UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ENGENHARIA MECÂNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA

CAIO FILIPE DE LIMA MUNGUBA

OTIMIZANDO O PROGRAMA DE MANUTENÇÃO DE SISTEMAS DE REFRIGERAÇÃO POR COMPRESSÃO MECÂNICA ATRAVÉS DE APRENDIZAGEM POR REFORÇO

Recife
2022

CAIO FILIPE DE LIMA MUNGUBA

OTIMIZANDO O PROGRAMA DE MANUTENÇÃO DE SISTEMAS DE REFRIGERAÇÃO POR COMPRESSÃO MECÂNICA ATRAVÉS DE APRENDIZAGEM POR REFORÇO

Dissertação apresentada ao Programa de Pós-Graduação em engenharia mecânica da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de mestre em engenharia mecânica.

Área de concentração: Energia

Orientador: Prof. Dr. Alvaro Antonio Ochoa Villa

Coorientador: Prof. Dr. Gustavo de Novaes Pires Leite

Recife

2022

Catalogação na fonte
Bibliotecário Gabriel Luz, CRB-4 / 2222

M966o Munguba, Caio Filipe de Lima.

Otimizando o programa de manutenção de sistemas de refrigeração por compressão mecânica através de aprendizagem por reforço / Caio Filipe de Lima Munguba. 2022.

121 f: il.

Orientador: Prof. Dr. Alvaro Antonio Ochoa Villa.

Coorientador: Prof. Dr. Gustavo de Novaes Pires Leite.

Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia Mecânica, Recife, 2022.

Inclui referências.

1. Engenharia mecânica. 2. Refrigeração. 3. Degradação. 4. Energia. 5. Manutenção baseada em condição. 6. Aprendizagem por esforço. I. Villa, Alvaro Antonio Ochoa (Orientador). II. Leite, Gustavo de Novaes Pires (Coorientador). III. Título.

UFPE

621 CDD (22. ed.) BCTG / 2023 - 15

Caio Filipe de Lima Munguba

OTIMIZANDO O PROGRAMA DE MANUTENÇÃO DE SISTEMAS DE REFRIGERAÇÃO POR COMPRESSÃO MECÂNICA ATRAVÉS DE APRENDIZAGEM POR REFORÇO

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Mecânica da Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, como requisito parcial para a obtenção do título de Mestre em Engenharia Mecânica. Área de concentração: Energia.

Aprovada em: 06/12/2022.

BANCA EXAMINADORA

_____

Prof. Dr. Alvaro Antônio Ochoa Villa (Orientador)

Universidade Federal de Pernambuco

_____

Prof. Dr. Gustavo de Novaes Pires Leite (Coorientador)

Instituto Federal de Pernambuco

_____

Prof. Dr. José Ângelo Peixoto da Costa (Examinador Interno)

Universidade Federal de Pernambuco

_____

Prof. Dr. Márcio José das Chagas Moura (Examinador Externo)

Universidade Federal de Pernambuco

AGRADECIMENTOS

Agradeço a Deus por todo suporte e oportunidades que me trouxeram até aqui. A meus familiares, que também participaram das alegrias dessa jornada. Ao professor José Carlos Charamba Dutra, que me confiou usar seu laboratório como uma alfândega de cálculos computacionais, aos Doutores Enrique Droguett e Gabriel San Martin, que consentiram empregar seu servidor para acelerar nossos experimentos. E a meus orientadores, que confiando esse desafio, permitiram-me explorar um mundo de pioneiros.

RESUMO

No mundo inteiro, os edifícios são responsáveis por cerca de 30% do consumo energético, e dentre os edifícios, aqueles que necessitam de sistemas de refrigeração, como supermercados e pequenas mercearias, figuram entre os com uso mais intensivo de energia. Dispositivos de refrigeração, sejam comerciais ou residenciais, correspondem por uma grande parcela das emissões do setor de energia. Autores mais conservadores têm estimado que nesse grupo, há um potencial de reduzir o consumo entre 5% e 15% apenas pelo aprimoramento das técnicas de detecção e diagnóstico de falhas. Por isso, aprimorar programas de manutenção através de tecnologias de informação e controle é uma das dimensões necessárias para alcançar metas de emissões e consumo. A oferta de tecnologias como o 5G, agora permitem que as redes suportem o trânsito de milhares de dados de equipamentos interagindo remotamente com algoritmos, como os de aprendizagem por reforço, por exemplo, de forma inteligente e até autônoma, através de interfaces de *Internet of Things* aplicados a *smart building* e *smart cities*. Nesse trabalho, um framework de aprendizagem por reforço foi usado para desenvolver uma política de manutenção para refrigeradores baseados em compressão mecânica. Primeiro, foi construído um *test bench*, que é o ambiente de avaliação do algoritmo de solução, e é constituído do freezer e da sua degradação. Em seguida, a política ótima de manutenção foi encontrada através da solução de um processo de decisão de Markov por um algoritmo de aprendizagem por reforço. Os resultados mostram que a aplicação do modelo de AR ao proposto *test bench* pode reduzir as emissões, o consumo, os custos de manutenção e aumentar a disponibilidade do sistema. Obteve-se que a aplicação da AR é inovadora e apresenta desafios, mas também é promissora frente às técnicas preventiva e corretiva.

Palavras-chave: refrigeração; degradação; energia; manutenção baseada em condição; aprendizagem por esforço.

ABSTRACT

Worldwide, buildings are responsible for 30% of energy consumption, and among buildings, those which intensively use refrigeration systems, such as supermarkets and grocery stores, also figure among the most energy-intensive consumers. Whether commercial or residential, refrigeration devices are responsible for a great part of net emissions. Based on careful measurements, it is possible to reduce energy consumption in these devices by 5% to 15% only by improving the detection and diagnosis techniques of breakdowns. Thus, enhancing maintenance programs has become a crucial area in energy management in recent years. Nowadays, the market has experienced a hike after smart systems, and the offer of new network interfaces applied to smart buildings and smart cities. It has allowed previously isolated devices to become smart devices, interacting with control algorithms smartly and, to some extent, autonomously. As a result, many researchers have applied reinforcement learning to operations and maintenance management. This work used a reinforcement learning framework to develop a maintenance policy for vapor compression refrigeration devices. Firstly, a test bench was built in which each component was assigned to be individually repairable and individually degradable in parallel and interconnected processes. Then, a reinforcement learning algorithm modelled the degradation process via the hidden Markov model and solved it as a Markov decision process. The agent proposed maintenance program for the test bench was successful in reducing emissions, energy use, and maintenance costs while increasing system availability. It was found that the AR frameworks applied to maintenance have a series of challenges but are innovative and can show promising results compared to traditional maintenance techniques, such as preventive and corrective ones.

Keywords: refrigeration; degradation; energy; condition based maintenance; reinforcement learning.

LISTA DE ILUSTRAÇÕES

LISTA DE QUADROS

LISTA DE TABELAS

LISTA DE SIMBOLOS

Dc	Degradação do componente / Intensidade de falha
j	Estado de Markov
C	Custo
T	Instante
r	Recompensa
p	Penalização
π	Política
V	Função valor
γ	Fator de desconto
Q	*Q-Value*
E	Expectativa de retorno com determinada política
P	Probabilidade de transição de $j\ a\ j'$ mediante ação a
R	Expectativa de retorno mediante transição de $j\ a\ j'$
X	Variável aleatória
n	Escalar ou período, instante
$o \mid \Theta$	Observação
w	Perturbação
ε	Vocabulário de símbolos emitidos pelos estados
N	Probabilidade de transitar de $j\ a\ j$'
H	Probabilidade de início
α	Escalar aplicado ao cálculo da política
Y^{Q}	Atualização do objetivo do *Q-learn*
Y^{DQN}	Atualização do objetivo do DQN
$Y^{doubleQ}$	Atualização do objetivo do DDQN
ρ	Redução de intensidade de falha
ζ	Constante de proporcionalidade
τ	Temperatura
q	Taxa de aquecimento

ϕ	Fluxo de calor
ϖ	Escalar aleatório
δT	Índice de atraso
$max\tau$	Temperatura de disparo do compressor
$min\tau$	Temperatura de desligamento do compressor
η	Eficiência com relação a ROP
v	Evaporador
h	Condensador
f	Fluido
g	Compressor
$c\mathrm{F}$	Potência elétrica dos componentes do freezer
Co	Consumo em kWh
Em	Emissões em g de Co_2
Ta	Consumo em €
e	Emissões do freezer
p	Tarifa elétrica
ν	Velocidade de degradação de um componente
κ	Fator de forma da distribuição de Weibull
φ	Degradação linearizada para a vida útil
elc	Vida útil esperada de um componente
ηg	Eficiência do compressor no tempo
βc	Custo de reparo
βr	Custo do técnico
Bc	Custo no instante *T*
ι	Tempo de manutenção
I	Tempo de manutenção
spi	Indicador de temperatura de descarregamento do compressor
f	Tensão
coi	Indicador de consumo
ϑ	Coeficiente de atraso
I	Corrente

ξ	Tempo de acionamento do compressor
Ii	Indicador de corrente
Ir	Corrente na ROP
dti	Indicador de diferencial de temperatura
fai	Indicador de falha
pd	Atraso aceitável antes de computar a falha
ap	Indicador de ação
$n_{treinamento}$	Número de episódios de treinamento
$n_{avaliações}$	Número de episódios de avaliação

LISTA DE ABREVIAÇÕES E SIGLAS

5G	Tecnologia de 5ª geração
IoT	*Ineternet of Things*
AR	Aprendizagem por reforço
AM	Aprendizado de máquina
API	*Aplication programing interface*
DDF	Detecção e diagnóstico de falhas
CBM	*Condition Based Maintenance*
TOP	Condição atual de operação
ROP	Condição de referência de operação
MTTF	*Mean Time To Failure*
BN	*Bayesian Networks*
CNN	*Convolutional Neural Networks*
GAN	*Generative Adversarial Networks*
BNC	*Bayesian Network Classifier*
LR	*Linear Regression*
BP	*Backpropagation Neural Networks*
RBF	*Radial Basis Function Neural Networks*
SVM	*Support Vector Machine*
VRF	*Variable Refrigerant Flow*
PCA	*Principal Component Analysis*
LDA	*Linear Discriminant Analysis*
AHU	*Air Handling Units*
VAV	*Variable air Volume*
ANN	*Artificial Neural Networks*
IFPE	Instituto Federal de Pernambuco
UFPE	Universidade Federal de Pernambuco
DQN	*Deep Q-learning*
DDQN	*Double Deep Q-learning*
HMM	*Hidden Markov Model*

MDP	*Markov Decision Process*
UCLA	Universidade da Califórnia em Los Angeles
CPU	*Central Processing Unit*
GPU	*Graphic Processing Unit*
TDP	*Thermal Design Power*
kG	Kilo Gramas
kWh	Kilo Watt Hora
CO_2	Gás carbônico

SUMÁRIO

1 INTRODUÇÃO

1.1 REFRIGERAÇÃO E INTERNET DAS COISAS, UMA JORNADA A CAMINHO DE MELHORES EXPERIÊNCIAS

Na era da Indústria 4.0, as tecnologias da Internet das Coisas (IoT) são aplicadas aos mais diversos sistemas. O objetivo da IoT é abrir canais de comunicação adicionais com o mundo físico, fazer conexões objeto-objeto e pessoa-objeto, e realizar percepções, identificações e gerenciamento inteligente de processos utilizando esses acessos à rede. Por exemplo, um dos propósitos dessas redes é o compartilhamento de informações entre o usuário final, o fornecedor do produto, o fabricante do equipamento e seu mantenedor. O usuário pode receber informações sobre o estado de seus equipamentos, aqui um freezer, enquanto o fornecedor ou construtor/mantenedor também pode acompanhar informações como temperatura interna e umidade, por exemplo (DONG, Z. *et al.*, 2020).

No setor de refrigeradores, por exemplo, o foco da IoT é atender principalmente a conveniência e conforto dos usuários. Estes dispositivos inteligentes costumam interagir com o usuário sobre os materiais embalados, coletar dados relevantes e emitir alertas antecipados sobre manutenção, ajudando-o a operá-los corretamente. A IoT geralmente utiliza uma inteligência artificial para monitorar o dispositivo e seus produtos com os dados obtidos da rede. Atualmente, é possível que dispositivos simples e de baixo custo, com o auxílio de técnicas de aprendizado de máquina, possam informar o desempenho do equipamento e auxiliar nas tarefas de automação aplicadas à operação e manutenção em ambientes residenciais e comerciais (CHANG *et al.*, 2020a).

1.2 REFRIGERAÇÃO, MANUTENÇÃO E ENERGIA

Em sistemas de refrigeração, as faltas costumam acontecer ao longo do tempo, normalmente iniciadas pela degradação de componentes eletrônicos. Mas quando estados de falha se tornam notáveis, seus níveis de severidade costumam ser elevados (ZHONG, C. *et al.*, 2019). A degradação de sistemas de refrigeração costuma ser ignorada até que resulte em algum impacto significativo no conforto ou performance do sistema, ative algum mecanismo de proteção ou resulte em excessivo

consumo energético (DEY; RANA; DUDLEY, 2018). Trazendo a números, de acordo com a pesquisa realizada por Knowles & Baglee (2012), se fosse possível diagnosticar estados de falha prematuramente, em seus estágios iniciais, poder-se-ia reduzir demandas energéticas de sistemas de refrigeração entre 10% e 35% no reino unido. E além, Behfar *et al* (2017b) argumenta que os custos de manutenção e operacionais associados às operações de supermercados e os lucros líquidos estreitos, da ordem de 1% a 3% (FMI, 2022), chamam cada vez mais atenção para o uso de métodos de detecção e diagnóstico de falhas baseados em estado.

Outro aspecto pode ser citado do estudo publicado por (DAVENPORT; QI; ROE, 2019), que observou que a manutenção dos refrigeradores pode ter peso significativo para a segurança e hábitos alimentares. Em um estudo piloto, foi encontrado que esquecer a data de vencimento é a maior causa de desperdício, mas a temperatura interna do refrigerador também é importante. Davenport *et al* (2019) então argumenta que reduzir em 3ºC a temperatura interna pode salvar cerca de 162.4 milhões de libras em perecíveis por ano. O mesmo trabalho também argumenta que há correlação direta entre o tempo de conservação e a temperatura interna, e que por isso, a oportunidade de rastreá-la e estabilizá-la ao longo da vida útil pode ser relevante para a otimização do armazenamento.

Todavia, mais métodos têm sido desenvolvidos para estudar sistemas de climatização do que refrigeração Wichman & Braun (2009), estes equipamentos, embora similares, podem ser até mais complexos em muitos aspectos, como por exemplo a carga térmica ou de refrigerante. Citando um estudo realizado por Assawamartbunlue & Brandemuehl (2006), estima-se que entre 15% e 20% de cargas de refrigerante se perdem todo ano no mercado de refrigeradores exclusivamente, o que implica que não só equipamentos como também as cargas, o que não ocorre para sistemas de climatização, estão à mercê da proatividade de profissionais dispostos a realizar o programa de manutenção e os diagnósticos corretos.

Para tudo isso, o modelo de manutenção ideal é capaz de compreender os sistemas observados a cada momento, e buscando os custos e a performance ideal. Esse modelo sugere que as tarefas de manutenção devem ser abalizadas por dados que indiquem o grau de degradação de componentes antes de sua falha. E o objetivo é evitar tarefas desnecessárias enquanto recomenda ações apenas quando sistemas

ou componentes realmente precisam ser intervencionados (YOUSEFI; TSIANIKAS; COIT, 2020).

1.3 A APRENDIZAGEM POR REFORÇO

A aprendizagem por reforço, AR, surgiu como um método de treinamento de redes neurais artificiais que propunha trocar o tradicional exemplo da aprendizagem supervisionada por um escalar de reforço, de onde se origina o nome. Essa substituição observa o chamado condicionamento pavloviano, fenômeno natural descrito por foi descrito por Ivan Pavlov, um fisiologista russo que recebeu o Prêmio Nobel de Medicina em 1904. Ele observou que, quando um estímulo, como o sino, era apresentado antes de um reforço, como a comida, o animal aprendia a associar o sino com a comida e começava a reagir ao sino sozinho. Isso é conhecido como condicionamento clássico ou condicionamento pavloviano. Ele é amplamente utilizado em psicologia e em outras áreas da ciência para entender como os organismos aprendem a partir de suas experiências. Na RL, o condicionamento pavloviano é utilizado para criar associações entre as ações do agente e as recompensas, possibilitando que o agente aprenda a tomar decisões que maximizam a recompensa ao longo do tempo. Isso é feito através da apresentação de reforços positivos ou negativos para incentivar ou desincentivar certas ações do agente (PAVLOV, 2010).

Com o passar do tempo, os interesses do campo da AR mudaram da neuro-robótica para a engenharia, onde excelentes resultados na solução de problemas de controle vêm sendo encontrados (KOPRINKOVA-HRISTOVA, 2014), pois em termos computacionais, o desejo da AR é maximizar a recompensa em busca de uma política ótima que definirá um conjunto de ações a serem tomadas quando determinado estado baseado nas observações for atingido (KNOWLES; BAGLEE; WERMTER, 2011). E foi exatamente por esse caminho que a AR foi introduzida aos sistemas de refrigeração e climatização, como explorado por (BARRETT; LINDER, 2015; BEGHI, Alessandro; RAMPAZZO, Mirco; ZORZI, 2017; DING; SUBIANTORO; NORRIS, 2021; WEI, T.; WANG, Yanzhi; ZHU, Q., 2017; ZHANG, D.; GAO, Z., 2019), e até previsão, como encontrado por (JANG *et al.*, 2021; LIU, Tao *et al.*, 2019a, 2019b).

De forma similar, os algoritmos de AR podem ser usados para determinar a frequência ótima de tarefas de manutenção e reparo com base no desempenho passado e no histórico de falhas do sistema. A AR também pode ser usada para

otimizar a alocação de recursos para manutenção e reparo, como determinar o número ótimo de técnicos e peças sobressalentes para manter em mãos, a exemplo das propostas de (HU, J. *et al.*, 2022a; KNOWLES; BAGLEE; WERMTER, 2011; KONGKIJPIPAT *et al.*, 2022; VALET *et al.*, 2022a; ZHANG, P.; ZHU, X.; XIE, M., 2021). Destarte, há a possibilidade da aplicação da AR também ao diagnóstico e detecção de faltas em sistemas de refrigeração visando otimizar cronogramas de manutenção e reparo do sistema, a exemplo da crescente literatura trazida como exemplo.

1.4 OBJETIVOS

O estudo apresentado por este documento pretende criar um agente de aprendizagem por reforço para gerir o programa de manutenção baseado em condição de um refrigerador.

O refrigerador, do tipo baseado em compressão mecânica de vapor, foi inspirado em um dispositivo comercial e calibrado para replicar um freezer existente e devidamente instrumentalizado, porém modelado computacionalmente para atender a dois objetivos principais: 1. Vida útil virtual de dez anos, 2. Compatibilidade com a interatividade requerida pela AR.

O agente controla o programa de manutenção decidindo sobre o funcionamento do freezer diante dos dados operacionais que recebe, e com isso, define quando será convocada a manutenção ou a substituição de algum equipamento defeituoso. E o refrigerador se mantém em seu estado operacional até que a ação decidida pelo agente seja cumprida ou outra ação seja requerida.

O sistema foi simulado em Python™, linguagem de programação bastante difundida na academia e com fortes aplicações nas áreas de inteligência artificial. O ambiente, ou *test bench,* foi construído usando a *application programing interface* (API) do Gym™, uma biblioteca Python™ desenvolvida pela OpenAI™, de código aberto, usada como protocolo de unificação para acoplagem e comparação de algoritmos de aprendizado a AR (GYM TEAM, 2022), e o agente foi criado através da biblioteca TensorForce™, um *toolkit* de aprendizagem por reforço profundo de código aberto e ênfase em pesquisa acadêmica (TENSORFORCE TEAM, 2022).

1.4.1 Objetivo geral

O principal objetivo é desenvolver um *test bench* para treinar e avaliar agentes de aprendizagem por reforço orientados a manutenção em refrigeradores baseados em compressão mecânica.

1.4.2 Objetivos específicos

1. Estabelecer a conexão entre a aprendizagem por reforço e o programa de manutenção baseado em condição aplicado a refrigeradores baseados em compressão mecânica de vapor;
2. Construir um *test bench* capaz de simular a operação do freezer e sua degradação ao longo da vida útil;
3. Construir um *test bench* capaz simular programas de manutenção corretiva, preventiva e preditiva;
4. Avaliar a performance do agente na criação do programa preditivo e comparar com os programas preventivo e corretivo;

1.5 JUSTIFICATIVAS

As contribuições da dissertação se baseiam em duas prerrogativas:

1. Ainda que as técnicas de AR sejam promissoras em termo de performance se comparadas as abordagens periódicas e corretivas de manutenção (CORREA-JULLIAN; LÓPEZ DROGUETT; CARDEMIL, 2020), o número de pesquisas aplicando AR a otimização da manutenção é inferior ao de técnicas baseadas nos paradigmas supervisionado e não supervisionado (YOUSEFI; TSIANIKAS; COIT, 2020). A primeira contribuição reside em ampliar o arsenal de análises e propostas contendo o método da AR no contexto da manutenção.
2. Para a refrigeração, a AR apenas foi aplicada como ferramenta de controle e previsão, como por exemplo no estudo de Ding *et al* (2021). E Behfar *et al* (2017) também argumenta que mais métodos de DDF, em geral, também foram desenvolvidos para aparelhos de ar-condicionado e chillers do que para sistemas de refrigeração comercial. Portanto, que embora os ciclos básicos de refrigeração sejam semelhantes, os aparelhos dos supermercados e os sistemas de ar-

condicionado variam em muitos aspectos, o que justificaria estudos específicos. Logo, a segunda contribuição está em adicionar um novo paradigma ao arsenal de estratégias de manutenção aplicadas a DDF em freezers, enquanto reduz o *gap* numérico apresentado por Behfar *et al* (2017).

Logo, baseado em trabalhos teóricos como o de Yousefi *et al* (2020), e frameworks de DDF aplicados a problemas reais, como os de Barde *et al* (2019); o de Koprinkova-Hristova (2014); Mahmoodzadeh *et al* (2020); Rocchetta *et al* (2019), esse trabalho mostrará um framework funcional que pode servir como ponto de partida para investigações mais profundas.

