocamento* (D): valor que determina qual será o deslocamento da janela.
- *Margem de aceite* (β): valor de margem que determina se o dado é uma novidade.

4. SONDE

- *Similaridade mínima* (α_0): valor mínimo de similaridade que o novo ponto deve apresentar quando comparado aos neurônios.

- **Métodos indicados para aparelhos com consumo cíclico ou não cíclico**

1. Agrupamento por Densidade

- *Margem de aceite* (β): margem que acrescida à média de consumo padrão determina o limiar para se considerar do mesmo grupo.

2 Coleta de dados

O principal objetivo desse trabalho é realizar a detecção de novidades e enviar alertas em tempo real para que desperdícios de energia possam ser evitados. Além disso, pretende-se também possibilitar a consulta do histórico de consumo de energia dos aparelhos monitorados.

O desenvolvimento desta pesquisa dividiu-se em duas etapas principais: (a) coletar o consumo de energia elétrica de diferentes aparelhos eletrônicos e (b) implementar e avaliar métodos capazes de processar dados contínuos em larga escala para detecção de novidades.

Para realizar a coleta do consumo de energia dos equipamentos eletrônicos, levou-se em conta que o procedimento deve ser feito de modo muito cuidadoso para não gerar problemas para os métodos, como redundância e falta de dados. Isso porque a coleta de dados não deve ser realizada apenas nos padrões dos usuários, mas também em situações anômalas (FILHO et al., 2014). Portanto, para garantir o que era padrão e o que era novidade, as coletas foram realizadas em dias diferentes para cada situação.

Para realizar a coleta dos dados, foi utilizado um wattímetro *Kill-a-Watt*, comercializado pela empresa P3¹, que foi fundido a uma placa transmissora *ZigBee* (dispositivo de rede sem fio), responsável pelo envio do consumo de energia (em watts).

Na outra ponta, para receber os dados, foi configurado um mini-servidor local com um mini-computador *Raspberry Pi*. A este mini-servidor conectou-se uma placa *ZigBee* em sua porta USB para agir como unidade receptora. Dessa forma, os dados são enviados para o mini-servidor e armazenados em um banco de dados *Postgress*. Cada amostra corresponde a média de consumo de energia elétrica (em watts) de cerca de 17 leituras realizadas em intervalos de dois em dois segundos.

A tabela de armazenamento apresenta a seguinte estrutura de dados: data e hora da coleta, média de consumo (em watts), identificador do coletor e temperatura em graus Celsius. A Tabela 1 apresenta um exemplo de como os dados são armazenados.

De maneira geral, os aparelhos eletrônicos podem ser classificados em duas categorias de acordo com o padrão de consumo de energia elétrica:

- *cíclicos* - aparelhos que durante o uso possuem padrões distintos de consumo que se repetem ao longo do tempo. Os ciclos podem ser bem definidos (quando apresentam apenas duas variações que se alternam) ou diversificado (quando apresentam diversas variações de comportamentos que se intercalam). Como por exemplo a geladeira

¹ Empresa P3: <http://www.p3international.com/products/p4400.html>, acessado em: jan/2015.

(bem definido) e a máquina de lavar roupa (diversificado). Neste trabalho apenas os ciclos bem definidos foram estudados; e

- *não cíclicos* - aparelhos que, quando ligados, apresentam sempre um padrão parecido de consumo ao longo do tempo, como por exemplo um computador e televisão.

Data	Horário	Sensor	Consumo	Temperatura
21-07-2015	12:36:37	1	89.92030356	22
21-07-2015	12:38:02	1	94.1234	22
21-07-2015	12:39:57	1	87.858773	22

Tabela 1 – Exemplo da estrutura da tabela de armazenamento da coleta do consumo de energia elétrica.

Para este trabalho foram monitorados e armazenados o consumo energético de uma geladeira, uma cafeteira, um computador e um equipamento de eletroforese capilar (CAPPILLARY..., 2015). A princípio, coletou-se o consumo de energia de uma geladeira com voltagem de 110 volts categorizada como aparelho cíclico, pois possui um padrão cíclico separado em estágios de consumos altos e baixos em determinados períodos. Ainda, para essa categoria de aparelhos cíclicos, também foi coletado o consumo de uma cafeteira. Após isso, coletou-se o consumo de energia elétrica de um computador para a categoria de aparelhos não cíclicos. E por fim, realizou-se a coleta de um equipamento eletroforese capilar, também pertencente a categoria de aparelhos não cíclicos, que apresenta um consumo de energia elétrica superior quando comparado ao consumo dos equipamentos residenciais.

A seguir são apresentados mais detalhes sobre cada coleta.

2.1 Geladeira

As coletas do consumo de energia da geladeira realizaram-se entre os meses de julho a setembro de 2014. O modelo utilizado foi uma Electrolux R280 110V. Através da análise das coletas de consumo de energia, é notório que ela possui o seguinte ciclo de consumo: (a) o estado de refrigeração dura cerca de 20 minutos com consumo médio de 160 watts; e (b) o estágio de manutenção da temperatura dura aproximadamente 40 minutos com consumo médio de 90 watts. Quando uma novidade ocorre, como por exemplo quando a porta da geladeira é esquecida aberta ou quando há uma sobrecarga excessiva de alimentos, o estágio de manutenção é reduzido (de cerca de 40 minutos para cerca de 20 minutos), fazendo com que o consumo aumente no ciclo.

Neste trabalho, a novidade foi gerada ao deixar a porta da geladeira aberta por duas horas seguidas. Ao dar início ao segundo ciclo na segunda hora da simulação, o

estágio de manutenção foi reduzido para aproximadamente 20 minutos, aumentando seu consumo para o estágio de refrigeração. A Figura 8 ilustra o comportamento padrão de consumo da geladeira.

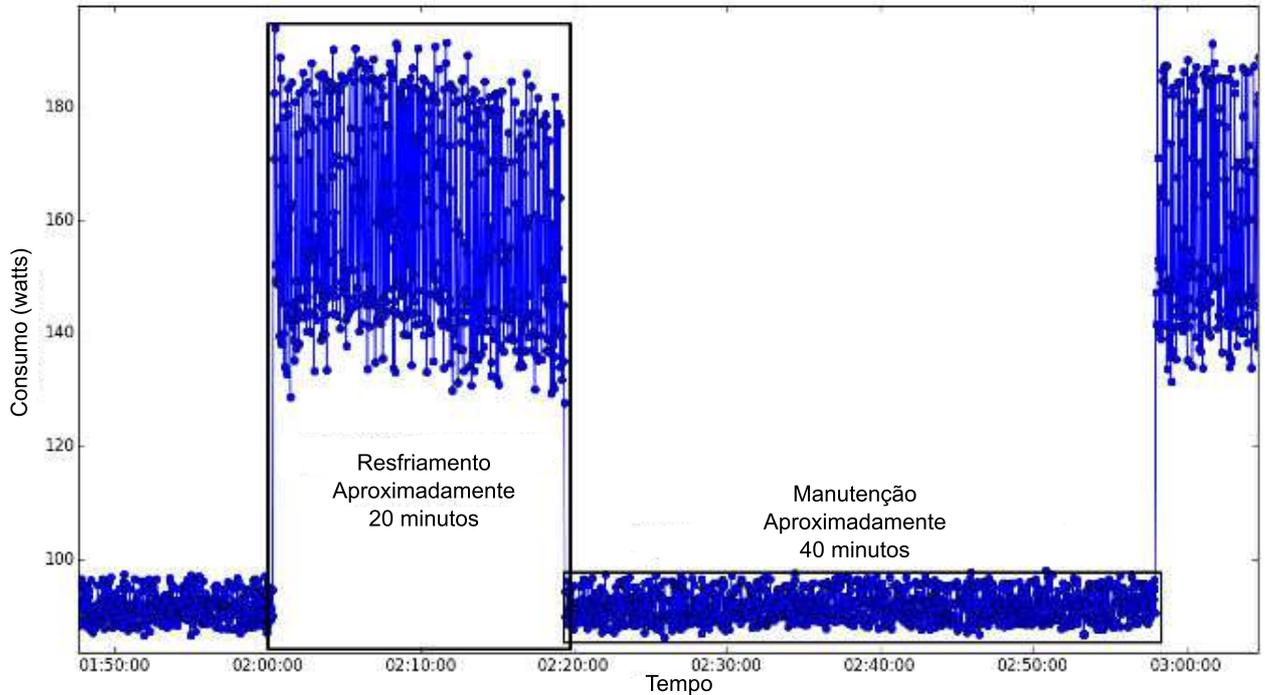


Figura 8 – Comportamento padrão do consumo de energia da geladeira.

2.2 Cafeteira

As coletas do consumo de energia da cafeteira Electrolux CMPRO 110V foram realizadas no mês de setembro de 2015. É possível notar que o consumo padrão do ciclo da cafeteira permanece entre 800 e 1.000 watts. Entretanto, quando a cafeteira é esquecida ligada por um longo período, ciclos de consumos baixos e altos se intercalam, ocasionando a novidade.

A novidade para a cafeteira foi gerada ao deixá-la ligada por cerca de uma hora após seu ciclo padrão. A Figura 9 ilustra o ciclo padrão da cafeteira.

2.3 Computador

As coletas de consumo de energia do computador evidenciam que o padrão médio de consumo é de 95 watts. Quanto maior é a carga de processamento, mais energia é consumida. As coletas realizaram-se no mês de dezembro de 2014 e utilizou-se um Macbook Pro de 13 polegadas Late 2011 para este propósito. O consumo do computador oscila sempre entre uma mesma faixa de consumo até que mais processamento seja demandado.

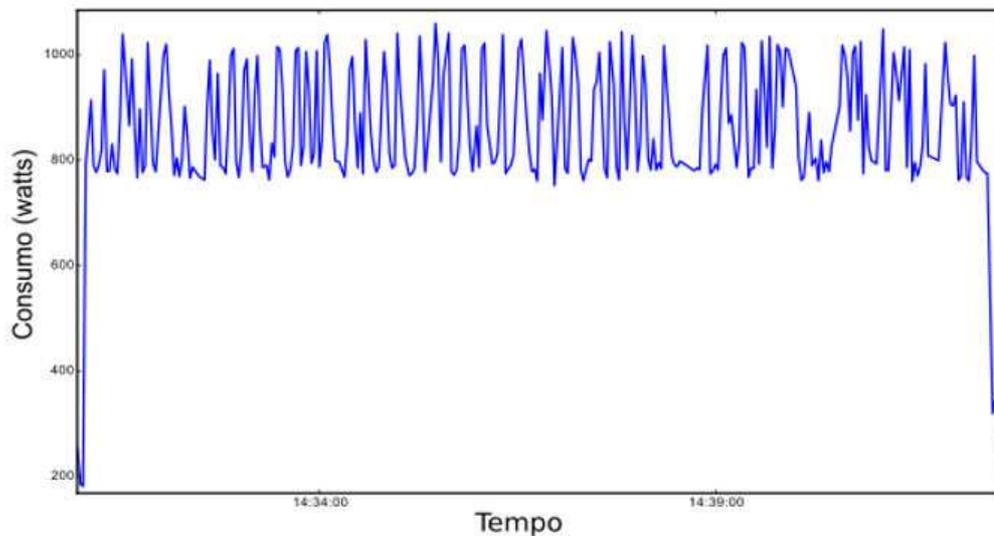


Figura 9 – Comportamento padrão do consumo de energia da cafeteira.