1.6 A ESTRUTURA DA DISSERTAÇÃO

Esse documento está estruturado como segue:

- No capítulo 2, uma revisão de estudos relacionados foi realizada com o objetivo de analisar as oportunidades do presente estudo.
- No capítulo 3, as bases teóricas são discutidas. Os fundamentos matemáticos relacionados ao agente e ao ambiente são levantados.
- No capítulo 4, características do ambiente utilizado nos experimentos são detalhadas, e as etapas de desenvolvimento são explicadas e justificadas.
- No capítulo 5, a metodologia de exploração dos resultados é apresentada. E todas as simulações que serão conduzidas durante a avaliação do agente são explicadas e justificadas.
- No capítulo 6, os resultados das simulações conduzidas são apresentados e discutidos.
- No capítulo 7, as conclusões que podem ser inferidas a partir dos resultados são apresentadas.

2 REVISÃO DA LITERATURA

2.1 REFRIGERAÇÃO E MANUTENÇÃO, AS POLÍTICAS E ABORDAGENS COMUNS

Os programas de manutenção para sistemas de refrigeração são, em linhas gerais, orientados a dados estatísticos de falhas para a realização de inspeções periódicas, e quando há a ocorrência de uma falha fora do esperado, o equipamento é reparado após a quebra. Então, se observarmos as estratégias de manutenção do ponto de vista do momento em que elas ocorrem, podemos listar três como principais: A corretiva, a preventiva e a preditiva (MOBLEY, 2002).

A manutenção corretiva parte da abordagem de que o sistema será intervencionado apenas quando houver a falha, quando o problema estiver instalado e for detectado pela impossibilidade de realizar suas funções normais. A estratégia corretiva, por esse motivo, costuma apresentar custos maiores e demandar mais tempo de reparo. Na refrigeração, a manutenção corretiva põe em risco a carga acondicionada, e em grandes empresas, põe estresse sobre trabalhadores e equipes que precisam gerir estoques e recuperar os aparelhos danificados o mais rápido possível. Estima-se que esse programa pode custar até três vezes mais que a manutenção preventiva (MOBLEY, 2002),(SCHWINDEN LEAL, 2019).

O programa preventivo é uma evolução que partindo de estatísticas de falhas propõe reparos regulares, com ou sem falha. Geralmente, assume-se que a degradação ocorre em intervalos chamados de MTTF, ou *mean time to failure*, calculados através da taxa de falhas de uma população de equipamentos similares. O MTTF pode ser considerado tanto para um equipamento quanto para subsistemas dele, ficando a cargo da equipe de manutenção gerir o programa. Todavia, ainda que a degradação seja retardada e o nível de desempenho recuperado a menores custos, não é possível evitar falhas completas, pois não há acompanhamento em tempo real. Os intervalos de inspeção baseados em estatísticas geram incertezas e negam a existência eventual de anormalidades que podem causar transtornos e no fim, a ocorrência de ações corretivas fora do programa (MOBLEY, 2002),(SCHWINDEN LEAL, 2019).

A solução preditiva parte da ideia de monitoramento constante do sistema, para com isso estimar o estado de degradação e evitar as dificuldades do programa corretivo e preventivo. A ideia é antecipar as falhas para aumentar o tempo entre as intervenções, reduzir os custos e aumentar a sobrevida (SCHWINDEN LEAL, 2019).O modelo preditivo tem avançado nos últimos anos devido ao desenvolvimento de sensores mais baratos, melhores interfaces de conexão entre máquinas e programas mais confiáveis para interpretação e gestão dos dados (KUŽNAR *et al.*, 2017).

Os métodos baseados em condição podem ser entendidos como métodos usados para reduzir o custo e minimizar as incertezas das atividades de manutenção já que serão apenas tomadas quando o componente demandar atenção (PENG, Y.; DONG, M.; ZUO, M. J., 2010). Então, de forma geral, pode-se entender como um método de tomada de decisão baseado na observação das condições de um sistema ou seus componentes (AHMAD; KAMARUDDIN, 2012).

A solução preditiva ainda é um campo de investigação para diversas equipes de desenvolvimento em indústrias e universidades. Hoje, considera-se que a maioria das pesquisas existentes sobre CBM, ou *condition based maintenance* assume que as manutenções preventivas devem ser realizadas quando as degradações dos componentes do sistema atingem níveis limite específicos na inspeção. No entanto, ainda é um desafio conciliar a CBM com a gestão de grandes sistemas multicomponentes, devido a existência de dependências complexas, não linearidades e relações indiretas que geram situações desafiadoras e até conflitantes para metodologias mais simples. Por isso, o CBM torna-se desafiador à medida que aumenta a complexidade de um sistema (ZHANG, N.; SI, 2020a).

A manutenção preditiva, baseada em condição, possui níveis de assertividade (MOBLEY, 2002):

1. A detecção de faltas, ou a capacidade de encontrar anormalidades;
2. O diagnóstico de falhas, ou a capacidade de determinar a causa da anormalidade;
3. O prognóstico da falha, ou a capacidade de inferir quando a falha ocorrerá.

Para a predição, todavia, não é necessário que todos os níveis estejam presentes, sendo o diagnóstico e o prognóstico aprimoramentos que partem da

capacidade de detectar anomalias. Na detecção e diagnóstico automáticos voltados a sistemas mecânicos de refrigeração, se o diagnóstico for fornecido, ele geralmente vem após ou concomitantemente à detecção, e a avaliação geralmente não é incluída (BEHFAR; YUILL; YU, 2017; KATIPAMULA; BRAMBLEY, 2005; WICHMAN; BRAUN, 2009).

Na arquitetura preditiva, a determinação da presença de uma falha em um determinado sistema é chamada de detecção. O diagnóstico de falhas envolve a determinação do tipo e/ou localização. Em algumas aplicações, o diagnóstico inclui o comportamento de uma falha que varia no tempo, mas isso não é comum nos sistemas de detecção e diagnóstico. A avaliação ou prognóstico da falha, por sua vez, é uma avaliação da gravidade da falha (BEHFAR; YUILL; YU, 2017; KATIPAMULA; BRAMBLEY, 2005).

Há três abordagens principais para a detecção e diagnóstico de falhas, que são a baseada em modelo, conhecimento e dados (ZHOU, Z.; LI, G.; *et al.*, 2020b). Os métodos baseados em modelos têm como vantagem a menor necessidade de dados e a simplicidade, todavia são os mais imprecisos. Os métodos baseados em conhecimento apoiam-se no domínio das partes interessadas e são inflexíveis quanto a aplicação em sistemas diferentes sem as devidas modificações, o que dificulta sua universalização. Já os métodos *data-driven*, vem tomando espaço por que dependem menos dos conhecimentos das partes envolvidas enquanto tem alta acurácia (ZHOU; WANG; et al., 2020). Dito isso, podemos explorar como os métodos *data-driven* tem adentrado o universo da manutenção na refrigeração.

2.2 REFRIGERAÇÃO, MANUTENÇÃO E APRENDIZAGEM DE MÁQUINA, UM PANORAMA

Desde a década de 80, significativos avanços têm sido feitos no campo da detecção e diagnóstico de faltas DDF em sistemas de refrigeração. Mas dentre as muitas técnicas, as chamadas *data-driven* tem conquistado espaço devido sua adequação as aplicações dos sistemas de refrigeração modernos (DEY; RANA; DUDLEY, 2018). Os métodos *data-driven* basicamente exploram a relação entre dados de entrada e as faltas em sistemas, e a diferença entre seus métodos reside no tipo de algoritmo matemático que dirige o sistema de interpretação de dados. Todavia, ao contrário dos métodos baseados em modelo e conhecimento, já

sedimentados, os métodos *data-driven* ainda estão em amadurecimento (ZHOU, Z.; WANG, Jiangyu; *et al.*, 2020).

Sistemas *data-driven* não necessitam de modelos explícitos para encontrar relações entre diferentes padrões de dados, e por consequência, encontrar componentes com faltas em sistemas (DEY; RANA; DUDLEY, 2018). Eles repousam sobre séries temporais obtidas por sensores sem nenhum conhecimento anterior sobre a física dos sistemas (LEI, Y. *et al.*, 2020). Com o crescimento da disponibilidade de estudos sobre otimização da manutenção, ficou evidente que a abordagem baseada em modelos esbarra em limitações derivadas dos parâmetros de hipóteses de calibragem quando aplicados a sistemas grandes e complexos (YOUSEFI; TSIANIKAS; COIT, 2020). As aplicações *data-driven* como o aprendizado de máquina, costumam superar essas dificuldades com boa performance, o que explica em parte o interesse na aplicação da aprendizagem de máquina a tarefas de tomada de decisão, inclusive para a manutenção (CORREA-JULLIAN; LÓPEZ DROGUETT; CARDEMIL, 2020).

O aprendizado de máquina é a área da inteligência artificial concernente ao desenvolvimento de máquinas que são capazes de aprender. Embora essa não seja uma definição perfeita, ela captura o ponto principal da aprendizagem, que está relacionado ao comportamento e não a composição do algoritmo. O aprendizado de máquina suplanta a programação tradicional indo além do acúmulo de informações, ele está relacionado a capacidade de realizar escolhas e criar regras novas automaticamente, e esse é o grande recurso desenvolvido. Como ferramenta capaz de interpretar determinados dados e encontrar padrões e desenvolver regras, o aprendizado de máquina torna-se particularmente interessante como assistentes de diagnóstico de dados ou programação de rotinas de manutenção por exemplo. O objetivo do aprendizado de máquina, nesse campo, é multiplicar os recursos humanos mimicando os comportamentos de um especialista.

De forma geral, existem três formas de aprendizado, a supervisionada, a não supervisionada e a AR. O aprendizado supervisionado funciona relacionando pares de informação de entrada e saída, e buscando quais as respostas o agente deve fornecer para determinados dados de entrada. Aqui é necessária a ação de um supervisor, que deverá identificar erros e realizar ajustes nos parâmetros da rede até que esses sejam eliminados por meio de treinamentos. Quando o sistema adquire

esse grau de amadurecimento, considera-se que ele aprendeu. O aprendizado não supervisionado não requer que sejam comparadas as entradas e as saídas, e não necessita de um supervisor. O agente então trabalha organizado os dados de entrada em categorias a seus critérios, gerando classes de dados e encontrando respostas a essas entradas. A aprendizagem por reforço funciona com base na tentativa e erro, com que um agente interagindo com uma base de dados ou meio, deverá buscar interpretá-lo para um fim baseando suas conclusões em um índice de desempenho.

Os métodos de aprendizado de máquina do tipo *data-driven* nos paradigmas por supervisionado e não supervisionado hoje são quase a completude das pesquisas sobre DDF em sistemas de refrigeração e ar-condicionado. Eles também são baseados no comportamento e geralmente enxergam o objeto estudado pela abordagem *black-box*, ou seja, apenas observando os parâmetros de entrada e saída, sem interagir com parâmetros intermediários e sem conhecer a física por trás deles (KIM, W.; KATIPAMULA, 2018b). Por isso, estão classificados como *process history-based*. Dentro dessa classificação, Kim & Katipamula (2018) argumenta que 63% dos trabalhos de abordagem *black-box* concentram-se em técnicas estatísticas, 25% em redes neurais e apenas 12% em outras técnicas baseadas em reconhecimento de padrões, uma área do aprendizado de máquina. Entre os trabalhos, Kim & Katipamula (2018) encontrou técnicas como *Support vector machine* e *Bayesian networks*.

Panorama semelhante foi encontrado por Lei *et al* (2020), Yan et al. (2020) e Sun *et al* (2019), que encontraram *Bayesian-Network, support-vector-machine e Generative Adversial Networks* aplicados a chillers, todos supervisionados e não supervisionados. Como se vê também em Dey, Rana & Dudley (2018), que como Zhou *et al* (2020), analisando os algoritmos do tipo *data-driven* aplicados a DDF em sistemas de refrigeração do tipo VRF, encontraram que há pelo ou menos cinco tecnologias: *decision-tree, support vector regression, clustering, shallow neural networks,* e *deep neural networks*. Todos com o suporte de sensores e realizando a DDF a partir de características operacionais dos componentes da máquina.

Hong et al (2020), em uma pesquisa estado-da-arte, obteve concernente a determinação de políticas de manutenção em equipamentos de edifícios, como sistemas de refrigeração comerciais, oito algoritmos de aprendizado de máquina, sendo aplicados a chillers PCA, SVM, *linear discriminant analysis* (LDA); *linear classifier, bayesian network classifier* (BNC), unidades de AHU o PCA, ANN, unidades

de VAV o PCA, BNC, unidades de VRF *decision trees*, unidades terminais o SVM e a sistemas completos as ANN.

Juntando todos os dados levantados por (KIM, W.; KATIPAMULA, 2018a), (YAN et al., 2020), (YAN; HUA, 2019), (ZHOU et al., 2020) e (HONG et al., 2020), encontramos o Quadro 1.

Quadro 1. Resumo dos métodos de AM aplicados a DDF.

Equipamentos	Algoritmos de AM	Referências	Ano
Refrigeradores	SVM	(REN *et al.*, 2008)	2008
Chillers	BN	(SHI, Z., 2018)	2018
	SVM	(YAN et al., 2014), (YAN, K. *et al.*, 2018), (YAN, K.; JI; SHEN, W., 2017), (LI, D.; HU, G.; SPANOS, C. J., 2016),	2014, 2018, 2017, 2016
	GAN	(DEY; RANA; DUDLEY, 2018), (ZHONG, C. *et al.*, 2019)	2018, 2019
	PCA	(COTRUFO; ZMEUREANU, 2016), (BEGHI, A. *et al.*, 2016), (LI, G. *et al.*, 2016)	2016
	LDA	(LI, D. *et al.*, 2016)	2016
	Linear classifier	(TRAN *et al.*, 2015), (ZHAO, X., 2015)	2015
	BNC	(BONVINI *et al.*, 2014), (HE, S. *et al.*, 2016)	2014, 2016
VRF	*Decision tree*	(LI, G. *et al.*, 2018), (LIU, Jiahui *et al.*, 2019a), (WANG, Jiangyu *et al.*, 2017), (LI, G. *et al.*, 2017)	2017, 2018, 2019
	Support vector regression	(LI, G.; HU, Y., 2018), (SUN, K. *et al.*, 2016), (HAN, H. *et al.*, 2011), (LIU, Jiangyan *et al.*, 2018), (LIU, Jiahui *et al.*, 2019b)	2011, 2016, 2018, 2019

	Clustering	(ZOGG; SHAFAI; GEERING, 2006), (CAPOZZOLI; LAURO; KHAN, 2015)	2006, 2015
	Shallow neural networks	(SUN, S. *et al.*, 2017), (SHI, S. *et al.*, 2018), (ZHOU, Z.; WANG, Jiangyu; *et al.*, 2020)	2017, 2018, 2020
	Deep neural networks	(FAN, C.; XIAO; ZHAO, Yang, 2017)	2017
AHU	BN	(NAJAFI *et al.*, 2012)	2012
	PCA, ANN	(LI, S.; WEN, 2014)	2014
VAV	PCA	(WU, S.; SUN, J. Q., 2011)	2011
	BNC	(XIAO *et al.*, 2014)	2014
Terminais	SVM	(DEY; RANA; DUDLEY, 2020)	2020
Todo o equipamento	ANN	(DU, Z. *et al.*, 2014)	2014

Fonte: O autor.

As revisões de Lei *et al* (2020) e Hong et al (2020) evidenciam particularmente que a academia tem privilegiado a climatização em detrimento a refrigeração, e essa é uma das justificativas utilizadas por Sun et al. (2021) para o desenvolvimento de sua pesquisa, por exemplo. Isso é perceptível pelo menor número de produções que tem essas tecnologias como alvo. Mas usando o controle de vocabulário da plataforma *engineering village,* é possível selecionar as seguintes contribuições realizadas entre 2020 e 2022 como relevantes no contexto desse trabalho, Quadro 2.

Quadro 2. Pesquisa bibliográfica entre 2020 e 2022 sobre AM com vocabulário controlado na plataforma *engineering village.*

Plataforma de pesquisa	Vocabulário controlado	Equipamento	Algoritmos de AM	Referências	Ano
Engineering village	*Maintenance; Refrigeration; Predictive*	Refrigeradores	BP Neural Network, RBF Neural Network	(CHANG *et al.*, 2020b), (BANSAL *et al.*, 2021)	2020, 2021
	Fault detection, Refrigeration,		CNN, SVM, LDA, LDA-SVM e PCA-SVM;	(SOLTANI *et al.*, 2022), (LEE, D.; CHEN, M.-	2022

	AM			H.; LAI, 2022)	
	Supermarket refrigeration, *Fault detection*		Pré-processamento	(SUN, J. *et al.*, 2021)	2021
	Fault diagnosis, *Refrigeration*		KNN, SVM, DT, RF e LR	(ZHANG, Z. *et al.*, 2020)	2020

Fonte: O autor.

Sobre os trabalhos destacados no Quadro 2, destaco que a pesquisa realizada por Sun et al. (2021), se propõe a contribuir para a redução do *gap* entre aplicações de climatização e refrigeração induzindo uma matriz de DDF para futuras explorações. No final, portas abertas, acúmulo de gelo, falha da válvula de expansão e falha do ventilador do evaporador foram encontradas como as falhas mais relevantes para o desenvolvimento de algoritmos de AM aplicados a DDF.

Além da identificação dos equipamentos relevantes, Zhang et al. (2020), propôs que uma das saídas para acelerar a produção de conhecimento acerca da identificação de faltas e universalização de modelos se dá pela aplicação de algoritmos de voto majoritário com KNN, SVM, DT, RF e LR em diferentes subconjuntos, preenchidos com temperatura, pressão, vazão, posição de válvula, potência etc. Em comparação com modelos individuais, o voto majoritário integrado alcançou uma precisão maior, até 99,58% na detecção de faltas de refrigerante, um tópico relevante no contexto das falhas em sistemas de refrigeração.

E de forma semelhante, Soltani et al. (2022), propõe que os algoritmos focados em identificação de falta devem ser julgados pela sua precisão, ou seja, tempo de computação e taxa de falso positivo. Por isso, este estudo visa investigar isso explorando vários classificadores de aprendizado superficial (CNN, SVM, LDA, LDA-SVM e PCA-SVM) e a seleção de recursos para, no final, propor qual se adequa melhor para uso industrial. No final, CNN e PCA-SVM não apresentam desempenho satisfatório, a precisão do SVM oscila em torno de 95%, enquanto LDA pode reduzir as taxas de falso positivo até zero.

Portanto, observando esse excerto de três anos em literatura exclusiva, em acréscimo as revisões realizadas por Kim, W. e Katipamula (2018a), Yan *et al.,* (2020),

Yan, Hua (2019), Zhou *et al.,* (2020) e Hong *et al.,* (2020),é possível identificar os seguintes pontos de convergência na literatura:

1. A generalização de modelos se tornou valorizada quanto a detecção de faltas, o que explica o avanço frente a utilização de *ensembles* e múltiplas técnicas em um único paper;
2. A detecção de faltas em nível de diagnóstico, é valorizada mais que a simples detecção;
3. A contenção das técnicas de aprendizado de máquina no conjunto de treinamento é altamente valorizada e o grande objetivo de modelos eficientes é superar essa limitação, isso é, a limitação de trabalhar apenas com dados já conhecidos;
4. Existem mais trabalhos explorando sistemas de HVAC, especificamente VRF, do que refrigeradores comerciais, mesmo que os últimos sejam relevantes no para os objetivos de redução das metas de emissões;
5. As despesas de rastreamento e censura são levadas em consideração e relevantes, e há trabalhos que se propõem apenas a aprimorar seleção de *features* de rastreamento.

Ciente disso, podemos explorar como a AR pode, como abordagem, contribuir para a dissolução desses pontos que a academia tem explorado nos últimos anos.

2.3 MANUTENÇÃO E APRENDIZAGEM POR REFORÇO, UMA PROPOSTA

Para o DDF, a AR é uma área da inteligência artificial concernente a sistemas *data-driven* e a solução de problemas de tomada de decisão sob incertezas. De acordo com Mahmoodzadeh *et al* (2020), as principais características que distinguem a AR no âmbito do DDF de outros algoritmos podem ser visualizadas abaixo:

- Pode prender dos dados históricos e interagir com dados online, suportando aplicações IoT;
- Pode lidar com dados sem implicações imediatas, ou seja, mesmo que as consequências de determinada ação não sejam imediatas, ainda é possível tomar decisões;
- Pode interagir e aprender de ambientes estocásticos, incorporando as incertezas do ambiente a sua tomada de decisão.

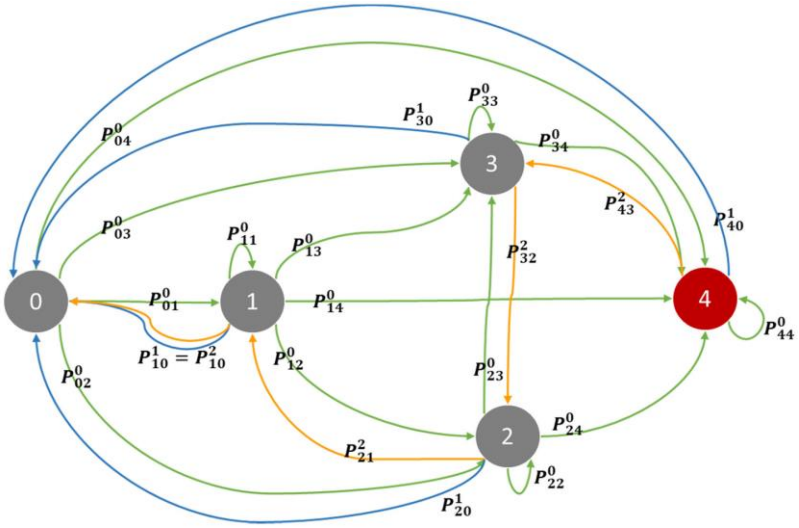
Embora os métodos de aprendizado supervisionado funcionem razoavelmente bem na identificação de falhas, eles não conseguem ter precisão em determinar o momento correto das ações de manutenção, pois é difícil quantificá-los e justificá-los durante o treinamento. Em contraste, na AR, a rotulagem explícita dos dados de treinamento não é necessária. O agente só precisa ser informado se a ação que ele executa é desejável, o que pode ser facilmente determinado com base nos dados de funcionamento. Por isso, a solução da AR de substituir a rotulação por um processo avaliativo durante o treinamento, onde o treinador só precisa fornecer uma avaliação das ações realizadas pelo agente com base no resultado pode vir a calhar em um problema de controle dinâmico domo a manutenção (LEI, Y. *et al.*, 2020).

Para a abordagem clássica da AR, tanto o ambiente quanto eventuais modelos de auxílio são considerados *black-box*, isso significa que o agente não tem suporte a interpretação de parâmetros intermediários, mas apenas os dados de entrada e saída (CORREA-JULLIAN; LÓPEZ DROGUETT; CARDEMIL, 2020), logo, para o agente, o ambiente é incerto. Destarte, o agente observa o comportamento do ambiente e é influenciado por ele já que as recompensas são derivadas de sua interpretação do meio. No âmbito da manutenção, o objetivo desta interação entre o agente e o ambiente, de onde as informações são podem ser obtidas por sensores, por exemplo, é gerar como saída a programação de manutenção baseada nos estados de degradação do ambiente, e a programação ideal, para o agente, será encontrar o modo de receber o máximo de recompensas positivas possível. O meio de obter essa recompensa pode ser definido como buscar a máxima extensão da vida útil, ou aumentar a disponibilidade, ou diminuir o consumo energético... de forma que a parte interessada pode recompensar ou punir o agente sempre que cada dessas tarefas obtiver êxito ou falha.

Um exemplo de framework de AR aplicada a manutenção pode ser encontrado em Yousefi *et al* (2020). Aqui, cada componente i degrada independentemente no tempo, e pode ser reparado individualmente. A degradação do componente é então organizada em múltiplos estágios. Quando esse componente é novo, sua degradação é zero, portanto $Dc_T = 0$. No estágio 4, entretanto, há falha. Para facilitar a compreensão, quando a degradação do componente está além de j_i^1, ele está no limiar da falha, quando ele está em j_i^2 e j_i^3, ele está no estágio 2, e quando ele está em j_i^2 e j_i^1, ele está no estágio 3. Do ponto de vista do tempo, sempre que um

componente no estado j não é reparado ele pode transitar para j' ou permanecer em j, o que é previsível, já que os componentes são fabricados para permanecerem o maior tempo possível sem serem afetados por falhas ou degradação. Todavia, também é possível retornar ao estado j estando em j'. A Figura 1 transforma essa relação em um fluxograma.

Figura 1. Framework do MDP para um componente.



Fonte: Yousefi *et al* (2020).

Dessa forma, para testar a adequação da AR ao problema de programação de manutenção, é necessário definir cenários indicativos que servem de base para experimentos simulados. Essas simulações envolverão um modelo de ambiente e um modelo de aprendizagem por reforço. O modelo de ambiente fornece ao módula AR uma indicação de uma condição atual, o agente de AR então decide se deve executar uma determinada tarefa de manutenção. Isso é semelhante ao cenário de controle ótimo descrito por (SUTTON; BARTO, 2018). Se a manutenção não for realizada, uma falha pode ou não ocorrer. Se o ambiente não falhar, um lucro é devolvido como recompensa. Se o sistema falhar, um custo de reparo é deduzido do lucro. Se o agente de AR decidir realizar a manutenção, o sistema não falhará, mas um custo de manutenção será deduzido do lucro. O custo de manutenção é consideravelmente menor do que o custo de falha, como é típico em cenários do mundo real. Assim, a cada passo de tempo, o módula AR deve decidir entre uma recompensa conhecida e moderada, realizando a manutenção ou arriscando nenhuma manutenção, o que pode resultar em uma recompensa alta no caso de nenhuma falha ou uma recompensa

baixa se o componente falhar (KNOWLES; BAGLEE; WERMTER, 2011). Já do ponto de vista prático, a implementação de frameworks de AR à manutenção costuma seguir o seguinte *pipeline* (AMARI; MCLAUGHLIN; PHAM, 2006; GINER *et al.*, 2021; KNOWLES; BAGLEE; WERMTER, 2011; LAMPRECHT; WURST; HUBER, 2021; MAHMOODZADEH *et al.*, 2020):

- Coletar dados sobre o desempenho e histórico de falhas do sistema: Neste passo, é coletado dados sobre a operação e o desempenho do sistema, bem como quaisquer falhas que tenham ocorrido. Esses dados são usados para treinar e testar o modelo RL;
- Definir o problema RL: Neste passo, o problema AR é definido em termos das ações que o modelo AR pode tomar, os estados do sistema, as recompensas ou penalidades associadas a diferentes ações e estados e quaisquer restrições ou limitações nas ações que podem ser tomadas;
- Treinar o modelo RL: Neste passo, o modelo AR é treinado usando os dados coletados no passo 1. Isso geralmente envolve o uso de um algoritmo RL, como Q-learning ou SARSA, para aprender uma política que maximize a recompensa ou minimize a penalidade;
- Testar e validar o modelo RL: Neste passo, o modelo AR é testado e validado usando um conjunto separado de dados para garantir que ele seja capaz de detectar e diagnosticar falhas no sistema de maneira precisa;
- Implementar o modelo RL: Depois que o modelo AR foi treinado e validado, ele pode ser implementado no sistema para detectar e diagnosticar falhas em tempo real. O modelo AR pode continuar aprendendo com os dados coletados durante sua operação e pode adaptar seu comportamento conforme necessário para otimizar o desempenho do sistema.

Compreendido isso, podemos observar para onde o desenvolvimento da literatura tem apontado nos últimos quatro anos usando o controle de vocabulário da plataforma *engineering village.* O primeiro objeto de observação deve ser o crescimento de 633% no número de publicações específicas na área entre 2019 e 2022, conforme Quadro 3.

Quadro 3. Pesquisa bibliográfica entre 2020 e 2022 sobre AR com vocabulário controlado na plataforma engineering village.

Plataforma de pesquisa	Vocabulário controlado	Referências	Ano	Quantidade
Engineering village	*(((((((maintenance & reinforcement learning) WN ALL)) AND (({reinforcement learning} OR {maintenance}) WN CV)) AND (({markov processes} OR {scheduling} OR {condition-based maintenance} OR {internet of things} OR {deterioration} OR {optimization} OR {fault detection}) WN CV)) AND ((2022 OR 2021 OR 2020 OR 2019 OR 2018) WN YR)) AND ({ja} WN DT))*	(LI; ZHONG; LIN, 2019), (BARDE; YACOUT; SHIN, 2019), (ANDRIOTIS; PAPAKONSTANTINOU, K. G., 2019)	2019	3
		(YOUSEFI; TSIANIKAS; COIT, 2020), (LIU, Yu; CHEN, Yiming; JIANG, T., 2020),(MAHMOODZADEH *et al.*, 2020), (ZHANG, N.; SI, 2020b), (ADSULE; KULKARNI; TEWARI, 2020), (WEI, S.; BAO; LI, Hui, 2020), (PARASCHOS; KOULINAS; KOULOURIOTIS, 2020)	2020	7
		(PENG, S.; FENG, Q. (May), 2021; PINCIROLI *et al.*, 2021; WANG, H.; YAN, Q.; ZHANG, S., 2021), (YANG, H.; LI, Wenchao; WANG, B., 2021), (CUI, P. *et al.*, 2021), (ANDRIOTIS; PAPAKONSTANTINOU, K. G., 2021), (YANG, Y.; YAO, 2021), (RUAN *et al.*, 2021), (LUO, Y., 2021), (MENG; BAI; JIN, J., 2021), (RENARD; CORBETT; SWEI, 2021), (CHENG, M.; FRANGOPOL, 2021), (CHEN, Jing; CHEN, Jia; ZHANG, Hongke, 2021; DAI, S. *et al.*, 2021; TANIMOTO, 2021; WANG, Xiao *et al.*, 2021; ZHANG, Y. *et al.*, 2021)	2021	16
		(CHENG, J. *et al.*, 2022; DU, A.; GHAVIDEL, 2022; FENG, M.; LI, Y., 2022; GAO, P. *et al.*, 2022; HU, J. *et al.*, 2022b; MOHAMMADI; HE, Q., 2022; NGUYEN *et al.*,	2022	19

		2022; ONG; WANG, W.; HIEU; *et al.*, 2022; ONG; WANG, W.; NIYATO; *et al.*, 2022; RUIZ RODRÍGUEZ *et al.*, 2022; VALET *et al.*, 2022b; WANG, Jing; LEI, D.; CAI, 2022; YANG, A. *et al.*, 2022; YANG, D. Y., 2022; YOUSEFI; TSIANIKAS; COIT, 2022; ZHANG, Huidong; DJURDJANOVIC, 2022; ZHAO, Yunfei; SMIDTS, 2022; ZHOU, W. *et al.*, 2022; ZHOU, Yifan; LI, B.; LIN, 2022)		

Fonte: O autor.

Removendo as abordagens puramente teóricas, é válido destacar, em ordem cronológica e de relevância de acordo com a plataforma *engineering village*, que o trabalho de Li, Zhong & Lin (2019) usa AR para otimizar a política de manutenção de motores aeronáuticos durante sua vida útil, com o objetivo de minimizar os custos de manutenção e aumentar a disponibilidade, e que obteve sucesso em demonstrar a efetividade dessa abordagem em estudos simulados provando que é capaz de superar abordagens tradicionais de manutenção principalmente em termos de custos e disponibilidade. De maneira similar, Barde *et al.,* (2019) propôs uma estratégia baseada em AR para otimizar a manutenção de uma frota de veículos militares, e mais uma vez obteve sucesso em demonstrar melhoras significativas em termos de custos de manutenção e disponibilidade.

Um ano após, Mahmoodzadeh et al., (2020) propôs uma estratégia de CBM para gasodutos secos usando AR. Os autores abordaram a otimização dos intervalos de manutenção baseando-se na condição para reduzir os custos de manutenção e aumentar a disponibilidade da estrutura. Ao final, foi demonstrada a efetividade do AR como programador, sendo capaz de atender as expectativas esperadas. Resultados semelhantes foram encontrados por Wei *et al.,* (2020), que propôs uma abordagem DRL para otimização da manutenção em infraestruturas de pontes e túneis. Os autores desenvolveram um modelo para aprender a política de manutenção baseada na minimização do risco estrutural e maximização do tempo de vida da estrutura. A

efetividade da abordagem proposta foi demonstrada em um estudo de caso envolvendo uma ponte.

Já Pinciroli *et al.,* (2021), propôs um modelo DRL baseado em PPO para a manutenção de uma fazenda eólica com múltiplas equipes de colaboradores. Os autores demonstraram a efetividade da abordagem através de uma simulação, mostrando significativos avanços em termos de eficiência da manutenção e redução de custos se comparado a estratégias tradicionais de reparo. O modelo DRL foi capaz de se adaptar a mudanças no contexto da fazenda eólica e otimizar o esquema de manutenção a disponibilidade da equipe de reparo. E Yang *et al.,* (2021), propôs uma abordagem AR para a otimização da manutenção preventiva de juntas em um sistema multiestado. A efetividade dessa abordagem foi provada através de um estudo simulado que comparou o agente aos programas tradicionais de manutenção do ponto de vista do custo e disponibilidade. O modelo de AR usou dados históricos para treinamento e foi capaz de se adaptar quando necessário aos modelos de produção para criar a programação da manutenção.

Em 2022, Mohammadi & He, (2022), desenvolveram uma abordagem DRL baseada em DDQN para otimizar o planejamento de manutenção visando reduzir o risco e aumentar a efetividade das intervenções em estradas de ferro. Para provar a abordagem, uma simulação de um trecho do trilho foi realizada e a posterior análise dos resultados demonstrou que a abordagem proposta não apenas reduziu os custos como também aumentou a segurança do trecho sob intervenção. Similarmente, Gao *et al.,* (2022) propôs aplicar o Q-learning ao planejamento da manutenção de um sistema de geração distribuída do tipo fotovoltaico. Do mesmo modo, o sistema foi modelado, e por comparação às abordagens já estabelecidas, os resultados apontaram melhoria na gestão dos recursos energéticos, disponibilidade do sistema e redução de custos.

Dito isso, pode-se entender o seguinte:

- Muitos dos trabalhos experimentais se concentram em demonstrar a eficácia das abordagens baseadas em AR para melhorar a eficiência da manutenção e reduzir o tempo de inatividade em vários tipos de sistemas, como sistemas industriais, turbinas eólicas, fábricas e tubulações;

- Alguns dos artigos avaliam o desempenho de abordagens baseadas em AR por meio de estudos de simulação, enquanto outros conduzem experimentos em sistemas do mundo real;
- Muitos dos artigos usam técnicas de aprendizado profundo, como redes neurais profundas ou processos gaussianos, para melhorar os recursos de aprendizado e generalização das abordagens baseadas em RL;
- Alguns dos artigos consideram a integração de informações ou restrições adicionais, como dados de custo ou disponibilidade de recursos de manutenção, na abordagem baseada em RL.

Logo, com base nesses trabalhos, pode-se sugerir que a potencial aplicação da AR na detecção e diagnostico de falhas em sistemas de refrigeração está em determinar a frequência ótima de reparo com base na performance e no histórico de falhas do sistema. AR também poderia ser usada para otimizar a alocação de recursos de manutenção, como determinar o número ideal de técnicos e peças de manutenção a se manter a disposição.

3 REFERENCIAL TEÓRICO

3.1 BASES DA APRENDIZAGEM POR REFORÇO

No aprendizado de máquina padrão, os paradigmas supervisionado e não supervisionado, os sistemas costumam extrair conhecimento de pares de informação, e o retorno costuma ser a função que pode ter gerado tais pares. Esses métodos são bons quando se pode fornecer dados corretos para ancoragem, ou quando se sabe a resposta do problema. Mas quando se busca completa autonomia, é requerido que o algoritmo aprenda apenas de outras fontes, como indicadores, recompensas ou reforços fornecidos pelo próprio ambiente (RIBEIRO, 1999).

A aprendizagem por reforço é o paradigma do aprendizado que baseia no conceito de sinal de reforço ou recompensa, que é oferecida quando determinado comportamento ocorre mediante uma ação. Esse processo não é guiado por um professor, mas é determinado pelo próprio algoritmo, suas interações e os efeitos de seus resultados (SUTTON; BARTO, 2018). Destarte, o algoritmo está interessado em maximizar uma função de retorno e o faz unicamente através do conhecimento obtido dos resultados de sua interação com o ambiente (KNOWLES; BAGLEE, 2012; YOUSEFI; TSIANIKAS; COIT, 2020).

Para isso, a aprendizagem por reforço possui quatro atributos distintos (SUTTON; BARTO, 2018), que são:

1. O aprendizado pela interação: O algoritmo aprende através da interação com um ambiente, entregando uma ação e recebendo um sinal de reforço que deve informar a qualidade da ação tomada;
2. O retorno atrasado: O objetivo do algoritmo não é maximizar a próxima ação, mas o somatório no infinito. Os objetivos dele, mais que solucionar problemas imediatos, é a solução global, e por isso, as ações são tomadas para maximizar o retorno total;
3. A orientação ao objeto: O algoritmo não requer detalhes do ambiente, e não precisa conhecê-lo para extrair a política ótima de ações. Só há interação dentro dos limites programados para maximizar um determinado comportamento.

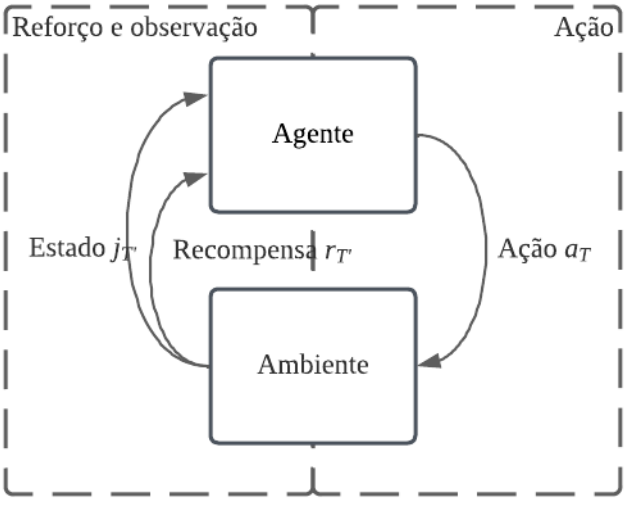
4. A investigação ou exploração: O algoritmo deve ser capaz de saber quando parar de explorar novas soluções e utilizar um conhecimento que ele já possui.

Outra grande diferença consiste em na AR, não ser necessário possuir um *dataset* construído. E isso ocorre porque nesse paradigma, o agente ou conjunto de agentes, que iniciam como *tabula rasa*, tomam uma ação a em cada estado j, são recompensados r ou penalizados p, se as ações tomadas aproximam ou afastam o estado j, do objetivo, que é a maximização de r. Assim sendo, o processo de determinação e aprendizado ocorre no tempo T, e é a forma como a política π é otimizada. Política essa que é extraída da uma função ação-valor V, que determina a validade da ação (RODRIGUES, 2018).

3.1.1 Elementos da aprendizagem por reforço

O sistema típico de aprendizagem por reforço possui dois elementos, o agente, constituído do algoritmo de aprendizagem, e o ambiente, constituído do problema a ser resolvido, Figura 2. A interação, como já apresentada, ocorre através de uma série de ações, recompensas e observações no tempo. Por isso, compreende-se que o agente deve ter a capacidade de alterar as condições do ambiente, e o ambiente de se adequar as novas informações recebidas (SEYR; MUSKULUS, 2019).

Figura 2. Relação agente-ambiente na aprendizagem por retorno padrão.



Fonte: O autor (2022).

Em detalhes, são quatro os componentes principais de um sistema AR padrão (SUTTON; BARTO, 2018).

1. O ambiente: Constituído do mapeamento de atributos observáveis, que pode ser obtido através de leituras de sensores, séries temporais e simulações computacionais e capaz de se moldar as ações dadas pelo agente. O ambiente não pode ser estático e apenas observável, como os conjuntos de dados usados no aprendizado supervisionado e não supervisionado (MNIH *et al.*, 2013).
2. A política: Normalmente expressa como π, representa a solução do algoritmo para o problema do ambiente. Ou seja, é o mapeamento a partir dos estados do ambiente, das ações que devem ser tomadas nesses estados a cada observação. A aprendizagem é a convergência de π a uma política ótima $\pi(j, a)$ (KUZNETSOVA *et al.*, 2013), ou ainda, a probabilidade de cada ação quando o agente está em um determinado estado.
3. O Retorno e o reforço: Reforço ou *recompensa* é um escalar do tipo $R_{T+1}(j, a)$, devolvido pelo ambiente e que serve de guia para a política. É de crítica importância para a formação da política, e alterá-lo pode significar alterar a função objetivo do agente. O algoritmo não sabe a quantidade máxima de Rt que pode receber, mas para todos os efeitos, sendo $Rt = \sum_{T=1}^{T=n} R$, convenciona-se que para $T = \infty$, $R = \infty$. Por isso, à função R_T, é necessário adicionar γ, ou fator de desconto $\gamma \in (0,1)$, que indica o peso das recompensas futuras frente as imediatas (SANUSI *et al.*, 2019); O Rt pode ser calculado como segue abaixo (YOUSEFI; TSIANIKAS; COIT, 2020):

$$R_T = r_{T+1} + \gamma r_{T+2} + \gamma^2 r_{T+3} + \cdots = \sum_{k=0}^{T} \gamma^k r_{T+k+1} \quad (1)$$

   Em que r_T é o reforço em T, e T o horizonte sobre o qual a política é determinada.
5. Funções de valor: O mapeamento do par ação-estado fica a cargo das funções de valor-ação e valor-estado. Estas funções só consideram o estado corrente, e são denotadas por $Q(j, a)$ e $V(j)$, respectivamente.

A função valor-ação, $Q(j,a)$, pode ser definido pela equação (2), parte do princípio de que para estabelecer π ideal, é preciso conhecer o valor de cada ação a em cada estado j. Logo, ela é a expectativa de valor ao tomar uma determinada ação em um determinado estado seguindo uma política (YOUSEFI; TSIANIKAS; COIT, 2020).

$$\begin{aligned} Q^{\pi}(j,a) &= E_{\pi}[R_T | j_T = j, a_T = a] \\ &= E_{\pi}[\sum_{k=0}^{T} \gamma^k r_{T+k+1} \,|\, j_T = j, a_T = a] \end{aligned} \tag{2}$$

Em que E_{π} é a expectativa de valor.

Já a função valor-estado, $V(j)$, equação 5, parte da equação de Bellman e das probabilidades de transição entre estados, e por consequência, probabilidade da acumulação dos reforços, equação 3 e equação 4 (YOUSEFI; TSIANIKAS; COIT, 2020).

$$P_{jj'}^{a} = P(J_{T+1} = s' | J_T = j, A_T = a) \tag{3}$$

$$R_{jj'}^{a} = E_{\pi}(R_{T+1} | J_T = j, J = j', A_T = a) \tag{4}$$

Em que $P_{jj'}^{a}$ é a probabilidade de transição, ou probabilidade de chegar em um estado j' partindo de um tempo T, tomando uma ação a, e $R_{jj'}^{a}$ é a expectativa de recompensas acumuladas, se partindo de um estado j, tomando uma ação a e chegando a um estado j'.

$$V^{\pi}(j) = E_{\pi}[R_T | j_T = j] = E_{\pi}[\sum_{k=0}^{T} \gamma^k r_{T+k+1} \,|\, j_T = j] \tag{5}$$

Em que $V^{\pi}(j)$ é o valor esperado de retorno para a permanência no estado $j_T = j$, e γ é a taxa de desconto. A expectativa E_{π}, 6, pode ser definida como o somatório de todos os estados e ações possíveis.

$$E_{\pi}\left[\gamma \sum_{k=0}^{T} \gamma^k r_{T+k+1} \,\middle|\, j_T = j\right] = \sum_{a} \pi_T(j,a) \sum_{s'} P_{ss'}^{a} \gamma E_{\pi}[\sum_{k=0}^{T} \gamma^k r_{T+k+2} | \, j_{T+1} = j'] \tag{6}$$

Logo, a equação de Bellman para a solução da função de valor-ação se torna a equação 7 (YOUSEFI; TSIANIKAS; COIT, 2020).

$$Q^{\pi}(j,a) = \sum_{j'} P^{a}_{jj'} \left(R^{a}_{jj'} + \gamma \sum_{a'} \pi(j',a')\, Q_{\pi}(j',a')\right) \quad (7)$$

Em que a' é a próxima ação e j' é o próximo estado. Isso significa que na aprendizagem por reforço, o valor de j é calculado obtendo o valor de j'.