Nessas coletas, primeiramente, foram coletados os consumos de energia com base no uso padrão do usuário, e após certo período foram iniciados programas que exigiam mais processamento do computador, o que resultou em picos no consumo de energia. O pico no consumo de energia teve duração de trinta minutos. A Figura 10 ilustra um trecho de dados com consumo padrão da coleta realizada para o computador.

2.4 Eletroforese Capilar

Foi monitorado um aparelho Eletroforese Capilar Agilent Technologies 7100 de 110v (CAPILLARY..., 2015) utilizado no laboratório de química da UFSCar Sorocaba. Ele tem a função de realizar a separação de biomoléculas e a quantificação desses compostos. O consumo padrão de energia elétrica desse equipamento oscila entre 100 e 350 watts, apresentando uma média de 200 watts. Não foi possível induzir uma novidade real para esse aparelho, devido ao seu uso constante no laboratório. Entretanto, de acordo com o especialista que o manipula, novidades são caracterizadas por picos de consumo de energia, indicando haver uma sobrecarga que possa danificar o aparelho. Dessa forma, foram introduzidas novidades sintéticas, representando picos de consumo de aproximadamente 500 watts. A Figura 11 ilustra um trecho de dados do consumo padrão da coleta realizada para o aparelho.

A Tabela 2 apresenta a sumarização das informações sobre as coletas realizadas.

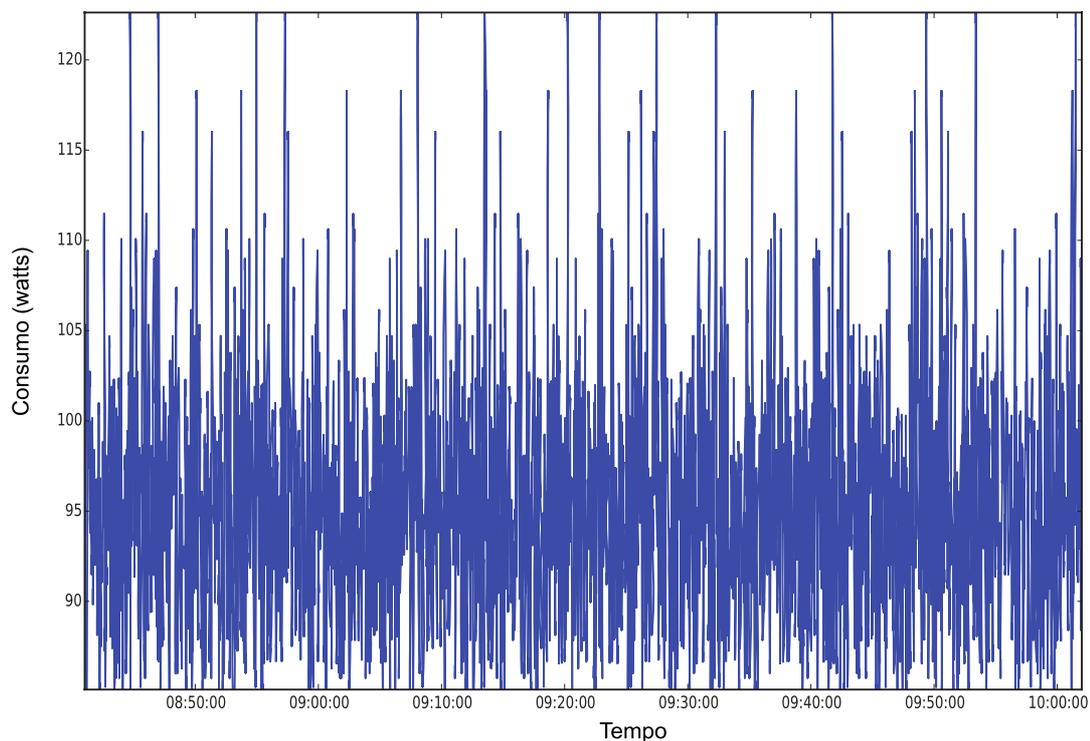


Figura 10 – Comportamento padrão do consumo de energia do computador.

Equipamento	Categoria	Tempo	Tipo de novidade
Cafeteira Electrolux CM-PRO 110V	Cíclico	2 dias	Esquecimento da cafeteira ligada
Geladeira Electrolux R280 110V	Cíclico	3 meses	Esquecimento da porta aberta
Computador Macbook Pro 13 Late 2011	Não Cíclico	2 dias	Sobrecarga de processamento
Eletroforese Capilar Agilent Technologies 7100 de 110v (CAPILLARY..., 2015)	Não Cíclico	2 dias	Sobrecarga criada sinteticamente

Tabela 2 – Dados sobre as coletas realizadas para cada equipamento eletrônico.

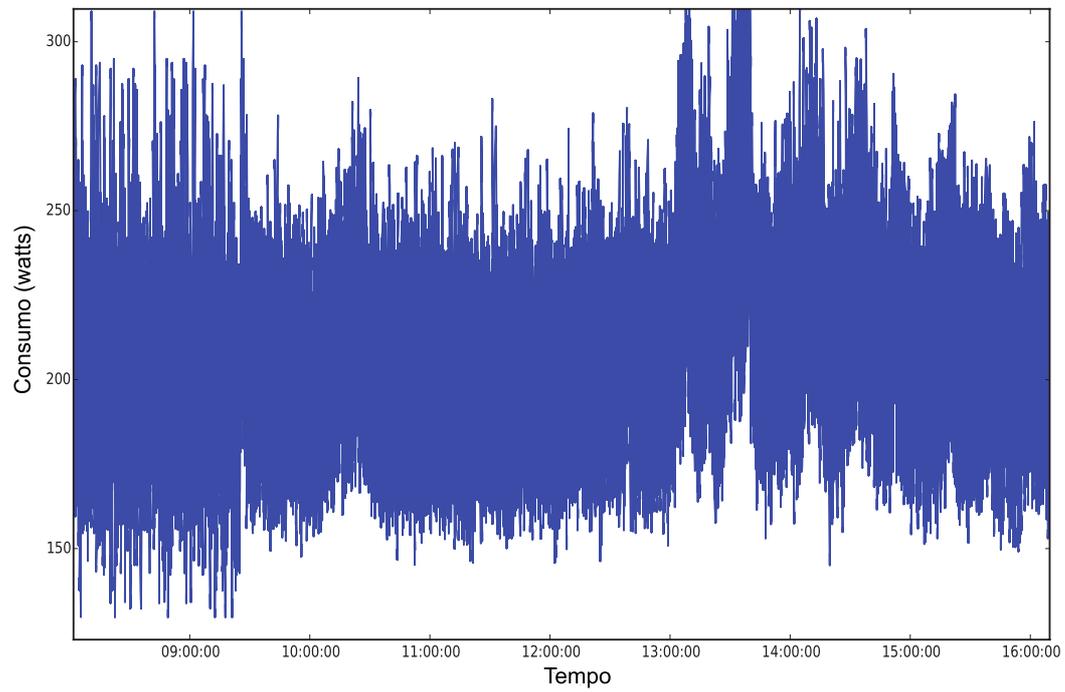


Figura 11 – Comportamento padrão do consumo de energia elétrica do Eletroforese Capilar.

3 Experimentação e resultados

Este capítulo descreve a metodologia utilizada nas experimentações. Os experimentos foram divididos em duas etapas: (a) fase de treinamento e (b) fase de teste. As bases de treinamento contém somente dados de consumo de energia elétrica coletados em um ambiente controlado para cada dispositivo eletrônico. Elas representam o padrão de consumo esperado para cada aparelho, livre de qualquer tipo de novidade. Por outro lado, as bases de teste apresentam leituras realizadas em cenários reais e cotidianos. Esporadicamente, novidades reais foram induzidas para avaliar o desempenho dos métodos implementados.

Os dados obtidos nas coletas realizadas com os aparelhos com padrão de consumo não cíclico foram normalizados para que melhores resultados fossem obtidos. A normalização foi feita por padronização, ou seja, as amostras ficaram com média zero e desvio padrão igual a um. Como a normalização altera a dispersão dos dados, ela não foi aplicada nos dados obtidos pelas coletas com aparelhos que possuem padrão de consumo cíclico, pois a dispersão é uma característica importante para definição dos ciclos.

Na Seção 3.1 são apresentadas as medidas de desempenho utilizadas para comparar os resultados obtidos nos experimentos. Os experimentos realizados são detalhados na Seção 3.2. Os resultados são sumarizados e discutidos na Seção 3.3. As parametrizações foram realizadas empiricamente, pois cada equipamento eletrônico apresenta um padrão próprio de consumo de energia elétrica sendo necessária a parametrização para cada situação. São reportados neste trabalho os melhores conjuntos de parâmetros encontrados nos experimentos realizados.

3.1 Medidas de desempenho

Para avaliar os resultados obtidos pelos métodos, foram utilizadas as medidas mais tradicionais empregadas na literatura:

1. *Novidades Detectadas (ND)*: percentual de novidades corretamente identificadas;
2. *Falsos Alarmes (FA)*: percentual de pontos incorretamente identificados como novidades (falsos positivos);
3. *Tempo Médio de Alarme (TMA)*: tempo médio, decorrido a partir da novidade, que o método levou para enviar o alarme para o usuário.

3.2 Experimentação

Nesta subseção, são reportadas as experimentações conduzidas neste trabalho.

3.2.1 Aparelhos com padrão de consumo cíclico

A seguir são apresentados os parâmetros que obtiveram os melhores resultados dentre uma faixa definida empiricamente, e os gráficos obtidos para cada experimento realizado para os métodos Média por Ciclo, Média por Estágio e Agrupamento por Densidade para os equipamentos Cafeteira e Geladeira.

3.2.1.1 Média por Ciclo

O parâmetro Margem de Aceite (β) foi definido empiricamente entre a faixa de 5% a 20%. O β que apresentou melhores resultados foi de 15%.

Neste experimento, os ciclos que apresentaram duração dentro de uma faixa de 15% aplicado ao tempo padrão de um ciclo completo ou tiveram uma média de consumo no ciclo dentro dos 15% da margem de aceite da média calculada não foram considerados como uma novidade.

Para a Geladeira, é possível observar na Figura 12 que os ciclos completos com comportamento padrão calculados no treinamento duram cerca de 60 minutos e são representados pelos “x”. Portanto, com β de 15% os ciclos que tiveram duração entre 51 minutos e 69 minutos não foram considerados novidades. Os círculos no gráfico da figura demonstram o ciclo com novidade. Como é possível observar, o ciclo que apresenta a novidade teve duração próxima de 36 minutos, sendo esse valor abaixo dos 15% de aceite para o tempo do ciclo completo. Este experimento obteve taxa de ND de 100%, de FA de 0% e TMA de aproximadamente 40 minutos, pois a novidade somente é identificada após o ciclo estar completo.

A Figura 13 ilustra um trecho de coleta realizada para a Cafeteira que apresenta as novidades. O método apresentou taxa de ND de 100% e FA de 0%. O TMA foi de 20 segundos, ou seja, a duração do novo ciclo iniciado.

3.2.1.2 Média por Estágio

Para o método Média por Estágio, 20% foi o valor do parâmetro Margem de Aceite (β) que apresentou melhor resultado dentre a faixa de 5% a 20% definida empiricamente.