Se as soluções possíveis de π devem retornar $max \sum_{T-0}^{T=\infty} R$ enquanto define $(a_T|j_T)$; quando aplicada a manutenção, π se torna a procura pelo mínimo custo C_T. Assim sendo, $Q^{\pi}(j,a)$ para cada estado pode ser a política ótima para que retorna à ação de manutenção que deve ser realizada naquele momento para aquelas observações, e que retorna $min \sum_{T-0}^{T=\infty} C$.

3.1.2 MDP, processo de decisão de Markov

A ideia básica da modelagem de processos estocásticos, como o MDP, é construir o modelo de um processo que começa a partir de sequências de eventos geradas pelo processo em si mesmo. Ao contrário de modelos determinísticos, os estocásticos não apresentam saídas certas dadas circunstâncias de entrada, mas possuem incertezas capazes de alterar os rumos do modelo, ou seja, permitem que aleatoriedades dirijam as possíveis saídas. Portanto, modelos de processos estocásticos, são sequências de eventos em que a saída, a cada momento, depende de probabilidades. Formalizando, um processo estocástico é definido como uma coleção de variáveis aleatórias $X = \{X_k : k \in T\}$ definidos em um espaço de probabilidade comum, tomando valores em estados comuns j, e indexados por T tal que n ou $[0, \infty]$, que corresponde ao tempo discreto ou contínuo. São classes importantes de processos estocásticos os processos de Markov e as cadeias de Markov (FRANZESE; IULIANO, 2019).

Considerando um conjunto de estados $J = \{j_1, \ldots, j_T\}$, a cadeia de Markov é um processo que começa em um desses estados e move sucessivamente de um estado a outro. Se a cadeia está em j e move-se para j', no próximo estado a probabilidade será $P_{jj'}$. Essa probabilidade é a matriz de transição da cadeia de Markov, e corresponde a probabilidade de um determinado estado ocorrer no próximo momento T. As cadeias de Markov são homogêneas no tempo, o que significa que as probabilidades de transição são estacionárias, dessa maneira, para P não é preciso

haver índice de tempo T. Outra característica é que na matriz de transição, o somatório de todas as probabilidades de uma mesma linha deve ser um, independente da distribuição (FRANZESE; IULIANO, 2019).

A principal característica de um processo de Markov é a propriedade de Markov, que implica que o processo não possui memória, isto é, que a distribuição para o próximo estado depende unicamente do estado corrente. Essa propriedade, estabelece três características distintas (FRANZESE; IULIANO, 2019):

a. O número de possíveis saídas ou estados é finito;
b. As probabilidades são constantes no tempo;
c. A propriedade de falta de memória é satisfeita.

Para Ribeiro (1999), o ponto de partida para compreender a propriedade de Markov é a equação 8.

$$o_{T+1} = f(o_T, a_T, w_T) \tag{8}$$

Em que o_T é a observação num momento T, a_t é a ação tomada e w_t é uma perturbação.

A condição de Markov estabelece que se o agente pode diretamente observar os estados do processo, então o estado sumariza todas as informações relevantes para um momento T e atende ao critério da observação Markoviana. Mas se as observações feitas pelo agente não são capazes de sumarizar todas as informações do processo, compreende-se que uma condição não-Markoviana, equação 9, pode tomar lugar (RIBEIRO, 1999).

$$o_{T+1} = f(o_T, o_{T-1}, o_{T-2}, \dots\ a_T, a_{T-1}, a_{T-2}, \dots, w_T) \tag{9}$$

Há também uma classe de modelos chamados de *hidden Markov88969 models* ou HMM. Neles, o movimento entre os estados é simultâneo ao processo de emissão de uma sequência sujeita à propriedade de Markov. Nos HMMs, a sequência de transição dos estados é escondida, o que significa que os estados não podem ser acessados diretamente, mas somente através da emissão de sequências que deles deriva. Logo, o HMM é definido como uma sequência de estados e probabilidade de

estados, transições, emissões e início, gerando a quíntupla $(j, \varepsilon, P_{jj'}, N, B)$ (FRANZESE; IULIANO, 2019), conforme descrição a seguir:

- $J = \{j_1, \dots, j_n\}$, os estados, sendo n o número de estados;
- $\varepsilon = \{\varepsilon_1, \dots, \varepsilon_n\}$, o vocabulário de símbolos emitidos pelos estados;
- $P_{jj'}: J \rightarrow [0,1] = \{P_{jj'}{}_1, \dots, P_{jj'}{}_n\}$, a distribuição inicial dos estados, ou a probabilidade de iniciar em cada estado;
- $N = \left(a_{jj'}\right) j \in J, j' \in J$, a probabilidade de transitar de j a j';
- $B = \left(b_{jj'}\right) j \in v, j' \in J$, em que ε é a probabilidade de emissão quando se está no estado j.

Os HMMs ajudam a modelar problemas em que os estados não são diretamente observáveis. A ideia central é que o HMM se torne um gerador de sequencias, que podem ser inclusive, sequencias de observações (FRANZESE; IULIANO, 2019).

Partindo dos princípios da cadeia de Markov, o processo de decisão de Markov, ou MDP, é um formalismo utilizado para descrever problemas de decisão sequenciais em ambientes totalmente ou parcialmente observáveis. Esses processos podem ser resolvidos através da aprendizagem por reforço (RODRIGUES, 2018). Para (RIBEIRO, 1999), esse formalismo, bem estabelecido matematicamente, simplifica a modelagem enquanto facilita a rastreabilidade dos resultados do agente. De acordo com (YOUSEFI; TSIANIKAS; COIT, 2020), usar o MDP para formular a degradação de componentes é uma maneira de otimizar dinamicamente o planejamento de manutenção enquanto se permite diferentes processos de degradação simultâneos. A solução do MDP passa por encontrar a política $\pi(j)$, para que o solucionador possa escolher a ação a quando estiver no estado j (das Neves Rodrigues, 2018). A política é matematicamente descrita na equação 10.

$$\pi(j) = \overset{argmax}{a \in A} \sum_{j \in J} P_a(j, j') | (j') \tag{10}$$

3.1.3 Q-Learning

O *Q-learning*, apresentado em 1989 por Watkins (1989), é uma evolução dos métodos baseados em diferenças temporais, que em resumo, realizam o cálculo dos

custos para uma dada política π. Esses métodos, que somam a grande maioria ainda hoje, são derivados do método da diferença temporal (TD) apresentado por Sutton em (1988), que já possuía duas vantagens sobre métodos convencionais de aprendizado e predição, o fato de ser incremental e a facilidade de computação. O *Q-learning* mantém essas duas características e as evolui removendo a necessidade de calcular os custos esperados de uma determinada política iterativamente. Nele, cada par (j, a), também denominado função Q, é armazenado em uma tabela, e a ação escolhida para cada estado, isto é, a 'política', é o maior valor da tabela para o par estado-ação. Dessa maneira, os problemas de decisão sequenciais são resolvidos através da expectativa dos reforços futuros quando tomada uma ação e seguida a política ótima. Assim, sob uma política π, o valor verdadeiro de uma ação a em um estado j é dado pela equação 11 (HASSELT, VAN; GUEZ; SILVER, 2015).

$$Q_\pi(j, a) = E[r1 + \gamma r2 + \cdots \mid J_0 = j, A_0 = a, \pi] \quad 11$$

Destarte, o valor ótimo de $Q_\pi(j, a) = max_\pi Q_\pi(j, a)$, e a política pode ser facilmente derivada dos valores ótimos selecionando o valor máximo para cada j. Quando, no entanto, não é possível aprender os valores de todos os estados separadamente, usa-se uma solução baseada numa série de observações. Assim sendo, o valor ótimo se torna $Q(j, a; \Theta_t)$, e dessa forma a atualização da tabela Q para uma ação a_t no estado j_t, recebendo r_T, e resultando no estado j_t, se transforma na equação 12 (HASSELT, VAN; GUEZ; SILVER, 2015).

$$\Theta_{T'} = \Theta_T + \alpha \left(Y_T^Q - Q(j_T, a_T; \Theta_T)\right) \nabla_{\Theta T} Q(j_T, a_T; \Theta_T) \quad (12)$$

Em que α é um escalar e Y_T^Q é definido como a equação 13.

$$Y_T^Q = r_{T'} + \gamma^{max}_a Q(j_{T'}, a; \Theta_T) \quad (13)$$

Essa atualização segue a ideia utilizada nos chamados gradientes descendentes, isto é, o valor corrente de $Q(j_T, a)$ (HASSELT, VAN; GUEZ; SILVER, 2015).

A obtenção da tabela Q, de acordo com Félix Júnior (2022) pode ser apresentada no seguinte pseudocódigo:

Algoritmo 1. Q-Learning.

Iniciar $Q(j, a)$ arbitrariamente:

Enquanto Q não convergir faça:

Iniciar j:

Para cada passo do episódio faça:

Escolher a partir de j usando a política derivada de Q;

Aplicar a;

Observar r, j;

$Q(j, a) \longleftarrow (1 - \alpha)Q(j, a) + \alpha[r + \gamma max_a Q(j', a)]$;

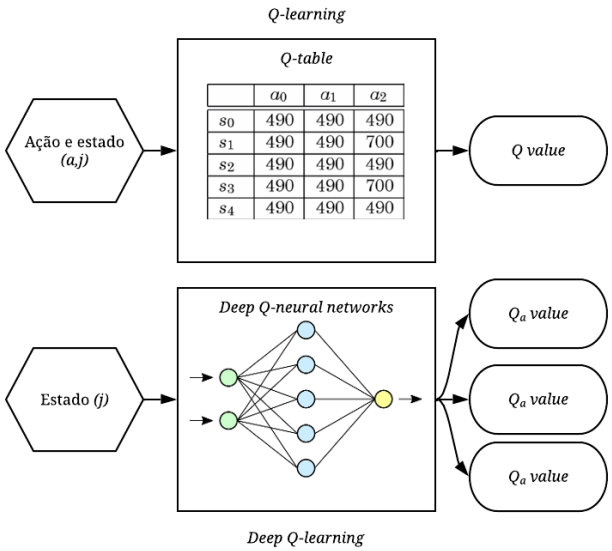
Fonte: Félix Júnior (2022).

Diante dos outros métodos de aprendizado de máquina, há três principais motivos que indicam o motivo da popularidade do *Q-learn*. De acordo com Ribeiro (1999), eles são o pioneirismo no uso desse método para problemas de controle com sucesso, segundo é o grau de autonomia no aprendizado, já que automaticamente encontra a ação ótima sem necessidade de cálculos intermediários ou modelos; e terceiro a sua performance melhor frente os outros algoritmos de aprendizagem por reforço.

3.1.4 Deep Q-Network e *Deep Q-learning*

Primeiramente, quando a AR é combinado a redes neurais, recebe o nome de *Deep*-AR. Construído sobre o *Q-learning,* o *Deep Q-network* ou DQN aprende a política ótima *Q* através de uma rede neural, que trata de aproximar a função que define (a_T, j_T), conforme Figura 3. Nota-se que nessa figura, na *Q-table,* o estado é denotado como s_k ao invés de j_k, enquanto a ação permanece como a_k.

Figura 3. Q-learning e Deep Q-learning.



	a_0	a_1	a_2
s_0	490	490	490
s_1	490	490	700
s_2	490	490	490
s_3	490	490	700
s_4	490	490	490

Fonte: O autor, inspirado em Kim S. et al (2019).

No *Deep Q-learning,* a rede neural recebe o estado como entrada e extrai como saída *Q-values* para cada uma das ações possíveis, e então a maior saída, se torna a ação a ser tomada. O formato da rede é flexível, e pode ter ou não camadas ocultas, assim como camadas de ativação, regularização e diversas estruturas diferentes, como por exemplo as chamadas redes convolucionais, normalmente usadas para estudos que envolvem reconhecimento de imagens (KIM, S. *et al.*, 2019),densas e completamente conectadas, usualmente aplicadas a problemas de controle como o de (CORREA-JULLIAN; LÓPEZ DROGUETT; CARDEMIL, 2020), e recorrentes, normalmente usadas para reconhecimento e processamento de linguagens.

Para um estado *n-dimensional* e um conjunto de ações *m-dimensional*, a rede neural é uma função $\mathbb{R}^n$ a $\mathbb{R}^m$. Além disso, outras características diferenciam o *Q-learning* do DQN:

1. O algoritmo faz uso do chamado *experience-replay;*
2. O algoritmo faz uso do chamado *target-network*;
3. A função de perda ou *loss-function* é a média do quadrado do valor previsto de *Q* e o valor alvo de *Q*.

O *experience-replay* é um dispositivo que ajuda o agente a não esquecer as ações predecessoras ré executando-as. Em períodos determinados, o algoritmo é

autorizado a retirar de um banco de experiências *batches* que alimentarão a rede neural, a atualizarão e ajudarão a estabilizar o processo de aprendizado. O chamado *target-network* segue o mesmo princípio usado no *Q-learning* padrão, com a exceção de que os parâmetros de Θ são copiados a cada T a partir da rede neural, se tornando Θ_T e posteriormente sendo fixados para todos os T seguintes. A diferença está na equação 13 (HASSELT, VAN; GUEZ; SILVER, 2015).

$$Y_T^{DQN} = r_{T,} + \gamma^{max}_{a} Q(j_{T,}, a; \Theta_T) \tag{14}$$

3.1.5 Double Deep Q-learning

De acordo com Hasselt (2015), o $maxQ$ do DQN e do *Q-learning* usam os mesmos valores para selecionar uma ação e avaliá-la, o que leva a sobre estimação de valores, criando vieses de confirmação. Para resolver isso, uma solução é desacoplar a seleção da avaliação, e essa é a ideia por trás do *Double Deep Q-learning,* DDQN. No entanto, o DQN não é capaz de resolver uma dificuldade relacionada ao viés da formação do par valor-ação no *Q-learning*. Tanto o DQN quanto o *Q-learning* são particularmente otimistas devido à falta de flexibilidade do aproximador da função valor (HASSELT, VAN; GUEZ; SILVER, 2015). Uma solução, que pode ter efeitos positivos na solução de problemas de manutenção, é a modificação da função do *target network* para a equação 15.

$$Y_T^{doubleQ} = r_{T,} + \gamma Q(J_{T,}, {}^{max}_{a} Q(J_{T,}, a; \Theta_T); \Theta'_T) \tag{15}$$

A grande diferença está na aplicação de um segundo aproximador Θ'_t. Isso significa que simultaneamente, duas funções valor aprendem, e cada uma se encarrega de definir um dos valores Θ. Para cada atualização da rede, um conjunto de pesos é usado para determinar a política e outro, o valor. Ou seja, a grande mudança está na atualização da função valor. Ainda observando a equação 15, a seção ligada ao *Q-learning* ainda estima os valores da política, mas apenas os pesos de Θ'_T são usados para avaliar a política vigente (HASSELT, VAN; GUEZ; SILVER, 2015).

A ideia do DDQN é manter o máximo possível do algoritmo do DQN intacto, para com o mínimo de alterações necessárias obter os benefícios esperados, e dessa

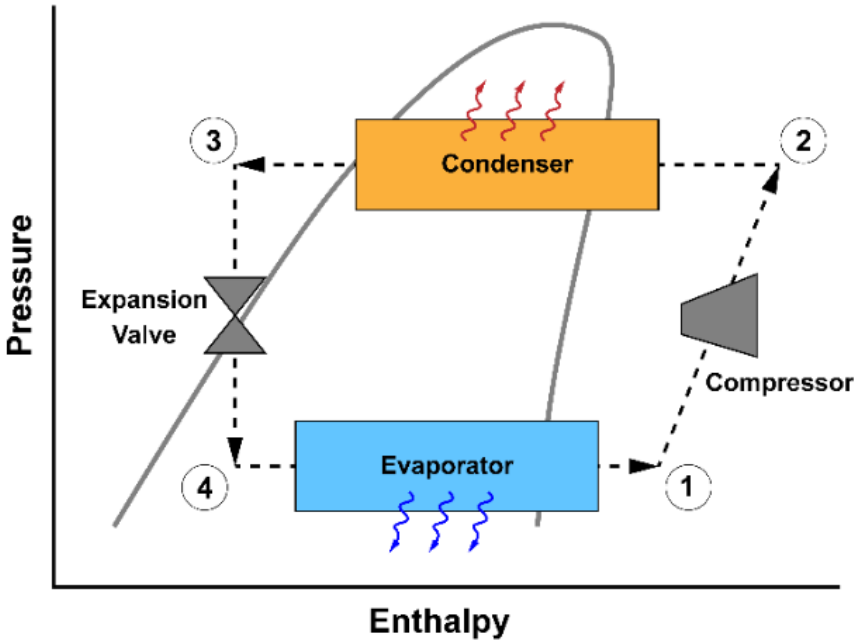
maneira, reduzir a carga computacional que seria derivada da aplicação de dois algoritmos de aprendizado, com duas redes neurais em paralelo sendo atualizados simultaneamente. Por esses motivos, nesse trabalho, o DDQN foi escolhido como o algoritmo de aprendizado do agente.

3.2 REFRIGERADORES BASEADOS EM COMPRESSÃO MECÂNICA DE VAPOR, PRINCÍPIOS E COMPONENTES NO CONTEXTO DO AMBIENTE

Sistemas de refrigeração realizam a transferência de calor de um ambiente para outro através de trabalho. Dessa maneira, eles conseguem reduzir a temperatura de um ambiente isolado, objetos ou mercadorias. Os refrigeradores, freezers e ar-condicionado funcionam com base nesse princípio básico, a diferença entre eles está no tamanho do recinto, nas características operacionais e tecnologias usadas (WANG, S. K. (Shan K.; LAVAN; NORTON, 2000).

Os refrigeradores baseados em compressão mecânica de vapor correspondem a maioria dos dispositivos de refrigeração de uso comercial e residencial. Eles costumam possuir quatro componentes principais, um compressor, um evaporador, um condensador e uma válvula de expansão, conforme a Figura 4.

Figura 4. Ciclo de refrigeração por compressão de vapor.



Fonte: Ding, Subiantoro e Norris (2021).

O principal elemento corresponde ao compressor, onde todo o processo inicia. No instante $1 \rightarrow 2$, o compressor recebe o gás na forma de vapor saturado seco vindo do evaporador e aumenta sua pressão. Nesse processo, o líquido ganha energia, que agora precisa ser dissipada. A dissipação dessa energia indesejada se dá no condensador, entre os instantes $2 \rightarrow 3$. Ele é uma interface de troca de calor que tem como principal objetivo resfriar o fluido refrigerante até chegar ao ponto de saturação 3, onde ele no estado de líquido saturado entrará na válvula de expansão ou tubo capilar. No instante $3 \rightarrow 4$, o fluido é expandido e inicia seu retorno ao estado de vapor. De $4 \rightarrow 1$, esse vapor passa por outra interface de troca de calor, agora voltada para o ambiente acondicionado, e recebe calor latente até retornar ao estado de vapor saturado seco. Quando volta para 1, já tendo recebido o calor do ambiente acondicionado, o fluido reinicia o processo, transportando a energia absorvida para o ambiente externo. É valido adicionar que os ciclos de compressão de vapor têm diversas não-linearidades e complexidades que os tornam difíceis de modelar e controlar (DING; SUBIANTORO; NORRIS, 2021).

De acordo com Knowles e Baglee (2012), os freezers possuem duas atividades de manutenção que são importantes para conservação de suas capacidades, que são checar as temperaturas de operação e a quantidade de poeira de detritos nos componentes. Mas além disso, os freezers costumam apresentar falhas em quatro componentes:

1. Falha de carga de fluido: Um dos principais problemas enfrentados pelos freezers é a perda de fluido refrigerante. Operando em quantidade superior ou inferior ao ideal, a performance diminui, o consumo aumenta e equipamentos mecânicos como juntas, selos e camisas podem ser danificados;
2. Perda de performance do compressor: O compressor é o coração do sistema. Em sistemas de refrigeração simples, eles são constituídos de um motor elétrico e uma bomba alternativa, e ambos perdem eficiência ao longo da vida útil. Essa perda costuma ser derivada da presença de gases indesejados no circuito, umidade interna, infiltrações, corrosões, falhas em protetores térmicos, perda de características do lubrificante e até impactos;

3. Perda de performance do condensador: O condensador, muitas vezes exposto, está sujeito ao acúmulo de sujeira nas superfícies de contato com a AR. Quando isso ocorre, seja por poeira ou corrosão, sua capacidade fica comprometida. Quando há a obstrução completa e não é possível realizar a troca de calor, o compressor pode ser danificado;
4. Perda de performance do evaporador: O evaporador deve ser mantido livre de obstruções, e de preferência, sem gelo. A presença de isolantes indesejáveis próximo a superfície de contato do evaporador leva a perda de eficiência do sistema e aumento do consumo energético.

Além desses componentes Knowles e Baglee (2012) também argumenta que é importante checar periodicamente o estado das vedações, que são importantes para a conservação da temperatura interna após a estabilização da temperatura.

3.3 DIAGNÓSTICO TERMODINÂMICO E MÉTODO DA RECONCILIAÇÃO NO CONTEXTO DO AMBIENTE

Dito tudo isso, o diagnóstico termodinâmico em sistemas de compressão mecânica tem como objetivo encontrar condições fora do funcionamento nominal, pois elas podem levar a falhas ou redução de performance. O objetivo é estabelecer condições de referência e com base nelas, averiguar as medições reais. A análise envolve todo o sistema, e tem como objetivo fornecer informações capazes de assinalar políticas operacionais que reduzam os custos de manutenção e operação, traduzindo em melhor consumo e maior confiabilidade (MENDES *et al.*, 2011).

Um dos métodos para obtenção do diagnóstico termodinâmico é o da chamada reconciliação (MENDES *et al.*, 2011). Nesse método termo econômico não há a descrição de fluxos exergéticos, consumo exergético unitário e indicadores de deterioração no ciclo térmico. Todavia, a complexidade do sistema é representada através de linearizações e artifícios matemáticos que convergem para normalmente para os produtos da instalação. O objetivo do modelo baseado em reconciliação é obter quais os elementos que interferem nas taxas de calor.

A partir da taxa de calor, o sistema estabelece uma condição de operação e teste, também conhecida como (TOP) ou condição atual, e busca restaurá-lo para a condição de referência, ou (ROP), através de ações de manutenção que analisa a

influência da avaria dos elementos nessas taxas de calor, que por consequência, alteram consumo energético e a performance no geral. A condição de referência do balanço térmico deve ser calibrada, no entanto, levando em consideração as condições geográficas da instalação ou a condição ISO (MENDES *et al.*, 2011).

Destarte, o método da reconciliação pode ser utilizado para comparação e identificação de discrepâncias em sistemas de refrigeração a partir de séries temporais de referência e medição. Entre as muitas abordagens para isso, pode-se citar, por exemplo, as baseadas em balanço de massa e modelo para predizer a performance esperada do sistema de refrigeração baseada em diversas variáveis de entrada, como a temperatura e a taxa de fluxo. Todavia, também é possível utilizar o método da reconciliação sem modelos ou balanços, simplesmente usando as séries temporais, e ancorando as medições em *benchmarks* e performance passada. Essa abordagem pode ser particularmente útil quando há limitações de informação ou a modelagem é muito complexa (DUMONT; QUOILIN; LEMORT, 2016; MARTÍNEZ-MARADIAGA; BRUNO; CORONAS, 2013).

Uma abordagem para construir modelos simplificados baseados em dados utilizando o método da reconciliação é a seguinte (CRACIUN; LECOQ; DIGAVALLI, 2019):

1. Dados de série temporal: Variáveis relevantes como temperatura, pressão, consumo etc.;
2. Processamento dos dados: O emprego de técnicas para filtragem e preparação dos dados, como análise de regressão ou algoritmos de aprendizado de máquina para identificar padrões e tendências;
3. Método de reconciliação: O ajuste de parâmetros do modelo ou o uso de abordagens orientadas a dados para ajustar o modelo aos dados;
4. A validação do modelo: A comparação do modelo com as medições reais para avaliar sua precisão e confiabilidade.

3.4 MODELAGEM DE REPAROS IMPERFEITOS NO CONTEXTO DO AMBIENTE

Como já apresentado, o agente de aprendizagem por reforço necessita interagir com o ambiente para construir a política. Quando essa política é relacionada a DDF, surge uma questão: como modelar os efeitos das ações nos componentes? Normalmente, essa ação, comissionada pelo agente, será performada por um técnico,

que mesmo aqui fazendo parte do *test bench*, é um terceiro componente. E como estimar os efeitos da ação do técnico no componente? Ambos os tópicos serão esclarecidos no item 4.4. Por hora, se apresentará a base teórica que dirige a abordagem desse trabalho.

De acordo com Doyen e Gaudoin (2004), duas suposições sobre a eficiência de reparos são conhecidas como o reparo mínimo, ou *as-bad-as-old,* ABAO, e o reparo perfeito, ou *as-good-as-new* AGAN. No modelo ABAO, entende-se que após o reparo, o sistema retorna ao estado em que estava antes dele, já no AGAN, sempre que o sistema é reparado, ele retorna ao estado de novo.

Uma das formas de modelar reparos do tipo ABAO, é através do método da redução da intensidade, também conhecido com *arithmetic reduction of intensity model,* ARI. Nesse método, cada reparo reduz a intensidade da falha dependendo do passado dos processos de falha; após a falha, a velocidade de desgaste é a mesma de antes da falha, e entre duas falhas, a intensidade de falhas é verticalmente paralela a velocidade de desgaste inicial (DOYEN; GAUDOIN, 2004).

Uma das modelagens da ARI, também conhecida como ARI_{∞}, considera que um reparo é capaz de reduzir a intensidade de uma falha no nível proporcional a intensidade de falha corrente. Esse princípio está exposto na equação 16.

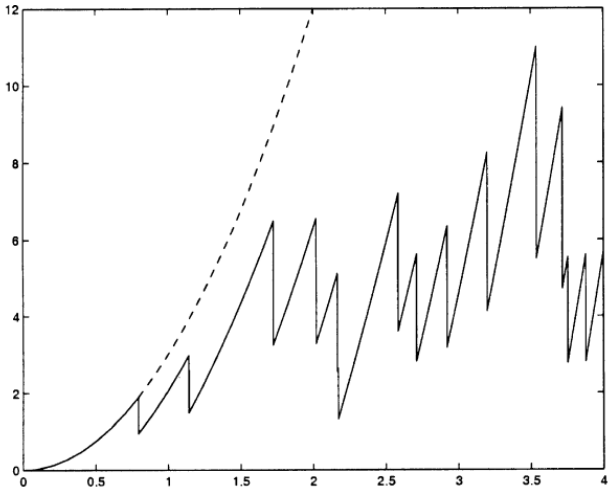
$$Dc_{T_i^+} = Dc_{T_i^-} - \rho Dc_{T_i^-} \tag{16}$$

Em que Dc é a intensidade de falha e ρ é um parâmetro de redução de falha $\rho \in \{0,1\}$. Nesse modelo, a intensidade de falha é computada conforme equação 17.

$$Dc_T = Dc_T - \rho \sum_{j=0}^{n_{T-1}} (1-\rho)^i \quad Dc(T_{n-j}) \tag{17}$$

De acordo com Doyen e Gaudoin (2004), se $n_T = 0, Dc_T = Dc(T)$, ou seja, o modelo segue a descrição ABAO. Para facilita a visualização, a Figura 5, desenha ambos a intensidade inicial, em pontilhado, e os resultados das intervenções para $\rho = 0{,}5$.

Figura 5. ARI_{∞} para $\rho = 0{,}5$.



Fonte: Doyen e Gaudoin (2004).

4 DESENVOLVIMENTO DA FERRAMENTA DE TREINAMENTO E AVALIAÇÃO

A ferramenta de treinamento e teste desse trabalho tem como base o refrigerador Fricon® modelo HFEB 311 C, Figura 6, que foi cedido e paramentado para estudos paralelos, e de onde se obteve os dados reais para calibragem do modelo utilizado no presente estudo. Esse freezer é uma unidade de congelamento voltada primariamente para a conservação de perecíveis que exijam baixas temperaturas com confiabilidade, como sorvetes, uma vez que a perda da capacidade de refrigeração pode levar a avaria da carga acondicionada.

Figura 6. Freezer Fricon HFEB 311.

Fonte: Fricon (2021).

O equipamento constitui-se de um freezer padrão de abertura superior munido de compressor, evaporador e condensador estáticos, termostato e demais itens de controle, operando em ciclo de compressão mecânica. Os detalhes de operação podem ser visualizados na Tabela 1.

Tabela 1. Informações gerais do freezer Fricon HFEB 311 C.

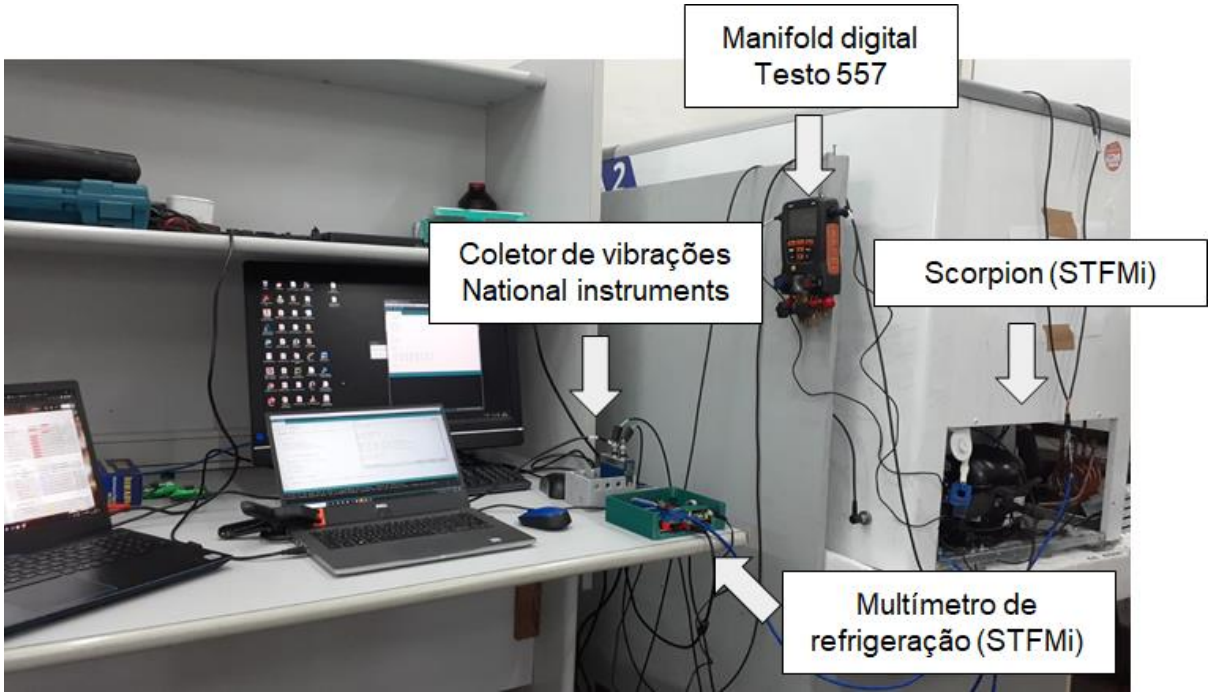
Elemento	Dado	Unidade
Capacidade bruta	311	Litros
Variação da temperatura	-30 a -18	ºC

Consumo de energia	2,4	kW/dia
Tipo de gás	R290	
Tampas	1	

Fonte: Fricon (2021).

No IFPE, o freezer recebeu sensores e sistemas de coleta de dados, como o coletor de vibrações, um manifold digital, um multímetro de refrigeração e dispositivos desenvolvidos no próprio instituto, Figura 7.

Figura 7. Aparato coletor de dados do freezer.



Fonte: Novaes (2022).

Com esse aparato, diversas medições invasivas e não invasivas foram extraídas, conforme Quadro 4.

Quadro 4. Medições invasivas e não invasivas obtidas do freezer.

Elemento	Medição invasiva	Medição não invasiva
Multímetro de refrigeração STFMi	Temperatura: Saída do condensador; Entrada do condensador; Meio do evaporador (1); Meio do evaporador (2);	Temperatura: Ambiente externo; Ambiente interno; Descarga do compressor;

	Entrada do evaporador; Saída do evaporador.	Sucção do compressor; Corrente.
Coletor de vibrações NI	-	Vibração do compressor; Vibração da carcaça.
Manifold digital Testo	Pressão alta; Pressão baixa.	-
Scorpion STFMi	Vibração triaxial.	-

Fonte: O autor (2022).

Medições invasivas são aquelas que precisam alterar elementos de funcionamento do freezer para serem obtidas. As pressões de alta e baixa, por exemplo, precisam ter acesso a ambientes selados, a saber, o condensador e o evaporador, respectivamente, alterando elementos do freezer. As não invasivas, por outro lado, não alteram elementos do freezer e em sua maioria, são obtidas através do contato com a superfície dos elementos.

4.1 O *TEST BENCH*

O programa proposto de manutenção baseada em condição é ancorado na aprendizagem por reforço. Para isso, a exemplo da pesquisa realizada por Mahmoodzadeh *et al* (2020), devem ser providos pelo sistema ambas as observações e o sinal de reforço. A melhor maneira de avaliar a performance de um algoritmo de aprendizagem por reforço é conhecendo os limites das ações e resultados.

Por isso, se construiu uma bancada de teste, ou *test bench*, onde os parâmetros do ambiente e do agente podem ser manipulados e estudados. Esse é um procedimento comum em estudos de AR, e tem por objetivo simplificar a integração entre o ambiente e o agente, que apresenta desafios superiores ao aprendizado não supervisionado e não supervisionado, como por exemplo a necessidade de interação para formação da política, e de algoritmos para estruturação de recompensas e pênaltis, ambos problemas de escopo aberto, por isso, o *test bench*:

1. Simula a degradação de componentes independentes do refrigerador usando um HMM, com emissões dependentes do estado e ajustáveis a múltiplas ações concorrentes de manutenção com precisão de cinco minutos;

2. Simula a operação do refrigerador e seus parâmetros de observação com precisão de cinco minutos;
3. Provê a interação entre o refrigerador e as ações do técnico de manutenção, fora do controle do agente e que se constituem de incertezas;
4. Aproxima o custo das manutenções e substituições e o estado de degradação do refrigerador.

Convém ressaltar que como no estudo conduzido por Mahmoodzadeh *et al* (2020), o agente não tem acesso ou conhecimento sobre o funcionamento interno do ambiente, por isso para ele o ambiente é tipo *'black-box'*. E nesse sentido, é razoável limitar o *test bench* nos seguintes sentidos:

O sistema corresponde a um refrigerador, que no instante $\Theta_{T,j} = 0$ é considerado novo e a degradação de cada um dos componentes habilitados desenvolve-se através do tempo e das emissões do HMM.

1. Independente do j_T, há degradação. A diferença é a intensidade com que ela ocorre e por consequência, quão rápido o componente atinge a falha completa, ou seja, deixa de performar sua função. A velocidade de degradação está sujeita a uma emissão gerada do estado do HMM, que será apresentada no item 4.4.2;
2. Dois tipos de manutenção são considerados, a saber, a preventiva e a substituição do componente defeituoso. A manutenção preventiva inclui um custo e a recuperação da performance daquele componente em um nível desconhecido após a intervenção do técnico. A substituição inclui o custo e a recuperação completa da performance daquele componente após a intervenção do técnico;
3. Sempre que um novo episódio é iniciado, considera-se que se trata de um novo refrigerador, com probabilidades e custos diferentes;
4. Os custos e a efetividade da manutenção corretiva, preventiva e preditiva são igualmente incertos até que o processo da manutenção ou substituição esteja concluído;
5. A cada cinco minutos, o agente decide por nada fazer, realizar uma manutenção preventiva ou substituir um componente. A ordem do agente é aplicada tão logo quanto possível pelo técnico. O período de

cinco minutos se dá em benefício da capacidade computacional disponível, e não é maior por causa do tempo compreendido entre as temperaturas de disparo e descarregamento do compressor do HFEB 311C.

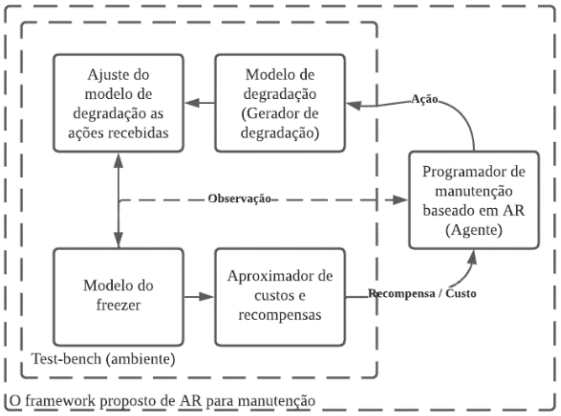
As seções seguintes detalharão a implementação desse *test bench*. Primeiro, será realizada a apresentação dos vários componentes e suas interações, e após os detalhes serão elaborados.

4.2 COMPONENTES DO *TEST BENCH*

O *test bench,* completamente construído em Python™, possui quatro componentes. O primeiro é o modelo do freezer, um conjunto de parâmetros operacionais que imitam o funcionamento do freezer. Posteriormente, os resultados são submetidos ao aproximador de custos e recompensas. Após, o modelo de degradação ajusta o estado do freezer às ações do programador de manutenção e, em seguida, o ajuste traduz os novos parâmetros de desempenho para o próximo iteração da operação do freezer. A Figura 8 apresenta um resumo do *test bench* proposto. O agente, reitera-se, não é parte dele, e interage com os quatro componentes dele a cada cinco minutos através do conjunto de ações, observações e recompensas. As escolhas de manutenção, reitera-se, são limitadas as ações a que o modelo de refrigerador pode se ajustar. Se o sinal de reforço ou custo/recompensa deve guiar a interpretação das ações do agente, as observações têm a função de resumir tudo o que acontece no sistema, fornecendo informações suficientes para que ele possa aprender e tomar decisões razoáveis. Nesse trabalho, definiu-se quatro dimensionalidades ou características como as observações:

1. A temperatura interna τint_T;
2. A corrente I_T;
3. O estado da porta do refrigerador;
4. A variável de tempo ϑ_T, que será apresentada no item 4.6.4.

Figura 8. O test bench proposto para a avaliação da interação entre a manutenção a proposta pelo agente e a degradação do refrigerador.



Fonte: O autor, inspirado em Mahmoodzadeh et al (2020).

4.3 O MODELO DO FREEZER

Como já apresentado, o agente de aprendizagem por reforço requer uma série de observações para definir suas ações ao longo do tempo. Parte dessas observações vem do gerador de dados, que corresponde ao que seriam leituras obtidas dos sensores adicionados ao refrigerador. Como também não é necessário que o agente conheça o funcionamento de um refrigerador, isto é, o programa de manutenção não possui modelo, é dispensável modelar pormenorizadamente as operações termodinâmicas do freezer. São necessários somente os dados de saída que seriam utilizados como observações.

Além do mais, a exploração do algoritmo e a repetição dos experimentos podem ter custo computacional extremamente alto. Para resolver isso, utilizou-se alguns artifícios matemáticos baseados no método já apresentado no item 3.3 mirando o mínimo de prejuízo à verossimilhança com o objeto real, mas buscando a máxima otimização dos recursos disponíveis.

4.3.1 Geração de dados de temperatura e corrente

O funcionamento de um freezer é repleto de transformações termodinâmicas e não linearidades. Mas a maioria dos parâmetros de funcionamento se relacionam de alguma maneira a temperaturas (ºC) e coeficientes de performance que reverberarão na potência elétrica (W) e temperatura interna (°C).

Os dados de base para calibragem foram extraídos do manual de funcionamento do freezer (FRICON, 2021) e estão disponíveis na Tabela 2.

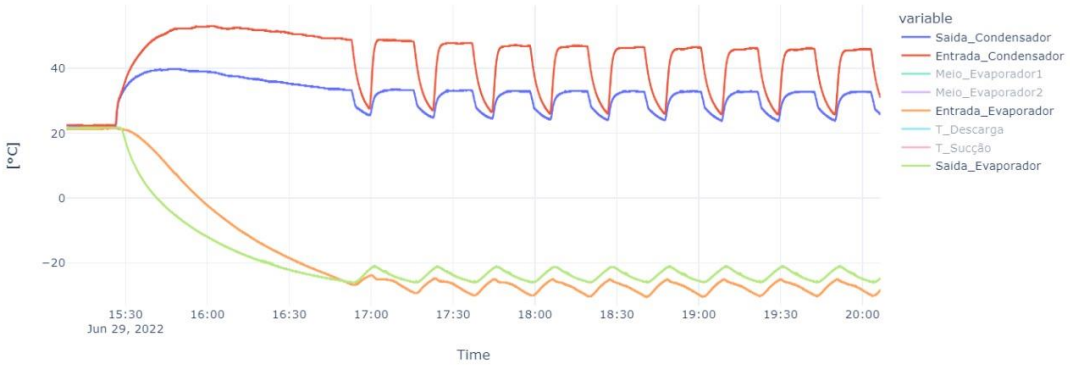
Tabela 2. Dados de calibragem do freezer Fricon HFEB 311 C.

Elemento	Dado	Unidade
Tensão	198 - 240	V
Potência do equipamento	211	W
Potência de iluminação	8	W
Corrente máxima	1,31	A
Temperatura de operação	-30 a -18	°C

Fonte: O autor (2022).

O comportamento do freezer foi extraído da unidade disponibilizada para testes em trabalhos correlatos no campus Recife do Instituto Federal de Pernambuco (IFPE). Confirmando que a unidade estava em estado de referência, isto é, pleno funcionamento sem anormalidades, realizou-se a medição do início do funcionamento até o desligamento do aparelho, dessa maneira obtendo o processo de refrigeração do interior do freezer e seu comportamento após a estabilização da temperatura. A seleção das características de observação priorizou as chamadas medições não invasivas, que aqui são a corrente de entrada do aparelho e a temperatura interna. Um recorte das medições de temperatura está disponível na Figura 9.

Figura 9. Medições de temperatura partindo da inércia por cinco horas de funcionamento do refrigerador.



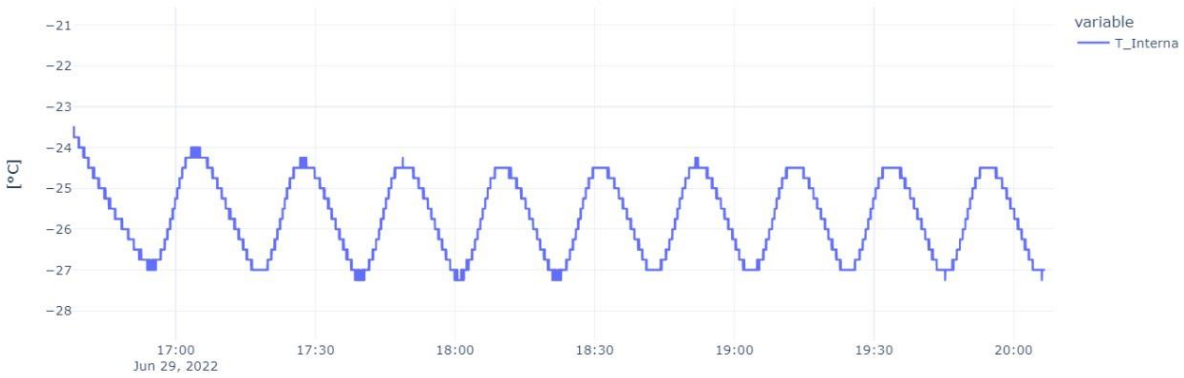
Fonte: O autor (2022).

Na Figura 9 podem ser visualizadas as temperaturas da entrada do condensador, em vermelho, da saída do condensador, em azul, da entrada do

evaporador, em laranja, e da saída do evaporador, em verde. Uma quarta medição foi adicionada ao interior do conservador em vias de obter a temperatura da AR acondicionado após a inércia da realização das trocas termodinâmicas.

Desde que o modelo parte do ponto de vista de que o agente tem conhecimento limitado acerca do funcionamento do freezer, e que a otimização do seu funcionamento depende principalmente da conservação dos estados ideais de temperatura e consumo, a partir da medição da temperatura do interior do freezer é que foram feitas as calibragens de funcionamento. Um recorte da seção de calibragem está disponível na Figura 10.

Figura 10. Recorte da oscilação normal da temperatura interna do freezer em bom estado de conservação após estabilização.



Fonte: O autor (2022).

Como o freezer HFEB 311 C não possui a tecnologia do compressor de velocidade variável, a medição disponível na Figura 10, do interior do freezer após a estabilização da temperatura, apresenta um ciclo. Ele corresponde aos momentos em que o fluxo de calor para o interior do aparelho é revertido pelo trabalho do compressor. Quando o compressor é ligado, isto é, a temperatura é maior que a temperatura de disparo $max\tau$, o trabalho do compressor transfere calor do interior para o exterior, e quando a temperatura é menor que a temperatura de descarregamento, $min\tau$, o compressor é desligado e por condução, através das interfaces de contato entre o interior e o exterior, o interior volta a ser aquecido até $max\tau$ ser novamente alcançada e o processo é reiniciado. Na condição de referência, tanto o aquecimento quanto o resfriamento possuem taxas estáveis, sinalizando o bom estado dos componentes envolvidos.