A Figura 14 ilustra o gráfico de um trecho do consumo da Geladeira que apresenta a duração média de cada estágio. É possível observar que um padrão encontrado no treinamento tem duração de 40 minutos e está representado entre 20:52 e 21:33, com um consumo próximo de 90 watts. O estágio de refrigeração, por sua vez, é representado

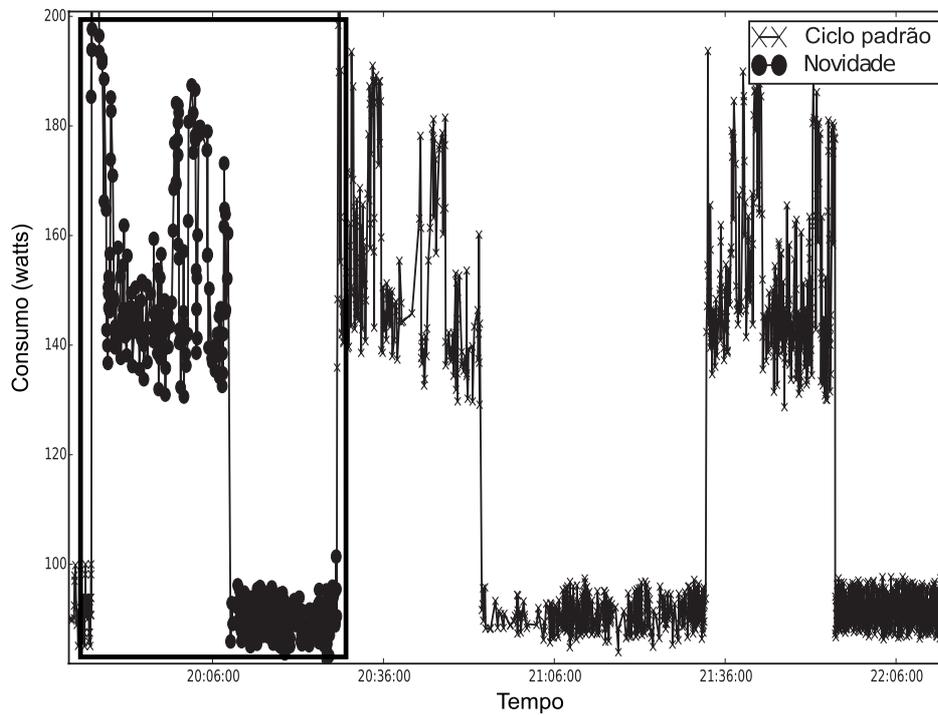


Figura 12 – Média por Ciclo - Ilustração de um ciclo com novidade detectada para a Geladeira.

entre 20:30 e 20:51, com consumo em cerca de 160 watts. A novidade foi identificada corretamente e é representada pelos pontos na figura em que o consumo de energia ficou próximo de 90 watts com a duração média próxima de 20 minutos, entre 20:09 e 20:30. O método apresentou a taxa de ND de 100%, porém, apresentou FA de 3%. O TMA foi de aproximadamente 20 minutos, isto é, o tempo de duração do estágio.

A Figura 15 ilustra um trecho dos estágios de consumo de energia da Cafeteira. O trecho representado por “x” ilustra o estágio de consumo alto. A Cafeteira não apresenta estágio de consumo baixo no seu funcionamento. O método apresentou ND de 80%, pois alguns casos de consumo baixo que não são demonstrados no gráfico foram considerados como padrão. A taxa de FA foi de 0%. E o TMA de 10 segundos, que corresponde ao tempo de duração de mudança do primeiro estágio.

3.2.1.3 Agrupamento por Densidade

O parâmetro que apresentou melhor resultado para os experimentos realizados para o método Agrupamento por Densidade foi de Margem de Aceite (β) de 30%, em uma faixa de 10% a 30%.

A Figura 16 apresenta o gráfico de um trecho de consumo coletado para a Geladeira. As centróides dos grupos encontrados são representadas por estrelas, e os pontos

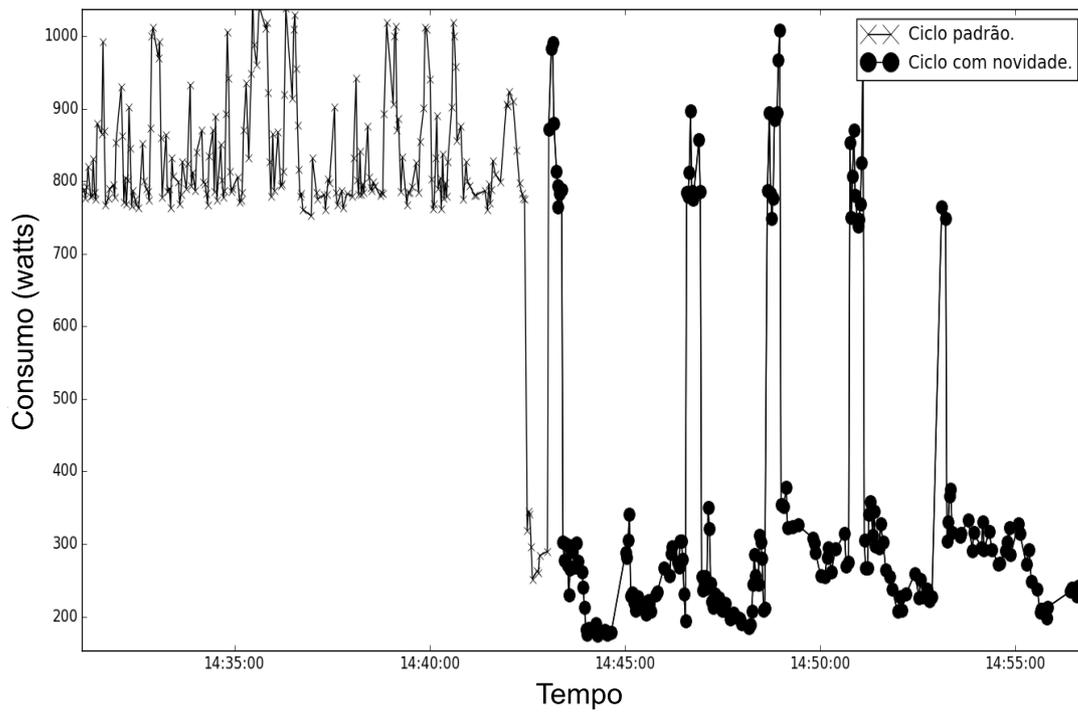


Figura 13 – Média por Ciclo - Ilustração de ciclos com novidades detectadas para a Cafeteira.

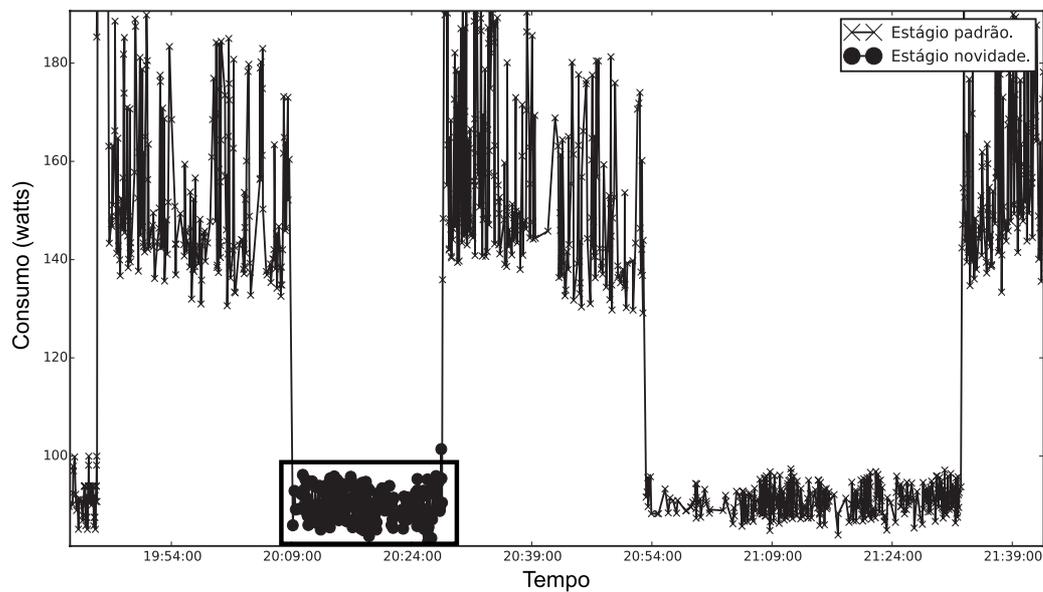


Figura 14 – Média por Estágio - Ilustração de um estágio com novidades detectadas para a Geladeira.

de novidades identificadas são representadas por círculos. Foi detectada uma novidade próximo às 20:30. O grupo padrão com centróide de 90 watts apresenta uma densidade

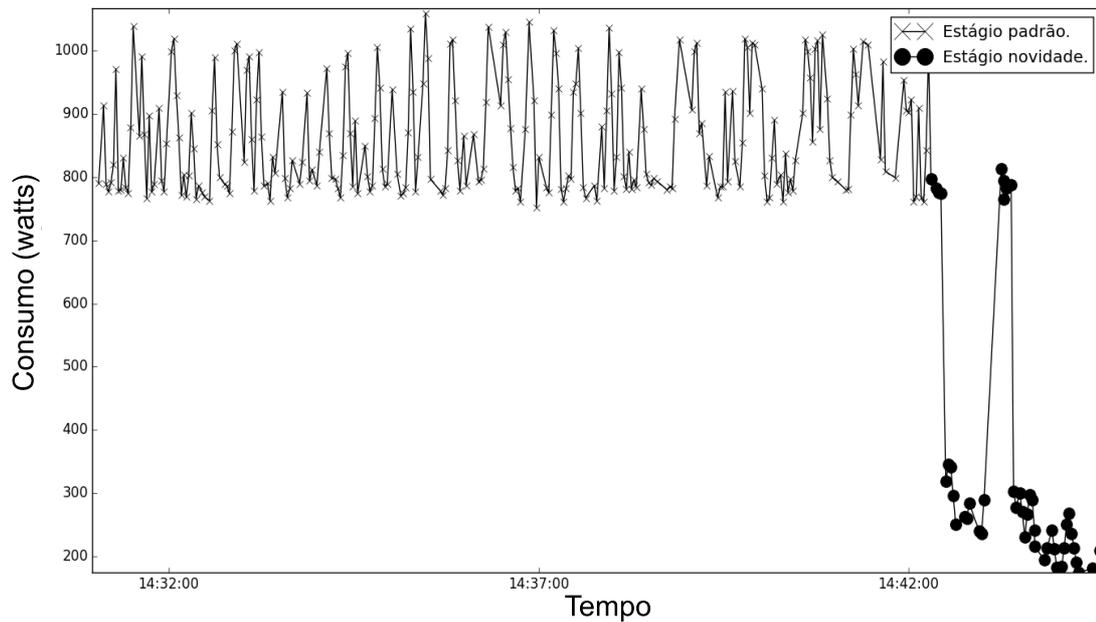


Figura 15 – Média por Estágio - Ilustração do estágio padrão e demais estágios com novidades detectadas para a Cafeteira.

próxima de 1.200 pontos. O grupo apontado como novidade apresenta a mesma média de consumo, porém, sua densidade é de aproximadamente 220 pontos. Este método identificou corretamente as novidades, apresentando taxa de ND de 100%, e uma taxa de FA de 3%. O TMA foi de aproximadamente 20 minutos, isto é, a duração média dos grupos.