A obtenção das taxas de calibragem parte do princípio utilizado na lei do resfriamento de Newton, de que a taxa de aquecimento de um corpo, ou seja, sua velocidade, neste caso em ºC/s, é diretamente proporcional à diferença de temperatura τ entre os corpos multiplicada a uma constante de proporcionalidade ζ. Esse conceito pode ser matematicamente expresso na equação 18 (MARUYAMA; MORIYA, 2021).

$$\frac{d\tau}{dT} = \zeta * (\tau - \tau_m) \tag{18}$$

Sendo conhecidos os momentos em que o fluxo do calor no interior é revertido pelo trabalho do compressor no HFEB 311 C, selecionou-se da medição dois momentos representativos, um para o aquecimento e outro para o resfriamento. Com eles, através da equação 19, foi possível obter as taxas de aquecimento e resfriamento em ºC/s.

$$q = \frac{(\tau - \tau'_n)}{\Delta T} \tag{19}$$

Em que q corresponde a taxa de aquecimento ou resfriamento em ºC/s, τ à temperatura inicial e τ'_n à temperatura final em ºC e ΔT ao intervalo de tempo em s. Sabendo que as trocas termodinâmicas para $\phi < 0$, sendo ϕ o fluxo de calor, não anulam as ineficiências de vedação do sistema, é necessário somar a taxa de aquecimento à taxa de resfriamento, e com isso, se obtém os dados da Tabela 3.

Tabela 3. Taxas de resfriamento e aquecimento após a estabilização do sistema.

Elemento	Dado	Unidade
$q \forall \phi < 0$	-0,012	ºC/s
$q \forall \phi > 0$	7,288*10-³	ºC/s

Fonte: O autor (2022).

O valor de q não é estático, e ele pode ser modificado por pelo ou menos os quatro motivos considerados nesse trabalho:

- Alterações na vedação: As taxas de aquecimento para o freezer fechado e aberto são diferentes, no algoritmo isso pode ser feito pela substituição de q por um escalar em um intervalo predefinido quando a

porta está aberta. Nesse trabalho, se considera que dr_T é falso, ou seja, a porta está fechada, e esse escalar é 0;

- Mudança na carga a ser refrigerada: Ao longo da vida útil, o freezer trabalhará para refrigerar cargas de volume, capacidade térmica e temperaturas inicial distintas. No algoritmo, isso pode ser feito pela manipulação de q em um intervalo predefinido. Nesse trabalho, se considera que q é multiplicado por um fator 1, ou, sempre estável;
- Degradação dos componentes: A degradação de componentes como o evaporador, o condensador e o compressor, por exemplo, reduzem $q \forall \phi < 0$. Esse tópico será detalhado no item 4.4;
- Estado de operação do aparelho: O respeito aos ciclos e as trocas termodinâmicas de acordo com o estado de q com relação a ϕ após as temperaturas $min\tau$ e $max\tau$.

Portanto, o q utilizado pelas simulações, que é o q_T, é descrito pelas equações abaixo:

$$q_T = \begin{cases} \dfrac{q * \varpi_{lT} * \eta v_T + \eta h_T + sf_T}{3} * \eta g_T, & se\ \phi < 0\ e\ f \neq 0 \\ q = q \forall \phi > 0, & se\ \phi < 0\ e\ f = 0 \end{cases} \quad (20)$$

$$q_T = \begin{cases} q = \varpi_{qT}, & se\ \phi > 0\ e\ dr_T = True \\ q * a * \big(1 + (1 - \eta i)\big), & se\ \phi > 0 \end{cases} \quad (21)$$

Em que q_T corresponde a taxa de transferência de calor ajustada no tempo, baseada em (TROTT; WELCH, 2000), é uma simplificação das operações termodinâmicas do freezer do ponto de vista energético, sendo ϖ_{lT} a manipulação na carga refrigerada, para $\varpi_{lT} \in (1{,}1.5)$; η correspondente ao percentual da performance relacionado ao estado inicial de v, evaporador, h, condensador e g, compressor, e sf a carga de fluído, para $sf \in (0{,}1)$. Para o cenário de aquecimento, $\phi > 0$, considera-se que quando dr_T é verdadeiro, ou a porta está aberta, ϕ é diretamente proporcional a uma taxa aleatória ϖ_{qT}, sendo $\{\varpi_{qT} \in R \big| 2 \leq \varpi_{qT} \leq 4\}$, e ηi a eficiência do isolamento.

Com essas taxas, é possível obter em que temperatura estaria o interior após determinado período em um regime de ϕ. Essa temperatura pode ser obtida através da equação 22.

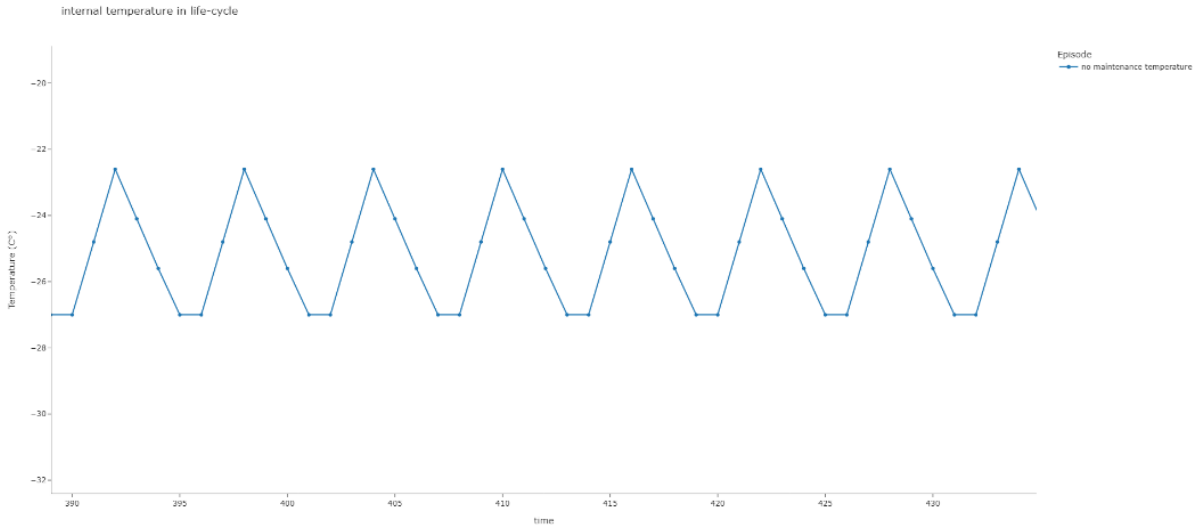
$$\tau' = \begin{cases} (q_T * \Delta T) + \tau, & se\ \tau - \tau' > q\ * 5; \\ (q_T * \delta\tau * \Delta T) + \tau, & se\ \tau - \tau' < q\ * 5 \end{cases} \quad (22)$$

Em que τ' corresponde a temperatura no próximo instante em ºC, τ a temperatura do instante atual em ºC, q_T a taxa de aquecimento em ºC/s, $\delta\tau$ um coeficiente de deflexão de q usado para ajustar a temperatura τ'a $min\tau$. A condicional é necessária por causa do intervalo de cinco minutos entre as leituras, que gera a necessidade de ajustar a curva da temperatura quando τ é muito próximo de $min\tau$. $\delta\tau$ por sua vez, é obtida pela equação 23.

$$\delta T = |\tau - min\tau| \quad (23)$$

Dessa maneira, obtém-se o comportamento semelhante ao observado na Figura 10 e disponível na Figura 11. Nela, os pontos representam as leituras, ligadas pela linha para melhor visualização. Nem sempre o fundo do vale e o topo do monte coincidem com o intervalo de leitura.

Figura 11. Recorte da oscilação normal da temperatura interna do freezer modelado em bom estado de conservação após estabilização.



Fonte: O autor (2022).

Na operação do freezer, a corrente de operação oscila à medida que componentes são adicionados a carga instantânea para um determinado T. O principal componente capaz de alterar a corrente é o compressor, e por isso, são previstas três condições, uma com o compressor ligado, outra com ele desligado e a

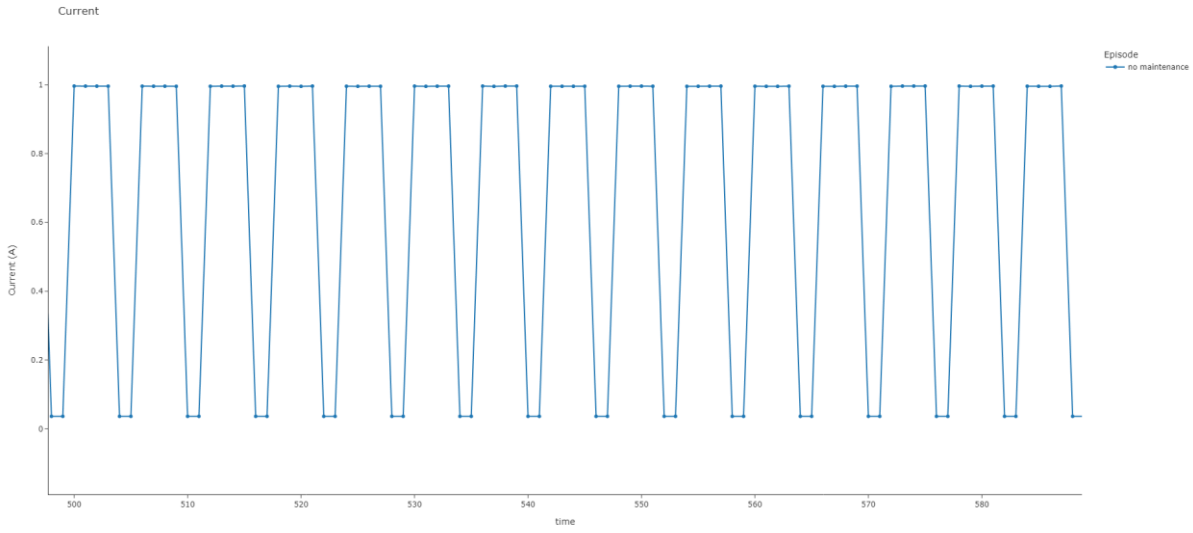
última para o freezer, por completo, desligado. De forma semelhante, de posse dos dados da Tabela 3, é possível modelar a oscilação da corrente na entrada do sistema a partir da equação 24.

$$I = \begin{cases} \dfrac{cg * \varpi_g + cd * \varpi_d}{f}, & se\ \phi < 0\ e\ f \neq 0 \\ \dfrac{cd * \varpi_d}{f}, & se\ \phi > 0 \\ 0, & se\ f = 0 \end{cases} \quad (24)$$

Sendo ϕ a taxa de aquecimento ou resfriamento em ºC, I a corrente em A, cg a potência elétrica do compressor, cd a potência elétrica dos demais componentes, f a tensão em V, ϖ_g uma taxa aleatória dependente do estado $j_T, \forall j_T = 0, \{\varpi_c \in R | 1 \leq \varpi_c \leq 1{,}001\}$ e $\forall j_T = 1, \{\varpi_c \in R | 1 \leq \varpi_c \leq 1{,}01\}$, e ϖ_d uma taxa aleatória constante ao longo da vida útil $\{\varpi_d \in R | 1 \leq \varpi_d \leq 1{,}001\}$.

Destarte, respeitando os intervalos de cinco minutos de resolução entre as leituras, obtém-se a Figura 12, onde semelhantemente, as leituras são representadas pelos pontos ligados por linha para melhor visualização.

Figura 12. Recorte da oscilação normal da corrente de entrada do freezer modelado em bom estado de conservação após estabilização.



Fonte: O autor (2022).

Estando o refrigerador acionado, com respectiva corrente e tempo de funcionamento definidos, pôde-se obter as emissões e consumos. As emissões, por indisponibilidade de dados locais, foram baseadas nos dados de emissões de CO_2

horárias por kWh projetados para a rede elétrica na Europa nos próximos vinte e dois anos (IEA, [s.d.]), interpolando no período compreendido entre 2018 e 2028.

Para usar a mesma base das emissões, o consumo energético usou o histórico da média de preços por kWh da mesma rede para os consumidores residenciais e comerciais dos últimos dez anos (EUROSTAT, [s.d.]), projetando-os para o mesmo período das emissões através do cálculo dos juros compostos.

Destarte, o cálculo do consumo está matematicamente representado na equação 25, da emissão na equação 26, e do custo em moeda corrente, na equação 27 (MUNGUBA *et al.*, 2020). Nota-se que por causa do intervalo de cinco minutos entre as leituras, é preciso limitar os resultados de Co, Em e Ta aos respectivos intervalos de medição. Isso se faz multiplicando o resultado pelo intervalo entre as leituras e dividindo por sessenta minutos. Dessa forma, se obtém os valores para cada intervalo.

$$Co = \frac{\frac{\sum cF_T}{1000} * 5 * 1}{60} \tag{25}$$

$$Em = e_T \frac{\frac{T * \sum cF_T}{1000} * 5 * 1}{60} \tag{26}$$

$$Ta = p_T \frac{\frac{T * \sum cF_T}{1000} * 5 * 1}{60} \tag{27}$$

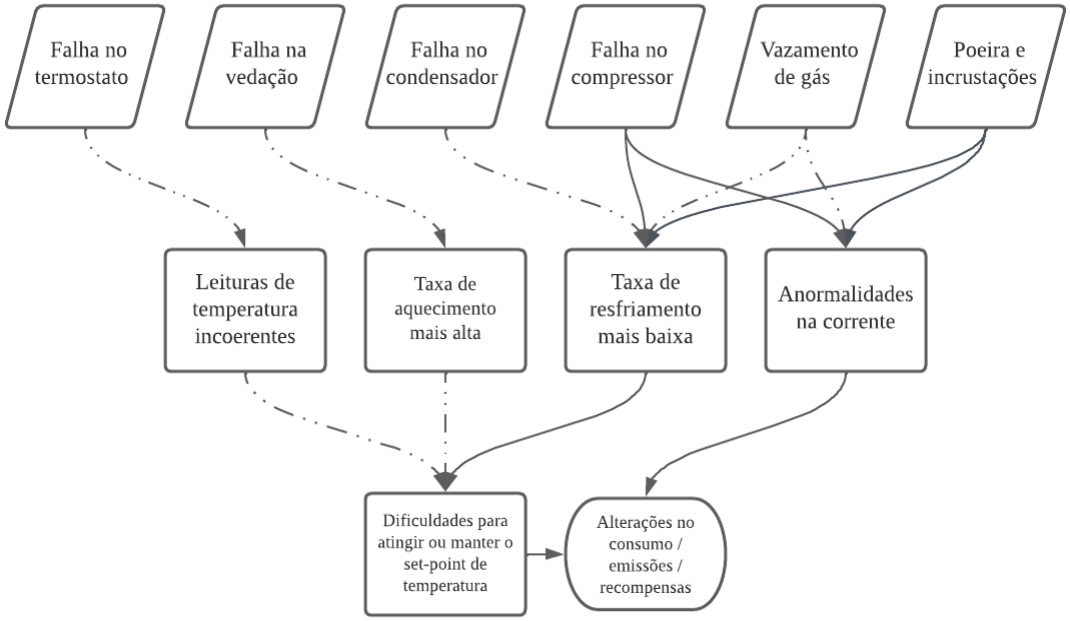
Sendo Co o consumo em kWh e cF_T a potência elétrica dos componentes do freezer para aquele instante, Em as emissões em g de CO_2, e_T as emissões para aquele instante, e Ta a tarifa em €, e p_T o preço do kWh para aquele instante.

4.4 O GERADOR DE DEGRADAÇÃO

Para esse trabalho, considera-se que a degradação do refrigerador ocorre para cada c_T, ou seja, para cada componente, independentemente, e a cada instante. Esse arranjo parte do princípio já apresentado no item 2.3 e explorado em detalhes no trabalho de (YOUSEFI; TSIANIKAS; COIT, 2020). A relação entre o *test bench* e as falhas está na Figura 13. Para cada alteração em c_T, representado pelas caixas superiores, há efeitos esperados. Mas mesmo o *test bench* sendo capaz de executar todas essas interações, optou-se, para esse trabalho, considerar apenas a perda da eficiência do compressor e a presença de poeira e incrustações. Essa decisão

objetivou melhorar a rastreabilidade dos resultados e reduzir o tempo computacional. Toda forma, nada impede que pesquisas posteriores habilitem todos os processos disponíveis no ambiente, e de posse do framework já obtido, explorem as capacidades do agente em lidar com níveis superiores de incerteza.

Figura 13. Elementos do freezer e suas conexões no *test bench*.



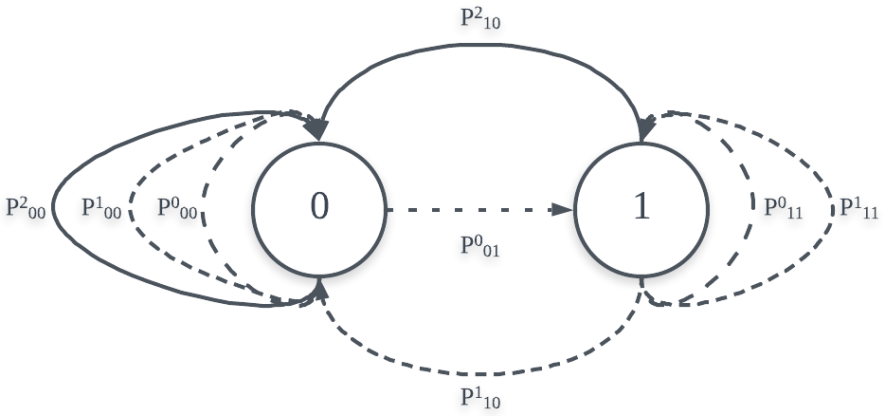
Fonte: O autor (2022).

Para explorar esse ponto, primeiro se apresentará a cadeia de Markov proposta para esse trabalho e em seguida, como se dão as observações derivadas do HMM.

4.4.1 Degradação e cadeia de Markov

Nesse trabalho, cada componente está degradando no tempo, a cada instante. Se o componente não é intervencionado, ele continua em j ou transita para j'. Mas é impossível retornar de j' a j sem a intervenção do agente. Assim, o estado do sistema no tempo T é $j_T = (j1_T, \dots jn_T)$, em que jc_T corresponde ao estado do componente c no instante T, $jc_T \in \{0,1\}$. É possível visualizar essa relação para cada jc_T na Figura 14.

Figura 14. MDP para c_T.



Fonte: O autor (2022).

A Figura 14 mostra a cadeia de Markov para c_T a qualquer momento a partir de T a T', incorporando três ações, nada fazer 0, manutenção 1 e substituição 2. As tracejadas indicam 0, as pontilhadas 1 e as contínuas 2. Se tratando de um HMM, cada um dos estados jc_T não está atrelado ao grau de degradação de cada componente c_T, mas sim a uma emissão, como já apresentado em 3.1.2, que define a velocidade com que cada c_T degrada a cada T. Por exemplo, em $jc = 0$, o componente possui velocidade de degradação $v_c = x$, podendo permanecer em $jc = 0$ por toda a vida útil ou transitar $jc_T = 0 \rightarrow 1$, e em $jc = 1$, $v_c = x + y$. Além disso, no próximo período de inspeção, jc pode mover-se livremente de acordo com as probabilidades estabelecidas para cada ação. Por exemplo, $\mathrm{P}^1{}_{10}$ mostra que há probabilidade de o estado regredir de j' a j mediante a ação 1, ou uma ação de manutenção.

Nesse trabalho, se considera que a transição $jc \rightarrow jc'$ é constante e pode ocorrer a qualquer instante, como definido pela matriz de Markov abaixo:

$$Pc_{j\rightarrow j}^{0} = \begin{bmatrix} 1 - 2{,}5 * 10^{-7} & 2{,}5 * 10^{-7} \\ 0 & 1 \end{bmatrix} \quad (28)$$

Em que $Pc_{j\rightarrow j}^{0}$ é a transição para a ação 0. O elemento $Pc_{j\rightarrow j,11}^{0}$ pode assumir qualquer valor $\in \{0{,}1\}$, e apenas dirige a probabilidade de transição de $jc \rightarrow jc'$ a cada

T. Nesse trabalho, $Pc_{0\to1}^{0} = 2{,}5 * 10^{-7}$ por escolha deliberada. Esse valor foi selecionado por limitar o número de transições a cada episódio de dez anos de simulação, ou 1.520.000 iterações, e pode assumir qualquer valor, sendo proporcional a probabilidade de ocorrência dessa transição a cada iteração, isso é, quanto maior, mais transições ocorrerão a cada episódio. Com o valor apresentado, a grande maioria dos compressores não muda de estado, apenas perde performance em ritmo natural, como se espera de um aparelho real. Apenas uma pequena proporção transita mais de uma vez, e um segundo grupo, que representa a maioria dos episódios em que há transição, transita apenas uma vez, independente do momento T. Isso é importante por que também é necessário observar o desempenho do agente enquanto o sistema não está em falha, mas continua a perder performance pelo desgaste natural, como ocorre à maioria dos freezers (NASCIMENTO; FLESCH, R. C. C.; FLESCH, C. A., 2020). Como já apresentado, esse trabalho considera que apenas o compressor está sujeito a degradações ao longo da vida útil, portanto apenas ele está sujeito a $Pc_{j\to j}^{0}$. Todavia, o modelo considera que cada componente possui um $Pc_{j\to j}^{0}$ independente, com sua respectiva taxa $Pc_{j\to j,11}^{0}$.

Já $Pc_{j\to j}^{1}$ é definido com base em Dc_T, e pode ser expresso na equação 29:

$$Pc_{j\to j}^{1} = \begin{bmatrix} 100 - Dc_T & Dc_T \\ 100 - Dc_T & Dc_T \end{bmatrix} \tag{29}$$

Em que $Pc_{j\to j}^{1}$ é a transição para ação 1, e $Pc_{j\to j,11}^{1}$ é a normalização de Dc_T, que será detalhada na seção 4.4.2.

Assim sendo, a possibilidade de transitar ou permanecer em j mediante $a = 1$ é indexada ao valor de Dc_T, ou a profundidade da degradação. Quanto maior a profundidade da degradação de um componente em j', menor a probabilidade de voltar a j mediante $a = 1$.

Já a ação 2 sempre resulta na transição $j' \to j$.

4.4.2 Degradação e emissões do HMM

Como já apresentado, cada c_T possui uma matriz de Markov. Mas nesse trabalho, os estados não são diretamente observáveis pelo agente, que tem acesso

apenas aos resultados das emissões do HMM, por isso, é preciso de antemão, definir dois pontos:

- A degradação inicia como 0 em $T = 0$, logo $Dc_0 = 0$;
- Dc_T progride independentemente de j_T e de c para todo c em que D_T está habilitado no algoritmo.

Nesse trabalho, a linearidade do processo de degradação é quebrada por uma emissão ε indexada ao estado de Markov corrente jc_T. ε_{cT} é gerado através da função '*weibull_min.rvs'* do módulo '*scipy. stats'* do Scipy, uma biblioteca *open-source* para Python ("scipy.stats.weibull_min — SciPy v1.9.2 Manual", [s.d.]).. Essa função gera *samples* a partir da distribuição de Weibull usando os seguintes argumentos:

- k: O parâmetro de forma da distribuição de Weibull. Este parâmetro determina a forma da distribuição, com valores maiores resultando em uma distribuição mais acentuada, nesse trabalho, os parâmetros adotados foram $k = 5$ para degradação normal e $k = 0{,}2$ para acelerada;

loc: O parâmetro de localização da distribuição de Weibull. Este parâmetro desloca a distribuição para a esquerda ou para a direita ao longo do eixo x, esse trabalho adota o valor padrão da biblioteca;

- $scale$: O parâmetro de escala da distribuição de Weibull. Este parâmetro estica ou encolhe a distribuição ao longo do eixo x, esse trabalho adota o valor padrão da biblioteca;
- $size$: O número de variáveis aleatórias a serem geradas. Isso pode ser um inteiro ou uma tupla de inteiros, especificando a forma do *array* de saída. Nesse trabalho, esse parâmetro é 1, logo a saída corresponde a um escalar apenas, que assume o valor de ε_{cT}.

Assim sendo, cada estado gera uma emissão, que multiplicada ao φc, equação 30, leva a degradação daquele componente naquele instante.

$$\varphi c = \frac{100}{elc} \tag{30}$$

Sendo φc o fator de degradação esperada e elc a vida útil esperada do componente. Nesse trabalho, adotou-se $elc = 2.280.000$ instantes, ou quinze anos, ou seja, que ao fim da simulação, o compressor ainda teria mais 1/3 de vida útil. Dessa forma, a diferença da degradação entre dois instantes se dá pela equação 31.

$$Dc_{T'} - Dc_T = Dc_{(T',\varepsilon_{cT})} * \varphi_c - Dc_{(T,\varepsilon_{cT})} * \varphi_c \quad (31)$$

Portanto, para j e j', o componente está sujeito a diferentes emissões de ε_{cT}, e essas emissões, multiplicadas pelo φc, definem o passo em que a degradação ocorre. Destarte, ambos os estados levam a falha completa, e a diferença reside na velocidade com que o sistema poderá alcançá-la. Para todos os casos, considera-se que a performance do compressor ηg_T é inversamente proporcional a Dg_T, isto é, à medida que Dg_T aumenta, ηg_T diminui, conforme equação 32.

$$\eta g_T = \frac{(100 - Dg_T)}{100} \quad (32)$$

4.5 AÇÕES

Nessa seção, os resultados possíveis das ações dadas pelo agente serão explicados. Como já apresentado, para cada c, duas intervenções são consideradas, a manutenção preventiva e a substituição. Então, a cada instante T, o agente decide se deverá desligar o equipamento e realizar uma intervenção ou seguir com a operação nas condições em que se encontra. A relação entre as ações e seus resultados está no Quadro 5, após, o processo será detalhado.

Quadro 5. A programação de manutenção para cada uma das ações.

Ação	Descrição	Comentário
Nada fazer	Nada é realizado	O sistema segue em operação independente das condições
Manutenção	O sistema é desligado O técnico performa atividades de manutenção	Há modificação na profundidade da degradação e no estado de Markov, mas o elemento reparado nunca

	preventiva no elemento indicado	recupera a performance inicial
Substituição	O sistema é desligado O técnico performa a substituição do elemento indicado	A performance inicial do elemento substituído é recuperada

Fonte: O autor (2022).

A quantificação das ações de reparo foi inspirada no método ARI_{∞},apresentado na seção 3.4, que propõe a redução na intensidade da falha como um modelo para expressar o efeito de uma determinada manutenção em um sistema simulado. De forma similar, esse trabalho propõe um fator de desconto fd aplicado à degradação ancorado no estado inicial de performance dos elementos e limitados pelo estado de falha completa.

Há também um segundo grau de liberdade que o agente deve controlar. Sendo o fd_c relacionado ao estado de degradação do componente que sofrerá intervenção, é razoável inferir que para componentes em performance muito próxima ao ideal, a intervenção preventiva, que possui intervalo de um ano, pode ser negligenciada ou ineficaz. Toda forma, o parâmetro Dc_T é minorado de acordo com as fórmulas no Quadro 6.

Quadro 6. Fatores de desconto e recuperação na performance por componente mediante ação.

Ação	Fator de desconto	Recuperação da performance
Nada fazer	$fd_T = 1$	Não há
Manutenção	$fd_T = \varpi_{Dc}$	Estocástico e condicionado a performance no momento T
Substituição	$fd_T = 0$	Completa

Fonte: O autor (2022).

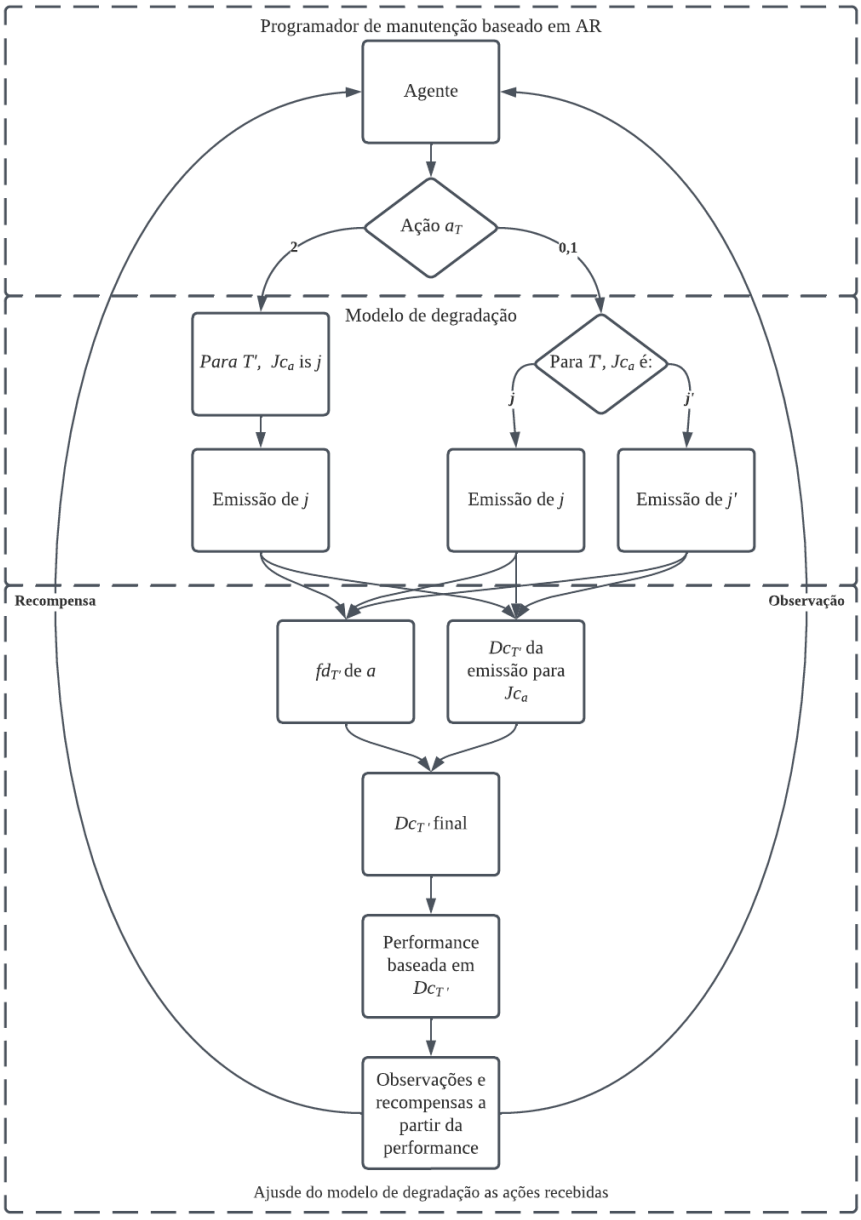
O ϖ_{Dc} é um parâmetro desconhecido e corresponde a um escalar obtido aleatoriamente a partir da performance ηc_T sendo $\varpi_{Dc} \in \{0, (1 - \eta c_T)\}$. Portanto, a relação entre a degradação do componente, sua performance e o fd_T pode ser apresentada pela equação 33.

$$Dc_{T'} = Dc_T * fd_T \quad (1)$$

Sendo Dc_T o nível de degradação no instante corrente e Dc_T, o nível de degradação no instante seguinte. Lembra-se também que a degradação só volta a ser contabilizada no reinício da operação do sistema, condicionada à ε_{cT} derivada de j_T após a intervenção.

Assim sendo, o processo de degradação e ajuste das ações pode ser apresentado por meio da Figura 15.

Figura 15. Processo de degradação e ajuste do ambiente.



Fonte: O autor (2022).

4.5.1 O aproximador de custos de intervenção

O aproximador de custos é fundamental para a formulação de problemas de decisão que utilizam a abordagem da a aprendizagem por reforço, devido principalmente, ao sinal de reforço enviado ao agente, que é o único elemento que indica a efetividade de suas ações (MAHMOODZADEH *et al.*, 2020). Destarte, são os objetivos do aproximador de custos:

1. Evitar a falha completa, que aqui corresponde a incapacidade de atingir a temperatura de descarregamento do compressor nas condições em que $f \neq 0,\ \phi < 0\ e\ dr_T = Falso$, e indica a perda da capacidade funcional do refrigerador;
2. Ampliar o período de funcionamento ideal e a disponibilidade do aparelho;
3. Reduzir os custos de manutenção.

Seguindo os princípios do Quadro 6, a partir de valores genéricos, quando $a = 0$, não são contabilizados custos referentes a intervenções, $\beta c_0 = 0$, pois é considerado que as manutenções, se não forem convocadas, não serão realizadas. Quando $a = 1$, o βc_1 é indexado ao Dc_T, sendo maior Dc_T, maior o βc_a, tal que $\beta c_a \in \{80, 250\}$. Já para $a = 2$, $\beta c_2 \in \{500, 700\}$. Para ambos os casos de intervenção, é adicionado o custo de acionamento do técnico definido como $\beta r_a \in \{80, 120\}$. Os valores de βc_a vem de valores típicos de manutenção obtidos no mercado no ano de 2022. O Quadro 7 sumariza a aplicação do β.

Quadro 7. Estimação do custo da intervenção para cada uma das ações.

Ação	Descrição	Comentário
Nada fazer	Nada é realizado	$Bc_{aT} = \beta c_0$
Manutenção	Ação preventiva pode ou não ser realizada	$Bc_{aT} = \frac{Dc_T}{100} * \left(max\beta c_1 - min\beta c_1\right) + min\beta c_1 + \beta r_a$
Substituição	Substituição é realizada	$Bc_{aT} = \beta c_2 + \beta r_a$

Fonte: O autor (2022).

Sendo Bc_{aT} o custo estimado da intervenção para um instante T mediante uma ação a.

4.5.2 O Aproximador de período sob intervenção

Após a intervenção, entende-se que o sistema é desligado, e a geração de dados e leituras é interrompida até que o reinício seja realizado por parte do técnico. O tempo em manutenção ou substituição também é contabilizado pelo modelo. Seguindo os princípios do Quadro 7, a partir de valores genéricos, quando $a = 0$, não há interrupção do funcionamento, portanto $\iota_0 = 0$, ou seja, 0 intervalos de cinco minutos. Quando $a = 1$, o ι_1 é indexado ao Dc_T, sendo maior Dc_T, maior o ι_1, tal que $\iota_1 \in \{6{,}18\}$. Já para $a = 2$, $\iota_2 \in \{18, 36\}$. O Quadro 8 sumariza a aplicação do ι.

Quadro 8. Estimação do custo da intervenção para cada uma das ações.

Ação	Descrição	Comentário
Nada fazer	Nada é realizado	$\mathrm{I}_{aT} = \iota_0$
Manutenção	Ação preventiva pode ou não ser realizada	$\mathrm{I}_{aT} = \frac{maxDc_T}{100} * (max\iota_1 - min\iota_1) + min\iota_1$
Substituição	Substituição é realizada	$\mathrm{I}_{aT} = \iota_2$

Fonte: O autor (2022).

Sendo I_{aT} o tempo em que o sistema estava indisponível sob intervenção, e $maxDc_T$ o equipamento mais degradado que sofreu intervenção em um determinado instante T.

4.6 RECOMPENSAS

Para calcular a recompensa de uma ação, a performance do sistema é levada em conta tanto para o instante passado quanto para o atual. É necessário que o sinal de recompensa ou reforço seja computado e enviado a cada *instante*. Os indicadores levados em conta para computá-lo são:

1. Se o refrigerador está ligado;
2. Se o compressor está ligado;
3. A diferença de temperatura entre os *instantes*;
4. A corrente;
5. A temperatura instantânea;
6. O estado j;
7. A ação.

A construção do R_T, o sinal de reforço ou *recompensa,* é um problema de escopo aberto, tanto no sentido das dimensionalidades, que nesse caso são sete, quanto na estrutura das funções, e para esse trabalho, não há referências diretas. Por isso, abaixo será detalhada a função que solucionou o problema da acoplagem da AR aos sinais do refrigerador no escopo da CBM.

4.6.1 Indicador de temperatura de descarregamento do compressor

É a recompensa por alcançar a temperatura de descarregamento do compressor de temperatura e é definido pela equação 34. Quando não há energia no sistema, $spi = 0$:

$$spi_T = \begin{cases} \dfrac{-\tau in_T - \tau out_T}{\tau min - \tau out_T}, & se\ f\ \neq 0 \\ 0, & se\ f = 0 \end{cases} \tag{34}$$

Em que spi é o indicador de temperatura de descarregamento do compressor, tin_T é a temperatura interna para o instante corrente, $tout_T$ é a temperatura externa para o instante corrente, e $tmin$ é a temperatura de temperatura de descarregamento do compressor. O spi recompensa o agente por manter a temperatura interna mais próxima à temperatura de descarregamento do compressor estabelecendo um índice de severidade para cada leitura. Esse índice corresponde a estrutura da própria função, que é uma equação da reta entre dois pontos, a saber, a temperatura inicial, correspondente a externa, e a temperatura alvo, ou a temperatura de descarregamento do compressor.

4.6.2 Indicador de consumo

É a recompensa ou penalização pelo tempo de funcionamento do compressor após alcançar a temperatura de descarregamento do compressor de temperatura, e é definido na equação 35. Algumas questões precisam ser explicadas aqui, a primeira é que o triplo condicional é definido em função da corrente, da tensão e da temperatura externa. Isso se dá por que quando o aparelho está desligado, a tensão é nula, mas quando está em operação e o compressor apenas está desativado, para $\phi > 0$, ela não é. Nesse caso, a corrente mínima repousa acima de $5 * 10^{-2}$. A terceira condição estabelece que quando o compressor está ligado e a temperatura interna atinge 90% da temperatura externa, uma folga para amortizar mudanças abruptas nas leituras de

τout_T entre T e T', um outro condicional é chamado, a saber ξ_T, que será explicado a seguir.

$$coi_T = \begin{cases} \begin{cases} coi = 0, & se\ \xi_T < 4 \\ coi = \vartheta_T, & se\ \xi_T > 4 \end{cases}, & se\ \tau in_T \leq \tau min_T * 90\% \text{ e } I \geq 0{,}05 \\ 0, & I \leq 0{,}05 \\ 0, & f = 0 \end{cases} \quad (35)$$

Em que coi é o indicador de consumo, ξ_T corresponde ao tempo de operação do compressor naquele instante, equação 36, e ϑ_T, equação 37, é um indicador de atraso no tempo esperado até alcançar o temperatura de descarregamento do compressor de temperatura. O condicional aplicado a ξ_T se justifica pela necessidade de neutralizar a inércia térmica que há no instante em que o refrigerador é iniciado, e a temperatura interna se encontra próxima da temperatura externa. Sem ele, coi penalizaria o R_T quando o refrigerador está em perfeito funcionamento, porém devido a inércia da carga térmica, demorou a refrigerar. A adição do ϑ_T adiciona uma segunda camada de tolerância ao coi, à medida que só se torna relevante para o cálculo de R_T quando começa a se repetir nos ciclos, o que significa que a anormalidade não foi capaz de ser resolvida pela operação do freezer. Ou seja, coi só é contabilizado pelo R_T quando ϑ_T repete-se nas condições já apresentadas na equação 35.

$$\xi_T = \begin{cases} \sum_{T}^{Tn} \mathrm{T}, & se\ \tau in_T \leq \tau min_T * 90\% \text{ e } I \geq 0{,}05 \\ 0, & se\ I \leq 0{,}05 \\ 0, & se\ f = 0 \end{cases} \quad (36)$$

$$\vartheta_T = \begin{cases} \sum_{T}^{Tn} \mathrm{T} * 0{,}5, & se\ \xi_T > 0 \\ 0, & se\ \xi_T = 0 \\ 0, & se\ \ f = 0 \end{cases} \quad (37)$$

4.6.3 Indicador de corrente

É a penalização pela corrente estar fora da condição de referência, e é definido pela equação 38:

$$Ii_T = \begin{cases} \begin{cases} (Ip - I_T) * 3{,}23, se\ Ir > i_T \\ (I_T - Ip) * 3{,}23, se\ Ir < i_T \end{cases}, & se\ I > 0{,}05 ef > 0 \\ 0, & se\ I \leq 0{,}05 \\ 0, & se\ f = 0 \end{cases} \quad (38)$$

Sendo Ii o indicador de corrente, Ir a corrente de referência para o compressor ligado em ampères, I_T a corrente no instante T em ampères, e f a tensão em volts. O valor de $3{,}23$ é um ajuste aplicado a diferença entre as correntes por causa do valor de R_T para $Ii = 0$. Aumentar ou diminuir seu valor significa alterar a intensidade com que o R_T é influenciado pelo Ii. Para o R_T usado nesse trabalho, obteve-se que esse valor conduz aos resultados desejados.

4.6.4 Indicador de diferencial de temperatura

É a bonificação ou penalização calculada a partir do diferencial de temperatura, é definida pela equação 39. Como as medições de τin_T são instáveis e variam a cada T, é necessário estabilizar o reforço para que o agente entenda que as mudanças, algumas até abruptas, nas observações são normais. Isso é feito pela junção da equação 39 com equação 34. Quando as duas são somadas, sendo elas equações da reta com início e fim iguais, porém de inclinações diferentes, formam um escalar estável para todo o espectro de temperatura de operação do freezer. E esse equilíbrio só é quebrado quando alguma das condicionais não é atendida, tanto na equação 39 quanto na equação 34.

$$dti_T = \begin{cases} \dfrac{(\tau in_{T-1} - \tau in_T) * \tau in_T + \tau out_T - \dfrac{min\tau - \tau out_T}{-1}}{\dfrac{min\tau - \tau out_T}{-1}}, & se\ \tau in_T \\ 0{,}5, \quad se \quad se\ \tau in_T > \tau min_T * 90\% \\ 0, \quad se\ f = 0 \\ \leq \tau min_T * 90\% \end{cases} \tag{39}$$

Em que dti é o indicador de diferencial de temperatura.

4.6.5 Indicador de transição de estado

É o indicador que bonifica o agente quando há transição $j(1 \rightarrow 0)$ após intervenção do agente durante o treinamento.

4.6.6 Indicador de falha

É o indicador que penaliza o agente quando há falha completa do sistema, e é definido pela equação 40.

$$fai_T = \begin{cases} penalty, & se\ f > 0\ e\ \tau in_T > \tau out_T * 90\%\ e\ ps > pd\ \ e\ dr_T = Falso \\ 0, & se\ f = 0 \end{cases} \quad (40)$$

Sendo fai o indicador de falha, ps, equação 41 um contador que pretende obter o tempo que seria esperado para o freezer escapar da condicional estabelecida na equação 40, e pd, equação 42 a quantidade de períodos em que $tint_T$ permanece acima de $tout_T$. O $penalty$ é um escalar predefinido, e serve como condição de encerramento do episódio de treinamento, o que significa falha do agente. Ou seja, o indicador de falha só é ativado quando o refrigerador atinge a temperatura externa estando ligado, o compressor acionado e a porta fechada.

$$pd_T = \begin{cases} \sum_{T}^{Tn} \mathrm{T}, & se\ \tau in_T > \tau min_T * 90\% \text{ e } I \geq 0{,}05 \\ 0, & se\ I < 0{,}05 \\ 0, & se\ \tau in_T \leq \tau min_T * 90\% \end{cases} \quad (41)$$

$$ps = \frac{\tau in_T - \tau max}{\frac{1}{2} * -1 * q\forall\phi < 0 - q\forall\phi > 0} \quad (42)$$

4.6.7 Indicador de ação

É o indicador que pretende evitar que o agente intervenha repetidamente no sistema. Uma vez que o agente tem o poder de causar o desligamento do refrigerador, se repetidamente ele o causa, ele é penalizado recebendo a subtração de um escalar. É definido pela equação 43. Esse indicador é necessário porque durante os experimentos, observou-se a tendencia do agente em convergir para a política não ideal de evitar penalidades mantendo o refrigerador desligado. Por isso, é necessário estabelecer limites para intervenções seguidas.

$$ap_T = \begin{cases} penalty, & se\ a_{T-1} = 1, & ea_T = 1 \\ penalty, & se\ a_{T-1} = 1, & ea_T = 2 \\ penalty, & se\ a_{T-1} = 2, & ea_T = 2 \\ penalty, & se\ a_{T-1} = 2, & ea_T = 1 \\ 0, & se\ a_{T-1} = 0, & ea_T = 2 \\ 0, & se\ a_{T-1} = 0, & ea_T = 1 \\ 0, & se\ a_{T-1} = 1, & ea_T = 0 \\ 0, & se\ a_{T-1} = 2, & ea_T = 0 \\ 0, & se\ a_{T-1} = 0, & ea_T = 0 \end{cases} \quad (43)$$

Em que ap_té a penalidade da ação no instante, ap_{t-1} é a ação no instante anterior, , a_té a ação no instante corrente, e $penalty$ é um escalar predefinido.