Para a Cafeteira, é possível observar as novidades detectadas nos grupos na Figura 17. O método apresentou taxa de ND de 100%, e FA de 3%. O TMA foi de aproximadamente 10 segundos, que corresponde a média de duração dos demais ciclos, pois o alerta somente é enviado a cada mudança de ciclo.

3.2.2 Aparelhos com padrão de consumo não cíclico

Nas subseções seguintes os experimentos conduzidos com os métodos Distribuição Gaussiana, *Exponentially Weighted Moving Averages*, Janela Deslizante, *Self-Organizing Novelty Detection* e Agrupamento por Densidade são apresentados, bem como os parâmetros definidos dentro uma faixa definida empiricamente que apresentaram melhores resultados.

3.2.2.1 Distribuição Gaussiana

O parâmetro Taxa de Aceite (ζ) de 0,001 e 0,00001 apresentaram melhores resultados dentre os experimentos realizados entre uma faixa de 0 a 0,5 para o Computador e o Eletroforese Capilar, respectivamente.

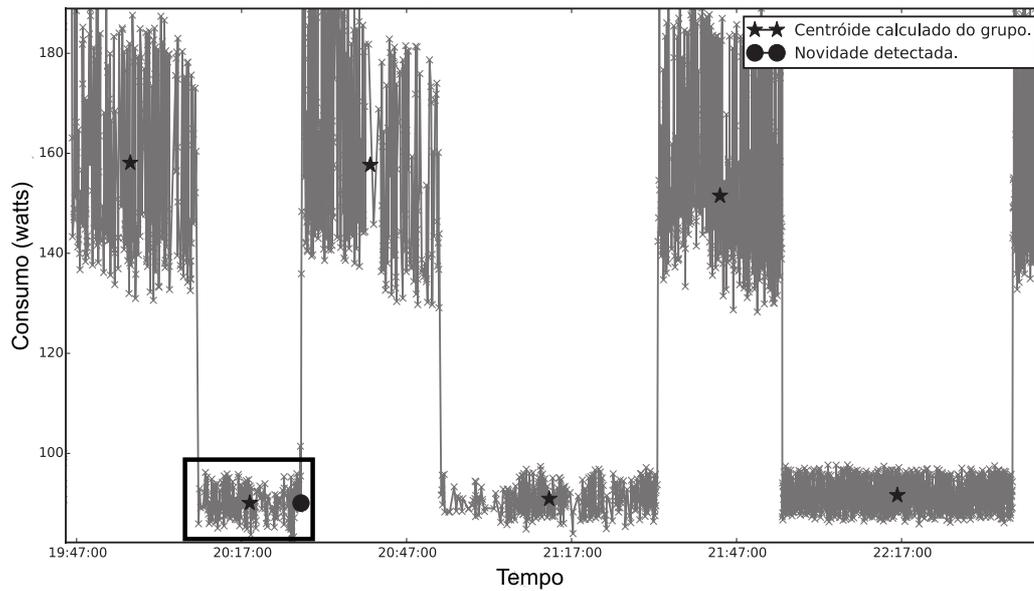


Figura 16 – Agrupamento por Densidade - Ilustração de novidade identificada nos grupos definidos para o consumo da Geladeira.

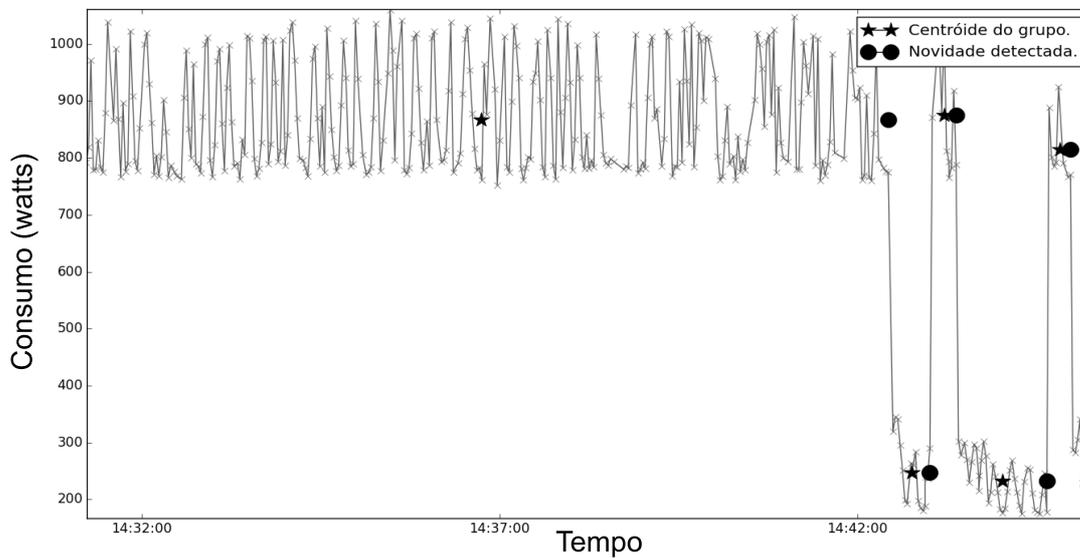


Figura 17 – Agrupamento por Densidade - Ilustração das novidades detectadas nos grupos para o consumo da Cafeteira.

A Figura 18 apresenta o gráfico de um trecho do arquivo de entrada gerado pela coleta realizada com o Computador. É possível notar que os pontos de novidades identificadas, próximas do meio-dia, apresentam a probabilidade de distribuição menor que a taxa de aceite, e por isso são categorizadas como novidades. Este método obteve taxa de ND de 100%, e taxa de FA de 0%. O TMA foi de aproximadamente 1 segundo, ou seja,

assim que o ponto foi calculado.

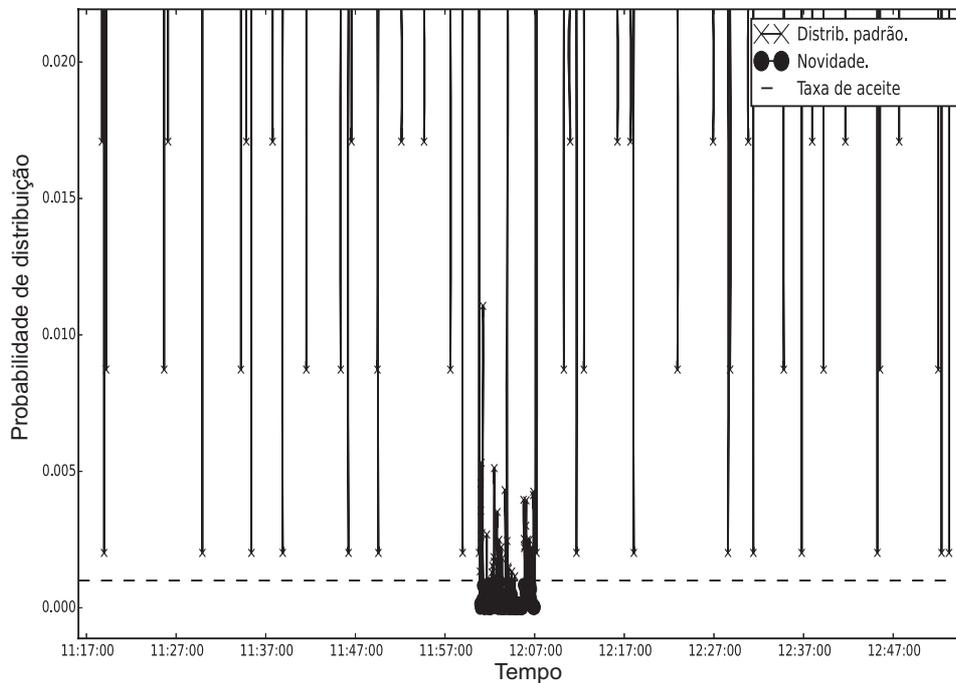


Figura 18 – Distribuição Gaussiana - Ilustração das novidades detectadas para o consumo da Computador.

A Figura 19 ilustra o gráfico para o resultado do experimento realizado no Eletroforese Capilar com a Distribuição Gaussiana. São consideradas novidades os valores que sejam inferiores a ζ de 0,00001. O trecho do gráfico da figura apresenta a detecção realizada corretamente próximo ao meio-dia. O método apresentou taxa de ND de 100% e apresentou um caso de falso-positivo próximo de 13:30, ficando com taxa de FA de 3%. O TMA foi de 1 segundo, pois a detecção é realizada a cada ponto.

3.2.2.2 Exponentially Weighted Moving Averages

Os parâmetros que apresentaram melhores resultados nos experimentos realizados para o método EWMA foram de Margem de Aceite (β) de 1 unidade de desvio padrão, Intervalo (I) de 5 minutos, dentre a faixa de 1 e 10 minutos, e Peso (ω) 5 de amostras passadas, dentre 1 e 10 amostras.

A Figura 20 apresenta um trecho do gráfico obtido com o Computador. Como é possível observar, no gráfico alguns casos de falso-positivos são apresentados nos horários: antes de 5h, entre 17h e 19h, e após às 19h. Entretanto, a detecção da novidade próximo ao meio dia foi realizada corretamente. Portanto, a taxa de ND foi de 100%, e de FA foi 4%. O TMA foi de apenas 1 segundo, pois a identificação é realizada a cada ponto.

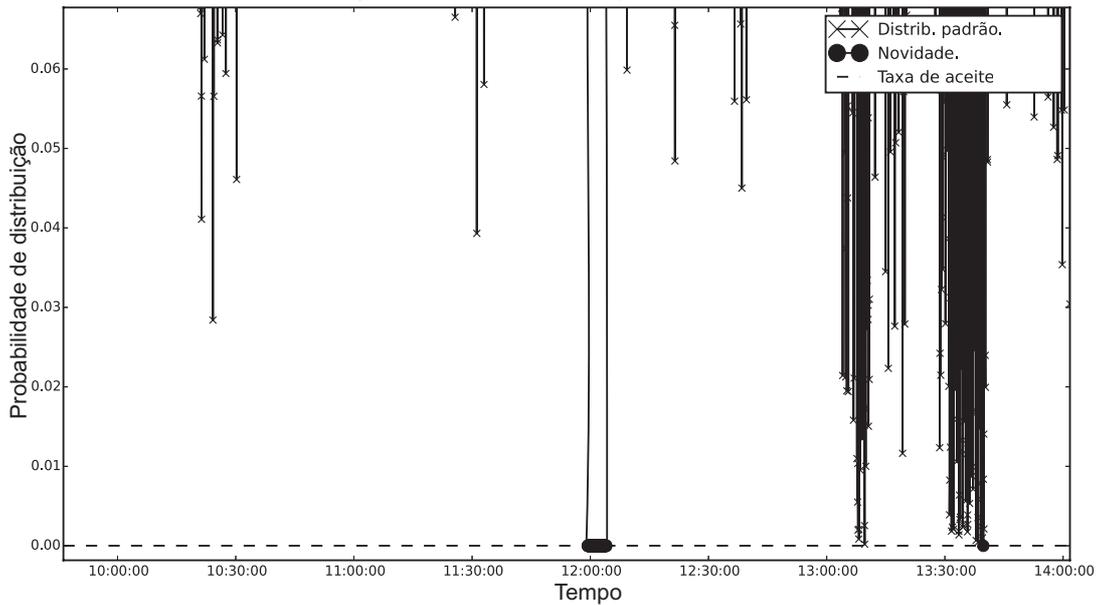


Figura 19 – Distribuição Gaussiana - Ilustração das novidades detectadas para o consumo do Eletroforese Capilar.

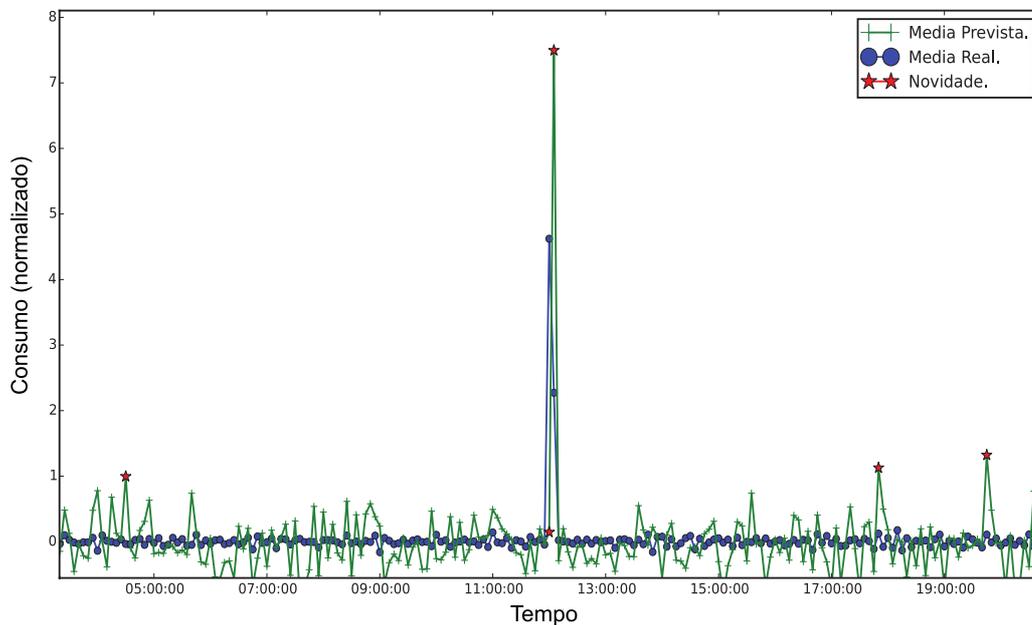


Figura 20 – EWMA - Ilustração das novidades detectadas para o consumo do Computador.

Para o Eletroforese Capilar, a Figura 21 demonstra um trecho do arquivo de coleta que apresenta a novidade. Como é possível observar no gráfico, a dispersão dos dados é mínima, e por isso o método apresentou alguns casos de falso-positivos próximo de 09:30

e entre às 13 horas e 13:30. A taxa de ND foi de 100%, e a de FA de 4%. O TMA é de um segundo, pois a detecção é realizada a cada ponto.

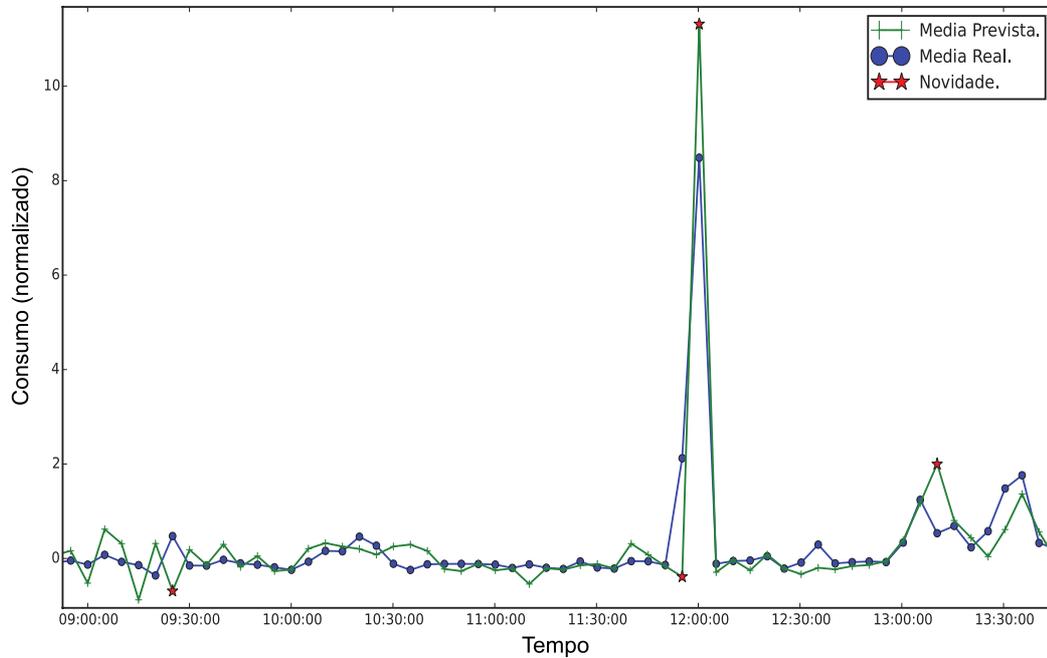


Figura 21 – EWMA - Ilustração das novidades detectadas para o consumo do Eletroforese Capilar.

3.2.2.3 Janela Deslizante

Os melhores parâmetros encontrados nos experimentos realizados para o método Janela Deslizante foram Margem de Aceite (β) de 1 unidade de desvio padrão, Tamanho da Janela (J) de 10 minutos e Deslocamento (D) de 1 minuto. Como é possível observar na Figura 22 para o Computador, os pontos identificados como novidades apresentaram uma dispersão muito grande, sendo superior ao aceite de uma unidade de desvio padrão. A taxa de ND foi de 80%, pois alguns não ultrapassaram o limite estipulado. A taxa de FA foi de 0%, e para esse cenário o TMA foi de 1 minuto, ou seja, o tempo do próprio deslocamento da janela.

No caso do Eletroforese Capilar, um trecho do gráfico é apresentado na Figura 23. A detecção de novidade foi realizada corretamente, porém detectou outros pontos que tiveram um consumo mais alto e não são novidades. A taxa de ND foi de 100%, e a taxa de FA de 5%. O TMA para este cenário depende do tempo de deslocamento da janela, portanto, 1 minuto.

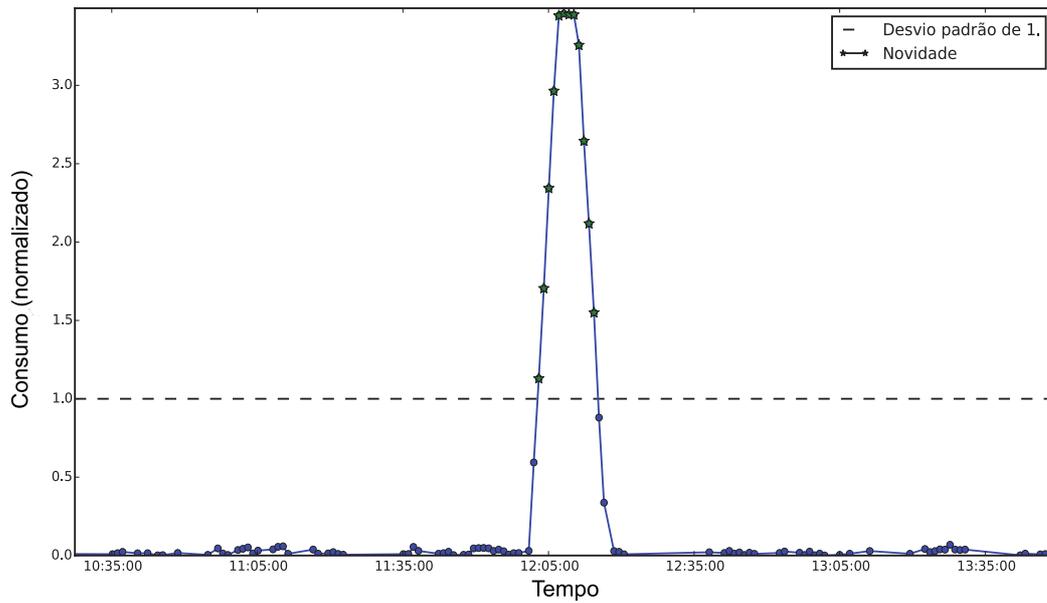


Figura 22 – Janela Deslizante - Ilustração das novidades detectadas para o consumo do Computador.

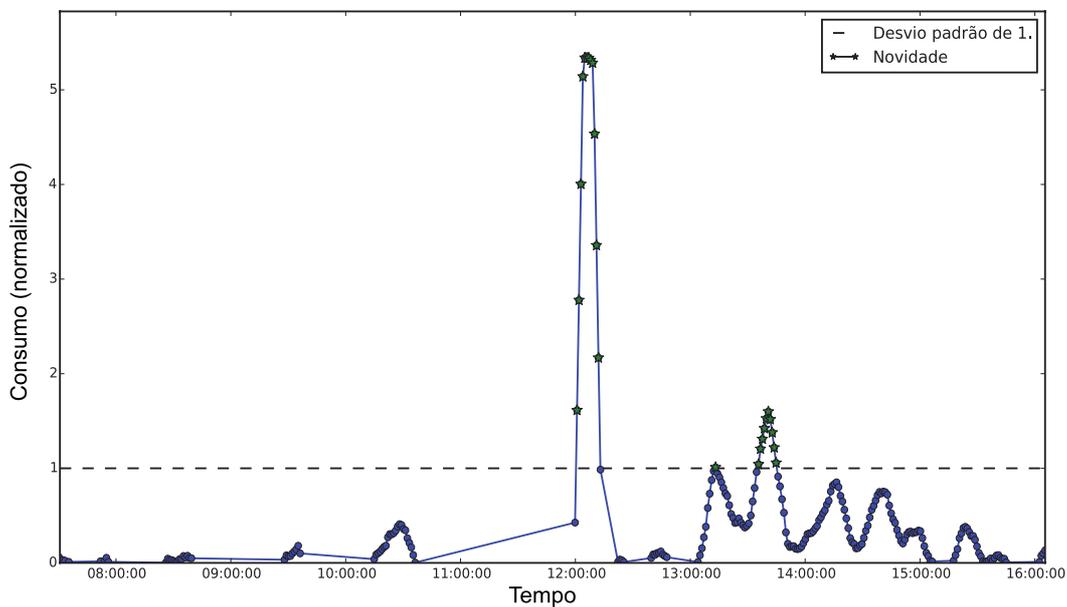


Figura 23 – Janela Deslizante - Ilustração das novidades detectadas para o consumo do Eletroforese Capilar.

3.2.2.4 Self-Organizing Novelty Detection

O melhor parâmetro encontrado dentre 0,0001 e 0,1 nos experimentos realizados para o método SONDE foi de 0,001 de Similaridade mínima (α_0).

Um trecho do gráfico obtido pelo experimento está demonstrado na Figura 24 para o Computador. Como é possível observar, os pontos identificados como novidade não tiveram representação por nenhum neurônio de comportamento padrão, ou seja, a similaridade ultrapassou o limite estipulado. O método apresentou taxa de ND de 100% e FA de 0%. O TMA foi de um segundo, pois a detecção é realizada em cada ponto.

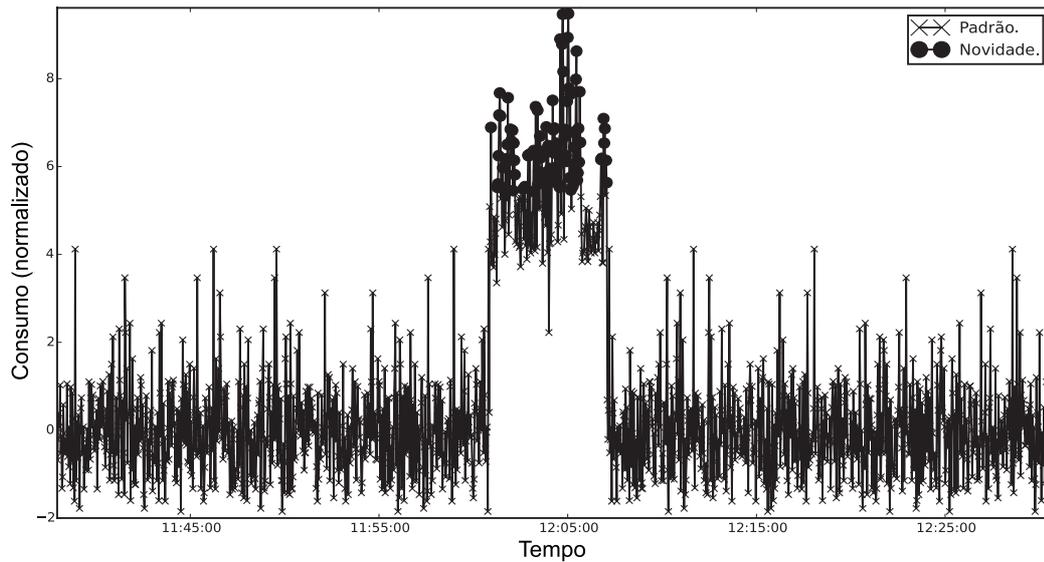


Figura 24 – SONDE - Ilustração das novidades detectadas para o consumo do Computador.

Para o Eletroforese Capilar, o método apresentou a detecção de novidade, próximo ao meio-dia. Um trecho do gráfico que apresenta a novidade é apresentado na Figura 25. O método apresentou taxa de ND de 100%, FA de 0%. O TMA foi de 1 segundo, pois a detecção é feita a cada ponto.

3.2.2.5 Agrupamento por Densidade

O parâmetro que apresentou melhor resultado para os experimentos realizados para o método Agrupamento por Densidade foi de Margem de Aceite (β) de 30%, em uma faixa de 10% a 30%.

A Figura 26 ilustra um trecho do gráfico para o experimento realizado para o Computador. É possível observar na figura que os pontos de picos formam um novo grupo que não tem média de consumo dentro da faixa estipulada e, portanto, é indicado como novidade. O método obteve taxa de ND de 100%. A taxa de FA foi de 0%, e o TMA de 2 minutos, ou seja, a mesma duração do evento novidade.

Um trecho do gráfico obtido com o experimento do Eletroforese Capilar está ilustrado na Figura 27. Observando o gráfico, é possível visualizar que quando o consumo de

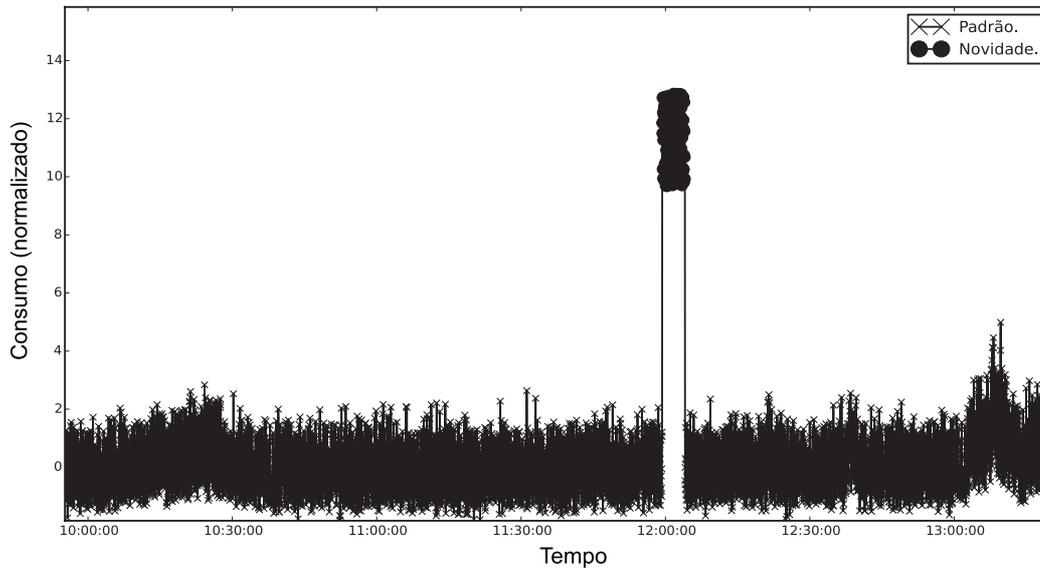


Figura 25 – SONDE - Ilustração das novidades detectadas para o consumo do Eletroforese Capilar.

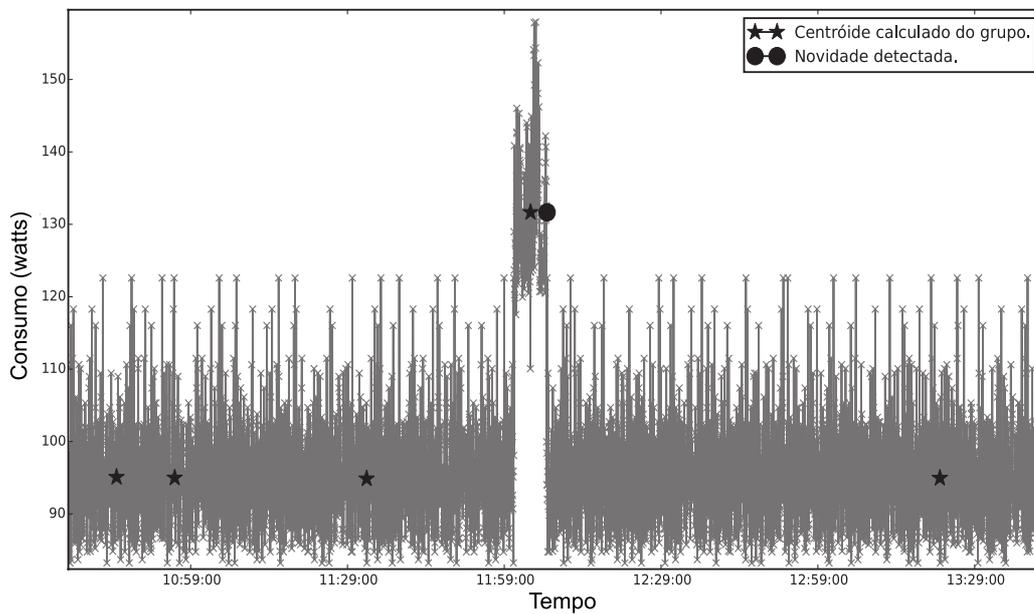


Figura 26 – Agrupamento por Densidade - Ilustração dos grupos identificado para o consumo do Computador.

energia apresenta o pico de consumo de energia próximo ao meio-dia, um novo grupo é formado. Porém, como este pico de energia não está presente no treinamento a detecção de novidade acontece por não haver grupos com esta média de consumo. O método apresentou taxa de ND de 100%, e FA de 0%, entretanto seu TMA depende da duração do

evento novidade, neste caso de 2 minutos.

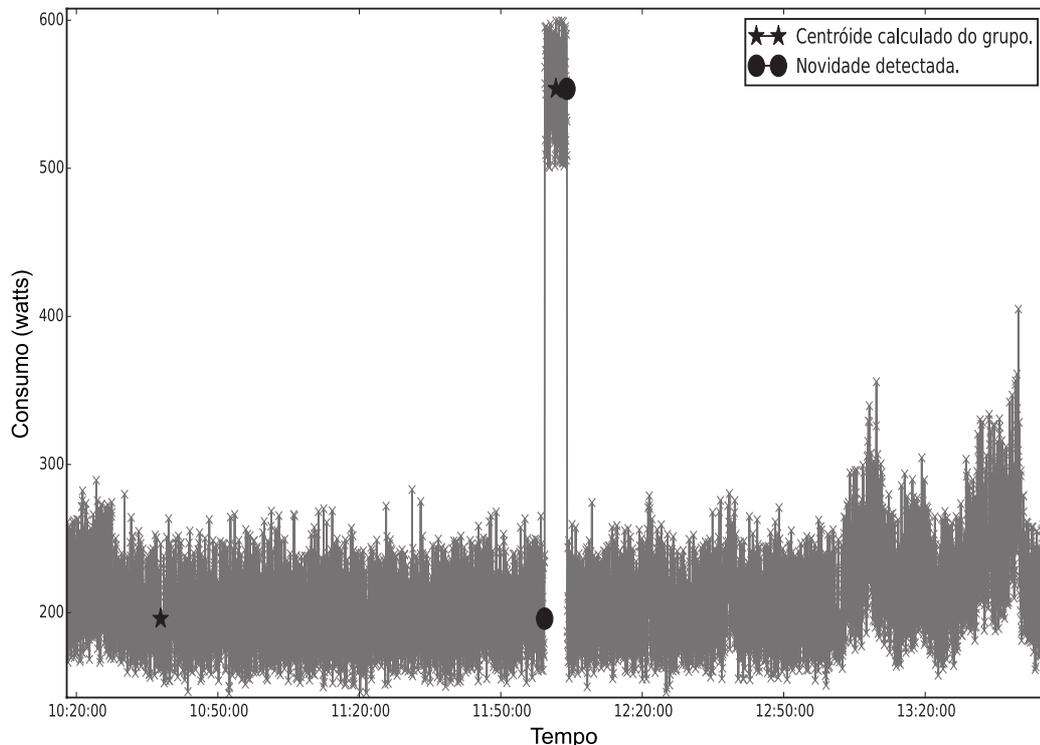


Figura 27 – Agrupamento por Densidade - Ilustração dos grupos identificado para o consumo do Eletroforese Capilar.

3.3 Avaliação dos experimentos

Nesta seção, apresentamos a sumarização dos resultados obtidos. Para facilitar a análise, os resultados estão agrupados de acordo com o padrão de consumo de cada equipamento eletrônico.

A Tabela 3 apresenta os resultados obtidos para os equipamentos que possuem padrão de consumo cíclico usando os métodos de Agrupamento por Densidade (AG), Média por Ciclo (MC) e Média por Estágio (ME). Nesse cenário, apesar dos três métodos avaliados terem conseguido detectar corretamente todas as novidades existentes, o MC apresentou desempenho ligeiramente superior em termos de acurácia, pois não emitiu nenhum falso alarme (falso positivo) para ambos os dispositivos. Contudo, é importante observar que o MC é o método que demora mais tempo para disparar o alerta para o usuário. Isso ocorre porque a detecção da novidade somente é realizada após a conclusão de um ciclo completo. Dessa forma, conclui-se que o MC é adequado para ser empregado no monitoramento de aparelhos que possuem ciclos curtos, como é o caso da Cafeteira. Por

sua vez, em cenários nos quais a novidade ou anomalia precise ser verificada rapidamente, é sugerido que os métodos AG ou ME sejam empregados, pois eles oferecem um bom balanceamento entre tempo de resposta e taxa de falso alarme.

Aparelho	Método	ND (%)	FA (%)	TMA
Geladeira	AG	100	3	20 (minutos)
	MC	100	0	40 (minutos)
	ME	100	3	20 (minutos)
Cafeteira	AG	100	3	10 (segundos)
	MC	100	0	20 (segundos)
	ME	100	0	10 (segundos)

Tabela 3 – Resultados obtidos pelos métodos adaptados para aparelhos com padrão de consumo cíclico.

A Tabela 4 apresenta os resultados obtidos para os equipamentos que possuem padrão de consumo não cíclico usando métodos de Agrupamento por Densidade (AG), Distribuição Gaussiana (GAUSS), *Exponentially Weighted Moving Averages* (EWMA), Janela Deslizante (JD) e *Self-Organizing Novelty Detection* (SONDE).

Aparelho	Método	ND (%)	FA (%)	TMA (seg)
Computador	AG	100	0	120
	GAUSS	100	0	1
	EWMA	100	4	1
	JD	80	0	60
	SONDE	100	0	1
Eletroforese Capilar	AG	100	3	120
	GAUSS	100	3	1
	EWMA	100	4	1
	JD	100	5	60
	SONDE	100	0	1

Tabela 4 – Resultados sumarizados dos métodos analisados para aparelhos com padrão de consumo não cíclico.

De forma geral, todos os métodos avaliados obtiveram bons resultados, com altas taxas de ND (a grande maioria detectou 100% das novidades) e baixas taxas de FA (abaixo de 5%). Contudo, os métodos GAUSS, EWMA e SONDE se destacam pelo baixo TMA (apenas 1 segundo). Isso ocorre porque esses métodos realizam a detecção da novidade ponto a ponto e, portanto, não apresentam atraso na emissão do alerta. O método JD depende do deslocamento da janela para a emissão de alerta, uma vez que a detecção

somente será realizada quando a média da janela no tempo atual for superior a margem estipulada. Nos exemplos do computador e do equipamento de eletroforese capilar, foram utilizadas janelas de 1 minuto, sendo o alerta enviado após esse período. Já o método AG somente realiza o envio do alerta quando o evento de novidade termina, pois sua detecção somente é realizada a cada mudança de grupos identificados e, portanto, seu TMA varia conforme o tempo de duração da novidade.

4 Sistema para detecção de novidade

Um sistema para realizar a detecção de novidade em tempo real foi desenvolvido. O sistema possibilita que a configuração dos parâmetros dos métodos seja realizada pelo usuário através de uma interface web. Os métodos implementados são executados em tempo real no *Raspberry Pi* e consultam continuamente a tabela de coleta para realizar a detecção de novidade. Uma vez que uma novidade é detectada, um e-mail é enviado para o usuário alertando sobre a mudança de comportamento. A Figura 28 ilustra a arquitetura do sistema.

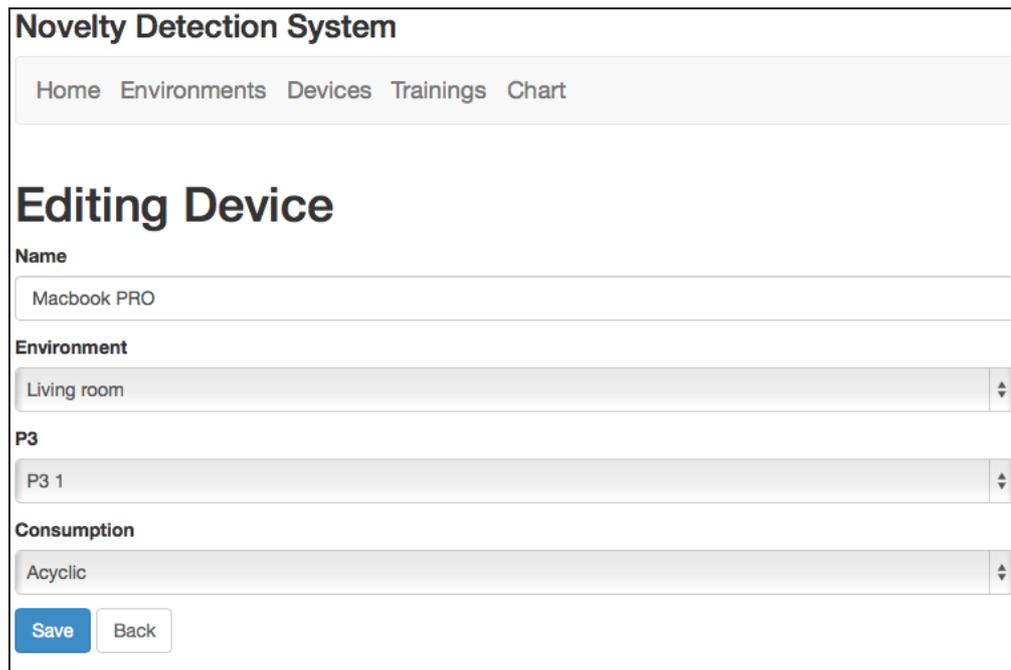


Figura 28 – Arquitetura do sistema criado para a detecção das novidades em tempo real.

Para avaliar o sistema criado, foi utilizado como caso de uso o consumo de energia elétrica do computador e o método Janela Deslizante. Dessa forma, quando o consumo de energia elétrica do computador ultrapassou a média de aceite estipulada para o método, houve a detecção de novidade e o e-mail de alerta foi enviado para o usuário. Não houve a necessidade de validação dos demais métodos pelo sistema, posto que todos foram avaliados nos experimentos realizados no Capítulo 3. A seguir são apresentadas as principais telas que compõe o sistema.

4.1 Cadastros

O sistema apresenta duas telas de cadastros básicos: (a) Ambiente: local onde o dispositivo se encontra (sala, quarto, cozinha); e (b) Aparelho. No cadastro de dispositivo, o usuário deve informar qual a categoria de consumo do dispositivo (cíclico, não cíclico ou genérico). De acordo com essa categoria de consumo o melhor método e as melhores parametrizações são configurados para o dispositivo, porém, o usuário pode alterar esses valores de modo que melhor satisfaça suas necessidades. A Figura 29 apresenta a tela de cadastro de dispositivos.



The screenshot shows a web interface for the 'Novelty Detection System'. At the top, there is a navigation bar with links for 'Home', 'Environments', 'Devices', 'Trainings', and 'Chart'. Below this, the main heading is 'Editing Device'. The form contains several fields: 'Name' with the value 'Macbook PRO', 'Environment' with a dropdown menu showing 'Living room', 'P3' with a dropdown menu showing 'P3 1', and 'Consumption' with a dropdown menu showing 'Acyclic'. At the bottom of the form, there are two buttons: 'Save' (in blue) and 'Back' (in white).

Figura 29 – Tela de cadastro de dispositivos.

4.2 Treinamento e monitoramento de dispositivos

Todos os dispositivos analisados neste trabalho, e todos os treinamentos realizados para as experimentações foram disponibilizados no sistema. Sendo assim, o usuário pode iniciar o monitoramento destes aparelhos que já apresentam o treinamento realizado pelos experimentos.

Para novos dispositivos é necessário realizar o treinamento para determinar os padrões de consumo. Quando o sistema está na fase de treinamento, todos os dados coletados são considerados como padrões e, portanto, o usuário deve garantir que esteja apresentado somente o uso padrão para aquele equipamento.

Caso o usuário deseje alterar seu padrão de consumo, deve-se realizar novo treinamento. Após finalizar esse novo treinamento e dar início a fase de monitoramento somente os dados coletados durante o novo treinamento serão considerados como padrão, sendo os demais treinamentos realizados descartados. Dessa forma, as novidades serão identificadas com o novo padrão aprendido. A Figura 30 ilustra a tela de controle de treinamento e monitoramento de dispositivos.

4.3 Gráfico

Os dados coletados podem ser visualizados em forma de gráfico. O gráfico possibilita a visualização do consumo médio por aparelho, e o usuário pode filtrar por hora, dia ou ano. A Figura 31 ilustra a tela de consulta de consumo médio, demonstrando o

Training module

Device	Training	Monitoring	Initial date	Final date	
Macbook PRO	No	No			<input type="button" value="Start training"/>
Refrigerator	No	No			<input type="button" value="Start training"/>

Figura 30 – Tela de controle de treinamento e monitoramento de dispositivos.

consumo no período de 10 dias.

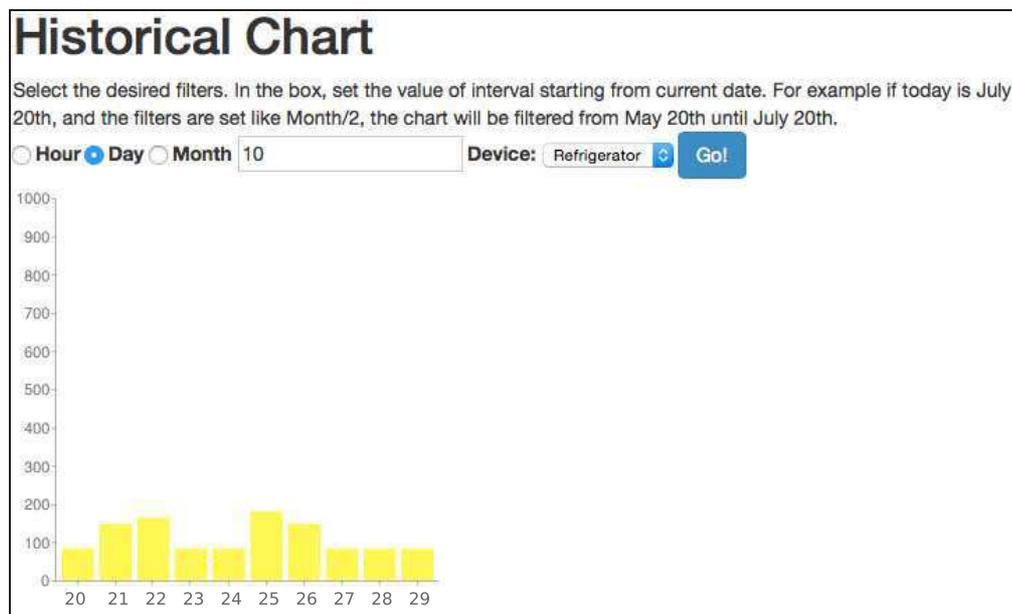


Figura 31 – Tela de visualização de consumo médio no período de 10 dias.

Conclusão

A economia de energia elétrica através de atitudes que auxiliem a evitar desperdícios são de suma importância nos dias atuais. A detecção de novidades com base no monitoramento do consumo energético dos dispositivos eletrônicos pode contribuir para a economia de energia, evitando desperdícios desnecessários. Além disso, através da detecção da novidade, falhas por mau funcionamento podem ser reparadas rapidamente evitando comprometer definitivamente a vida útil do aparelho.

Apesar de existir, na literatura, diversos métodos para detectar anomalias, na maioria dos casos, essas técnicas são adaptadas e especializadas para cada aplicação. Dessa forma, ainda não há um consenso de quais métodos são mais recomendados em cenários genéricos, principalmente nos quais os dados de análise são representados por séries temporais.

Nesse cenário, foram implementados e avaliados sete métodos tradicionais presentes na literatura. O principal propósito foi analisar suas capacidades de detecção de novidades com base no monitoramento contínuo e *online* do consumo de energia elétrica de aparelhos eletrônicos. Os resultados demonstraram que os métodos Média por Ciclo, Média por Estágio e Agrupamento, indicados para monitorar aparelhos com padrão de consumo cíclico, apresentaram alto índice de acerto, e baixa taxa de falso alarme. Entretanto, o método Média por Ciclo deve ser utilizado com cautela, pois, apresentou maior tempo médio de disparo de alarme após a ocorrência da novidade devido a necessidade do ciclo estar completo para a detecção e, portanto, sua aplicação é recomendada em aparelhos com ciclos de consumo de pequena duração.

Os demais métodos, avaliados no monitoramento de aparelhos com padrão de consumo não cíclicos, também apresentaram alta taxa de acerto e poucos falsos positivos. Contudo, os métodos Distribuição Gaussiana, SONDE e EWMA, por terem apresentado baixo tempo médio de alarme, foram os mais eficientes dentre as técnicas avaliadas.

De modo geral, todos os métodos obtiveram bons resultados e comprovaram sua eficiência no cenário empregado. Com a utilização do sistema de detecção de novidade proposto, espera-se que haja economia de energia através de medidas tomadas pelo usuário após recebimento de alertas das novidades detectadas.

Trabalhos futuros

Em relação aos trabalhos futuros, podemos citar:

- Melhoria do sistema: nesta versão, as interfaces para o usuário final estão simples e, portanto, requerem uma melhoria visual e interativa facilitando o uso do sistema;
- Experimentação com novos aparelhos: Neste trabalho foram analisados apenas dois aparelhos por categoria de consumo devido a limitação de apenas um wattímetro P3 estar disponível. Entretanto, é desejável que novos aparelhos sejam incorporados. Uma vez que os algoritmos estão implementados, deve-se aplicá-los no monitoramento de outros aparelhos, bastando para isso realizar a coleta do consumo de energia. Com isso, a base de dados aumenta e novos cenários podem ser considerados;
- Desenvolvimento de hardware próprio: para não depender de um hardware de coleta externo, um novo sensor poderia ser desenvolvido para que seja embutido internamente nos dispositivos;
- Estudo da velocidade do processamento: espera-se que futuramente todos os equipamentos eletrônicos possam divulgar seu consumo de energia. Se isso acontecer, o equipamento receptor dos dados deverá ser capaz de processar uma grande quantidade de dados, e alternativas escaláveis devem ser elaboradas para estes cenários de uso;
- Análise de outros métodos para aparelhos com consumo cíclico diversificado; e
- Estudo sobre o comportamento de aparelhos *stand-by*: ao verificar o padrão de uso por horário espera-se que o usuário seja alertado sobre possíveis economias em horários que o dispositivo não esteja em uso, porém este permanece consumindo em modo *stand-by*.

Produção científica

As seguintes publicações e submissões foram realizadas no período do projeto.

Publicações

Thamires C. Luz, Tiago A. Almeida, Fábio L. Verdi. *Detecção de Novidades em Aparelhos Eletrônicos através do Monitoramento do Consumo de Energia*. XII Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2015), 2015.

Submissões

Aceito para publicação: Thamires C. Luz, Tiago A. Almeida, Fábio L. Verdi. *Towards Novelty Detection in Electronic Devices based on their Energy Consumption*. 14th International Conference on Machine Learning and Applications (IEEE ICMLA), 2015.

Referências

AGGARWAL, C. C. *Data streams: Models and algorithms (advances in database systems)*. 1. ed. [S.l.]: Springer, 2006. Citado 2 vezes nas páginas 26 e 28.

AGUAYO, L.; BARRETO, G. A. Novelty detection in time series through self-organizing networks: An empirical evaluation of two different paradigms. *Neural Networks, Brazilian Symposium on*, IEEE Computer Society, Los Alamitos, CA, USA, v. 0, p. 129–134, 2008. ISSN 1522-4899. Citado na página 28.

ALBERTINI, M. K.; MELLO, R. F. de. A self-organizing neural network for detecting novelties. In: *Proceedings of the 2007 ACM Symposium on Applied Computing*. New York, NY, USA: ACM, 2007. (SAC '07), p. 462–466. ISBN 1-59593-480-4. Disponível em: <<http://doi.acm.org/10.1145/1244002.1244110>>. Citado 4 vezes nas páginas 28, 29, 30 e 35.

ANEEL - Bandeiras Tarifárias - Agência Nacional de Energia Elétrica. 2015. Disponível em: <<http://www.aneel.gov.br/area.cfm?idArea=758>>. Acesso em: agosto 2015. Citado na página 23.

BABCOCK, B. et al. Models and issues in data stream systems. In: *Proceedings of the Twenty-first ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. New York, NY, USA: ACM, 2002. (PODS '02), p. 1–16. ISBN 1-58113-507-6. Disponível em: <<http://doi.acm.org/10.1145/543613.543615>>. Citado na página 26.

CAPILLARY Electrophoresis - Agilent Technologies. 2015. Disponível em: <<http://www.chem.agilent.com/en-US/products-services/Instruments-Systems/Automated-Electrophoresis/7100-CE-System/Pages/default.aspx>>. Acesso em: agosto 2015. Citado 3 vezes nas páginas 40, 42 e 43.

CHANDOLA, V.; BANERJEE, A.; KUMAR, V. Anomaly detection: A survey. *ACM Comput. Surv.*, ACM, New York, NY, USA, v. 41, n. 3, p. 15:1–15:58, jul. 2009. ISSN 0360-0300. Disponível em: <<http://doi.acm.org/10.1145/1541880.1541882>>. Citado 3 vezes nas páginas 25, 26 e 28.

CHOU, J.; TELAGA, A. Real-time detection of anomalous power consumption. *Renewable and Sustainable Energy Reviews*, v. 33, n. 0, p. 400 – 411, 2014. ISSN 1364-0321. Citado na página 29.

DATA - The World Bank. 2015. Disponível em: <<http://data.worldbank.org/>>. Acesso em: janeiro 2015. Citado na página 23.

DING, X. et al. An experimental evaluation of novelty detection methods. *Neurocomputing*, v. 135, n. 0, p. 313 – 327, 2014. ISSN 0925-2312. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0925231213011314>>. Citado na página 28.

EPE - Empresa de Pesquisa Energética. 2015. Disponível em: <<http://www.epe.gov.br/Paginas/default.aspx>>. Acesso em: agosto 2015. Citado na página 23.

- FELLER, W. *An Introduction to Probability Theory and its Applications*. 3rd. ed. [S.l.]: John Wiley & Sons, 1968. Citado 2 vezes nas páginas 30 e 32.
- FILHO, G. P. R. et al. NodePM: A remote monitoring alert system for energy consumption using probabilistic techniques. *Sensors*, v. 14, n. 1, p. 848, 2014. ISSN 1424-8220. Disponível em: <<http://www.mdpi.com/1424-8220/14/1/848>>. Citado 4 vezes nas páginas 26, 29, 34 e 39.
- GAMA, J. *Knowledge Discovery from Data Streams*. 1st. ed. [S.l.]: Chapman & Hall/CRC, 2010. ISBN 1439826110, 9781439826119. Citado na página 27.
- GAMA, J.; GABER, M. M. *Learning from Data Streams: Processing Techniques in Sensor Networks*. 2st. ed. [S.l.]: Springer-Verlag New York Inc, 2007. Citado na página 27.
- HANSUN, S. A new approach of moving average method in time series analysis. In: *New Media Studies (CoNMedia), 2013 Conference on*. [S.l.: s.n.], 2013. p. 1–4. Citado 4 vezes nas páginas 28, 30, 32 e 33.
- HAYAT, M.; HASHEMI, M. A dct based approach for detecting novelty and concept drift in data streams. In: *Soft Computing and Pattern Recognition (SoCPaR), 2010 International Conference of*. [S.l.: s.n.], 2010. p. 373–378. Citado na página 29.
- JUN, W.; YING, Z. A survey and performance evaluation on sliding window for data stream. In: *Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on*. [S.l.: s.n.], 2011. p. 654–657. Citado na página 33.
- LIAO, T. W. Clustering of time series data—a survey. *Pattern Recognition*, v. 38, n. 11, p. 1857 – 1874, 2005. ISSN 0031-3203. Citado 3 vezes nas páginas 28, 29 e 35.
- LIMTHONG, K. et al. Unsupervised learning model for real-time anomaly detection in computer networks. *IEICE Transactions on Information and Systems*, E97-D, p. 2084–2094, 2014. Citado na página 31.
- MARKOU M.; SINGH, S. Novelty detection: a review—part 2: neural network based approaches. *Signal Processing*, v. 88, n. 2499–2521, 2003. Citado na página 28.
- MITCHELL, T. M. *Machine Learning*. 1. ed. New York, NY, USA: McGraw-Hill, Inc., 1997. ISBN 0070428077, 9780070428072. Citado na página 27.
- NGUYEN, H.; WOON, Y.; NG, W. A survey on data stream clustering and classification. *Knowledge and Information Systems*, Springer London, p. 1–35, 2014. ISSN 0219-1377. Citado 2 vezes nas páginas 28 e 29.
- PAIVA, E. R. d. F. *Deteção de novidade em fluxos contínuos de dados multiclasse*. Tese (Doutorado) — Tese de Defesa de Doutorado. Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC-USP), 2014. Citado na página 26.
- PIMENTEL, M. A. F. et al. A review of novelty detection. *Signal Processing*, v. 99, n. 0, p. 215 – 249, 2014. Citado na página 28.
- RUSSELL, S.; NORVIG, P. *The Artificial Intelligence*. 3rd. ed. [S.l.]: Prentice Hall Press, 2010. Citado na página 27.

- SCIPY. 2015. Disponível em: <<http://www.scipy.org>>. Acesso em: agosto 2015. Citado na página 32.
- SILVA, J. A. et al. Data stream clustering: A survey. *ACM Comput. Surv.*, ACM, New York, NY, USA, v. 46, n. 1, p. 13:1–13:31, jul. 2013. ISSN 0360-0300. Citado 3 vezes nas páginas 28, 29 e 30.
- SPINOSA, E. J. *Detecção de novidade com aplicação a fluxos contínuos de dados*. Tese (Doutorado) — Tese de Defesa de Doutorado. Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC-USP), 2008. Citado 2 vezes nas páginas 25 e 29.
- TAI, S.; LIN, C.; CHEN, Y. Design and implementation of the extended exponentially weighted moving average control charts. In: *Management and Service Science, 2009. MASS '09. International Conference on*. [S.l.: s.n.], 2009. p. 1–4. Citado na página 30.
- WANG, Y.; ZHAO, C.; CHANG, C. Anomaly detection using sliding causal windows. In: *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*. [S.l.: s.n.], 2014. p. 4600–4603. Citado na página 28.