4.6.8 Função de recompensa

Ao final, a função de recompensa a ser maximizada pelo agente é definida pela equação 44.

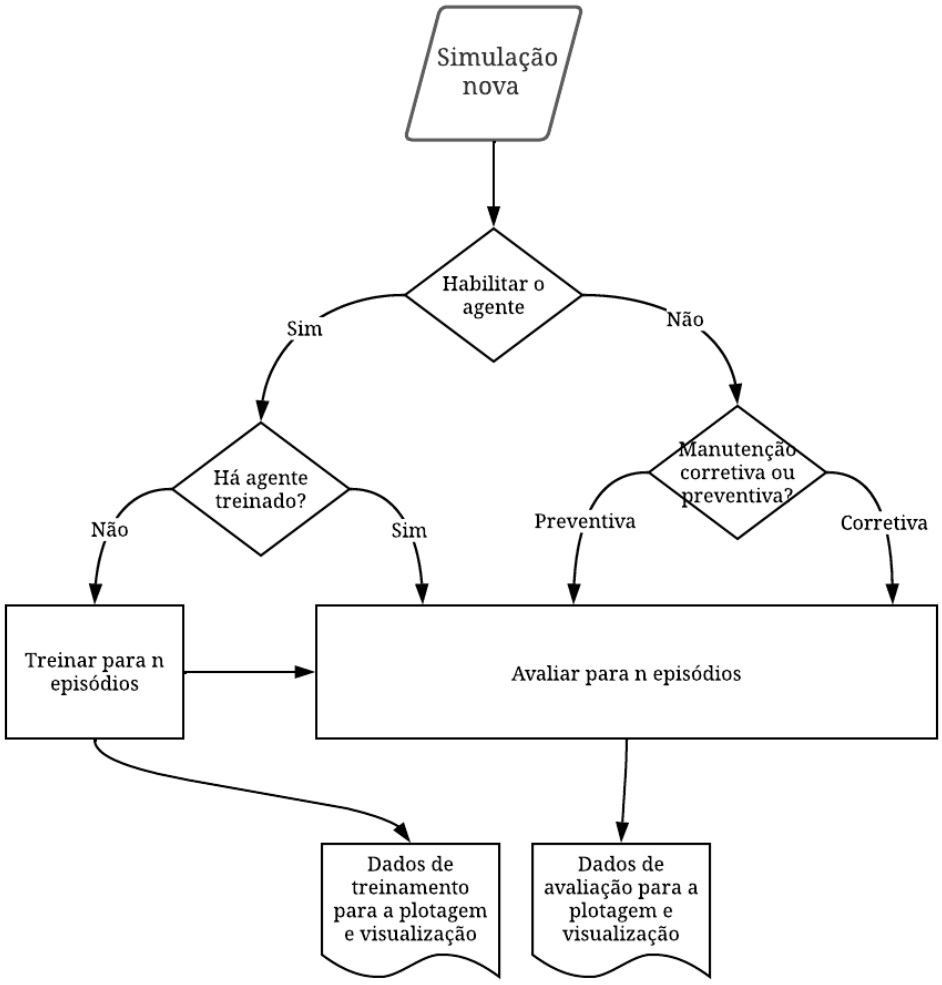
$$R_T = spi_T + coi_T + dti_T - li_T - fai_T - ps_T - ap_T - Bc_{aT} - \beta r_a \quad (44)$$

Em que R_T corresponde a função de recompensa.

5 METODOLOGIA

Nessa seção, será descrito como se avaliou a performance do agente de aprendizagem por reforço. Primeiro, uma série de parâmetros de funcionamento serão propostos, esses parâmetros serão a base de comparação. Em seguida, serão introduzidos dois programas de manutenção, o corretivo e o preventivo, periódico. Eles serão a base de comparação do agente. Após, o agente, hiper parametrizado é treinado até alcançar o limite de episódios pré-definido. E por fim, ele será avaliado, comparado com os outros programas de manutenção. Esse processo está disponível na Figura 16.

Figura 16. Método de geração de dados e avaliação do agente.



Fonte: O autor (2022).

5.1 TREINAMENTO E AVALIAÇÃO

Na etapa de treinamento, são definidos os parâmetros da rede neural do DDQN, os graus de liberdade do ambiente e uma janela de operação de três meses da vida útil do refrigerador. Essa janela deve conter uma transição $jc(0 \rightarrow 1)$, que é o problema que o agente deverá resolver. A solução desse problema representa o êxito do agente, o contrário, a falha. Nesse trabalho, o número de episódios $n_{treinamento} = 500$ foi o suficiente para obter resultados satisfatórios por parte do agente.

Na etapa de avaliação, a política estabelecida na etapa de treinamento será testada ao longo da vida útil do refrigerador, com todos os graus de liberdade disponíveis, e sem necessariamente a transição $jc(0 \rightarrow 1)$ acontecer, já que essa também será livremente definida pelo algoritmo. Aqui, os programas corretivo e preventivo estarão sujeitos aos mesmos parâmetros, e com isso, pretende-se avaliar a performance do agente frente as soluções concorrentes.

O programa de manutenção corretiva considera que qualquer outra forma de manutenção do sistema é negligenciada até que haja a quebra, isto é, algum componente falhe completamente e por extensão, só ocorre quando $fai_T > 0$. O programa preventivo intervenciona o sistema em intervalos de um ano, independente do estado, mas também permite que ações corretivas sejam realizadas. O comissionamento das ações em ambos os programas se dá por um gatilho adicionado ao algoritmo, sempre que a condição de acionamento da manutenção corretiva é alcançada ou o intervalo da manutenção preventiva encerra, uma ação é enviada usando a mesma interface e semântica usados pelo agente.

O agente tem um objetivo principal, maximizar as recompensas. Mas isso deve reverberar em outros três aspectos: evitar falhas completas, reduzir o consumo e reduzir os custos de manutenção. Por isso uma métrica de performance foi adicionada para cada um desses aspectos, para verificar a capacidade do agente atingir esses objetivos quantitativamente:

1. Se houve ou não falhas completas ao longo dos episódios;
2. O consumo/emissão sobre n episódios rodados;
3. O custo de manutenção sobre n episódios rodados;
4. O tempo sob manutenção sobre n episódio rodados.

Como métrica de performance usou-se a mediana dos resultados, os limites, os percentis, e o coeficiente de variação sobre n episódios rodados. Dessa maneira, permite-se observar o comportamento das aleatoriedades do modelo e eventuais erros do agente. Nesse trabalho, o número de episódios $n_{avaliações} = 50$ para todos os programas.

5.2 CONFIGURAÇÃO DOS EXPERIMENTOS

Os algoritmos de DRL crescem em demanda computacional no mesmo ritmo em que aumenta o tamanho das redes neurais que eles fazem uso. Hoje, é sabido que além disso, o tempo computacional dos experimentos de AR é muito dependente da estrutura do ambiente e da quantidade de iterações necessárias nos episódios de treinamento e avaliação. Logo na execução dos primeiros experimentos desse trabalho, ficou evidente a impossibilidade de realizá-los em um microcomputador de uso pessoal. Esse problema foi resolvido utilizando o servidor que pertence ao grupo de pesquisa do Dr. Enrique Lopez-Droguett da Universidade da California em Los Angeles - UCLA. Ele possui as seguintes configurações:

Quadro 9. Configurações do servidor em que os experimentos foram executados.

Sistema operacional	Ubuntu Server™
CPU	AMD Ryzen™ Threadripper™ PRO 3955WX
Número de núcleos	32
Memória	128 GB
GPU	2 x Nvidia Tesla™ M60
Armazenamento	1 Tb

Fonte: O autor (2022).

Os algoritmos de AR fazem bom uso de três características em específico, o processador voltado para aplicações profissionais, de alto *clock* e TDP, da memória, e da GPU, utilizada pelo Tensorflow™, a biblioteca base do Tensorforce™, para aceleração dos cálculos da rede neural do DRL. Para comparação, será apresentada a configuração do microcomputador local de apoio em que parte dos experimentos foi realizada:

Quadro 10. Configurações do microcomputador de apoio para comparação.

Sistema operacional	Windows 11™
CPU	AMD Ryzen™ 5700U
Número de núcleos	16
Memória	20 GB
GPU	AMD Radeon™ Vega 8
Armazenamento	1 Tb

Fonte: O autor (2022).

O servidor da UCLA foi acessado de maneira remota através do servidor do laboratório de cogeração da Universidade Federal de Pernambuco, UFPE. Nele foi estabelecida a conexão SSH através da qual o Pycharm™ pôde executar os experimentos remotamente.

6 RESULTADOS

Essa seção apresentará os resultados da etapa de treinamento e avaliação realizados. Na primeira etapa, serão discutidas as particularidades do treinamento, os hiper parâmetros usados e o que se pode obter de seus resultados. Em seguida, serão apresentadas as performances encontradas para os cenários sob manutenção corretiva e preventiva. Na terceira etapa, os resultados do agente serão incorporados e então se poderá discutir a qualidade do programa proposto por ele.

Para comparar o agente à base, ambos são testados em condições dispares do treinamento. Durante a avaliação, o agente não é treinado e não há atualização dos parâmetros da rede neural e nem ações randômicas. Nessa etapa, o algoritmo do agente trata de repetir a política que durante o treinamento obteve o melhor retorno no longo prazo, e por isso, não há exploração.

As recompensas recebidas pelo agente, assim como as emissões, consumo e custos de manutenção para toda a vida útil esperada do refrigerador serão comparadas. É relevante explorar toda a vida útil por causa dos objetivos do trabalho. Espera-se que o agente, de forma independente, seja capaz de manusear corretamente as demandas de um freezer, não por um período, mas em todo o tempo em que ele estiver comissionado. Isso enrobustece os resultados perante qualquer janela de avaliação.

6.1 O APRENDIZADO DO ALGORITMO DE AR

A arquitetura e os hiper parâmetros do agente usado nesse trabalho estão disponíveis na Tabela 4.

Tabela 4. Hiper parâmetros e arquiteturas.

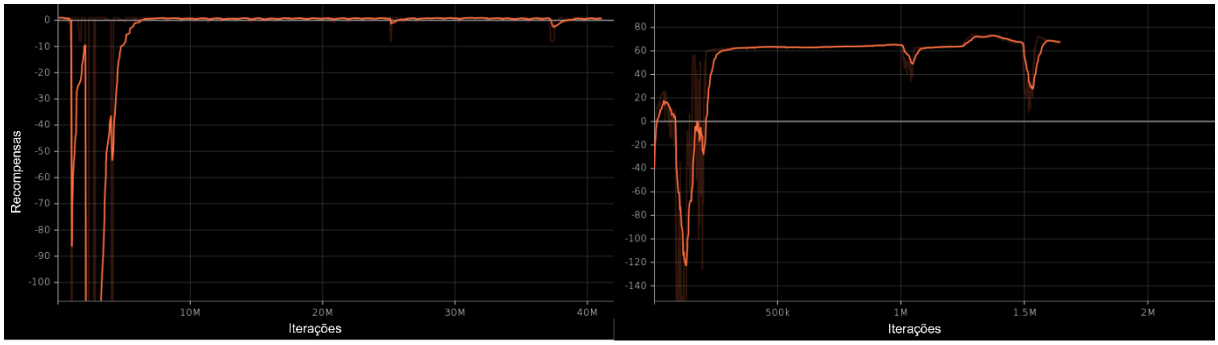
Frequência de atualização	0,25
Taxa de aprendizado	$1x10^{-4}$
Arquitetura da rede	Rede densa, camadas: Entrada: 10, sem ativação *Hidden*: 48,32,32,24; ativação ReLu

	Saída: 16, sem ativação
Tipo de agente	Double DQN
Tamanho do Batch	100
Fator de desconto	0,99

Fonte: O autor (2022).

Comparado ao microcomputador de apoio, o tempo de treinamento no servidor foi reduzido a cerca de ¼, mas ainda assim, ele custou aproximadamente 100 horas. Como o tempo computacional, mesmo com a ajuda das configurações apresentadas, é alto, sugere-se que o treinamento seja acompanhado da geração de logs ou relatórios de rastreio, que a cada momento retornam a performance do agente. A biblioteca Tensorforce™ é capaz de retornar essas informações, e esse foi um dos motivos da sua escolha. A impossibilidade da geração dos logs pode atrasar o processo de correção de erros e impossibilitar, devido ao alto consumo de tempo, a realização de pesquisas acadêmicas nessa área. Desses logs, se obteve os seguintes resultados:

Figura 17. Resultados do aprendizado DDQN. A esquerda, o gráfico por instante, a direita, o acumulado.



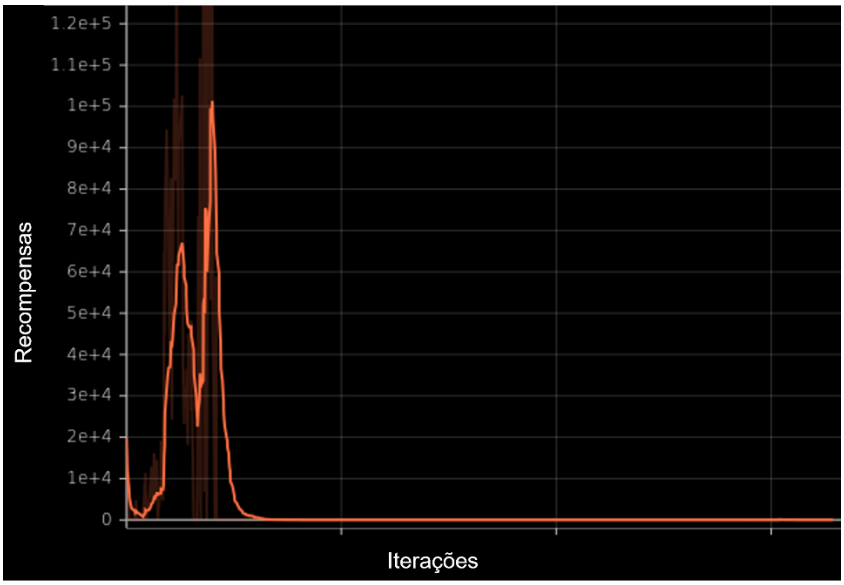
Fonte: O autor (2022).

A Figura 17 mostra as recompensas obtidas pelo DDQN durante o período de treinamento. Esse é o principal indicador de aprendizado quando se avalia algoritmos de aprendizagem por reforço. A esquerda da figura, visualizamos que o agente no início do processo, realiza o trade-off *exploration x exploitation*, explorando soluções possíveis para o problema, das quais grande parte, claro, sem sucesso. Ao encontrar uma solução capaz de maximizar as recompensas, ele passa a explotá-la, o que leva a subida e consequente estabilização das recompensas. Isso significa que houve

sucesso na formação da ‘política’. Como se trata do aprendizado, a exploração nunca será nula, e isso fica evidente quando na margem direita de ambas as figuras, encontramos dois pontos de deflexão provocados por tentativas frustradas de melhorar a ‘política’ vigente. Como a taxa de atualização da rede é de 0,25, as figuras da esquerda e direita possuem escalas diferentes.

A avaliação de quão boa uma ‘política’ é costuma ser evidenciada por meio da ‘função de perda’ na aprendizagem por reforço. Esse também é um bom indicador de aprendizado. A recompensa negativa é uma ‘perda’ que o algoritmo tentará otimizar através de políticas de gradiente descendente. Ou seja, quanto menor a perda, melhor os resultados. Quando há sucesso no aprendizado, o gráfico da perda tende a zero após o período de *exploration x exploitation*. No caso do DDQN, o gráfico de perda confirma o que está na Figura 17, mostrando sucesso na formação da política.

Figura 18. Perda do DDQN durante o aprendizado.



Fonte: O autor (2022).

6.2 ESTABELECENDO A BASE DE COMPARAÇÃO

Sob as mesmas condições de avaliação do agente, e em total grau de liberdade, foram rodados os programas corretivo e preventivo. Como já apresentado na seção 4.5, no programa corretivo, compreende-se que a manutenção do refrigerador é negligenciada até que por causa do grau de degradação do compressor, a temperatura interna se aproxime da temperatura externa, e o programa preventivo segue um cronograma de inspeções periódicas de custo atrelado a degradação no

intervalo entre as inspeções e efetividade calculada nos moldes do método apresentado na seção 4.4.2. A Tabela 5 apresenta os resultados na forma das métricas de performance.

Tabela 5. Resultados do programa corretivo e preventivo.

Dado	Modo de ação	Recompensas	Emissões kG de CO2	Consumo kWh	Custo de reparo $	*Downtime repair* (horas)	Falhas completas
Mediana	Corretivo	-49830680.8	3589.6	1608.45	0	1,93	Sim
	Preventivo	859071.815	3395.7	1520.44	1084,85	1,21	Sim
5-95 percentil	Corretivo	-50285953.39 | -49621689.2	3402.6 | 3583.6	1522.71 | 1608.57	0 – 2048,79	0 – 3,08	
	Preventivo	-50069662.24 | 953568.927	3360.6 | 3586.0	1503.15 | 1608.69	880 – 3036,94	0,69 – 2,13	
Limites	Corretivo	-50359618.5 | 876570.14	3370.5 | 3581.2	1509.75 - 1608,61	0 – 2381,56	0 – 3,08	
	Preventivo	-50197543.77 | 975153.19	3356.4 - 3588.8	1499.03 | 1608.74	880 – 3165,013	0,56 – 2,59	
Coeficiente de variação	Corretivo	-0.51	0.018	0.018	2,20	0,46	
	Preventivo	-2.59	0.022	0.022	0,46	0,36	

Fonte: O autor (2022).

Os resultados mostram que para ambos os programas, houve ao menos uma falha completa em um dos cinquenta episódios testados. Exceto pelo custo de manutenção, o programa preventivo apresenta melhores níveis de consumo e emissão, o que também se reflete nas recompensas. O programa preventivo é capaz de evitar que o sistema se deteriore a ponto de não ser capaz de atribuir recompensas positivas pelas métricas apresentadas em 4.6.

A recompensa negativa significa que o sistema esteve operando fora dos padrões de referência tempo o suficiente para que todo o ganho fosse revertido, sem, no entanto, apresentar falha completa. Essa situação põe em risco a conservação da mercadoria e reverbera negativamente no consumo e emissão. Observa-se também que em episódios do programa preventivo, houve falhas severas que ocorreram após a última verificação periódica, e por isso, não receberam intervenção. Para esses casos, também houve recompensa negativa.

Observa-se que para quase todos os casos, o programa corretivo apresenta maior coeficiente de variação, já que não há controle sobre a performance do refrigerador. Ao final de quinhentos episódios, há grande diversidade nas condições operacionais, o que também está de acordo com o apresentado nas seções 2.1 e 2.2, e argumenta favoravelmente quanto a construção do modelo.

A grande variabilidade fica mais evidente na mediana, percentis e limites do custo de reparo. Nos episódios em que não houve falha completa, seu custo foi 0, e o custo de reparo sobe na mesma proporção em que se repetem as falhas. Quanto ao programa preventivo, convém lembrar que cada inspeção periódica, de custo e tempo estabelecido pelos critérios da seção 4.5.1, infere em custos para o sistema. O que explica por que há, no fim das contas, mediana, percentis e limites maiores para esse programa.

6.3 AVALIAÇÃO DO PROGRAMA DE MANUTENÇÃO PROPOSTO

Primeiro, o agente de *Q-learn* foi treinado nas condições já apresentadas. Após, o desempenho dele foi testado sob as mesmas circunstâncias aplicadas aos programas corretivo e preventivo. Comparado ao microcomputador de apoio, o tempo de avaliação no servidor também foi reduzido a cerca de ¼, ainda assim, ele custou aproximadamente 18 horas. Ao final, os seguintes resultados foram encontrados, Tabela 6.

Tabela 6. Resultados do programa do agente.

Dado	Modo de ação	Recompensas	Emissões kG de CO2	Consumo kWh	Custo de reparo $	*Downtime repair* (horas)	Falhas completas
Mediana	Preditivo	904409,3	3404,5	13985,9	717,58	1,061	Não
5-95 percentil	Preditivo	879078,1 | 938541,2	3387,0 | 3416,5	13915,4 | 14033,3	219,4 | 2562,65	0,60 | 1,57	
Limites	Preditivo	831245,3 | 955269,7	3376,1 | 3448,1	13871,5 | 14160,1	219,4 | 3335,2	0,55 | 1,61	
Coeficiente de variação	Preditivo	0,02	0,03	0,003	0,78	0,28	

Fonte: O autor (2022).

Não houve falha completa em nenhum dos cinquenta episódios sob gestão do DDQN. Nota-se que para todos os indicadores, o agente apresenta melhores resultados. Esse resultado implica que há grande potencial em reduzir os custos de manutenção, as emissões e aumentar a confiabilidade quando se usa um agente de aprendizagem por reforço para gerir o sistema de manutenção.

A inexistência de recompensas negativas, assim como a ocorrência de valores mais altos, implica que o agente obteve sucesso na sua tarefa exposta na seção 2.3. E os demais valores argumentam a favor da função de recompensa exposta na seção 4.6.8. Isto é, o agente apresenta menor mediana nas emissões, no consumo e no tempo em reparo, enquanto, por outro lado, maior recompensa.

Também se observa que o agente entrega menor coeficiente de variação que os programas corretivo e preventivo. Isso se deve ao fato de o programa de manutenção do agente ser baseado na condição, não ser limitado por modelos pré-definidos munidos de limites, mas de estratégias construídas puramente do aprendizado. Assim, ele pode livremente intervir quando considera que aquele momento é o que trará a maior recompensa no longo prazo.

Todavia, o custo de manutenção destoa. De acordo com a construção do *test bench*, especificamente na seção 4.5.1, sempre que o agente convoca a intervenção,

a priori, o custo é computado. Logo, se o agente solicita duas vezes a manutenção em um curto período, como por exemplo $a_T, a_{T''} = \{1,0,1\}$, ainda que a ação seja executada apenas uma vez, ela será duplamente contabilizada, e o custo da segunda incorre como *pênalti* para o algoritmo. Sabendo disso, há duas considerações a se fazer:

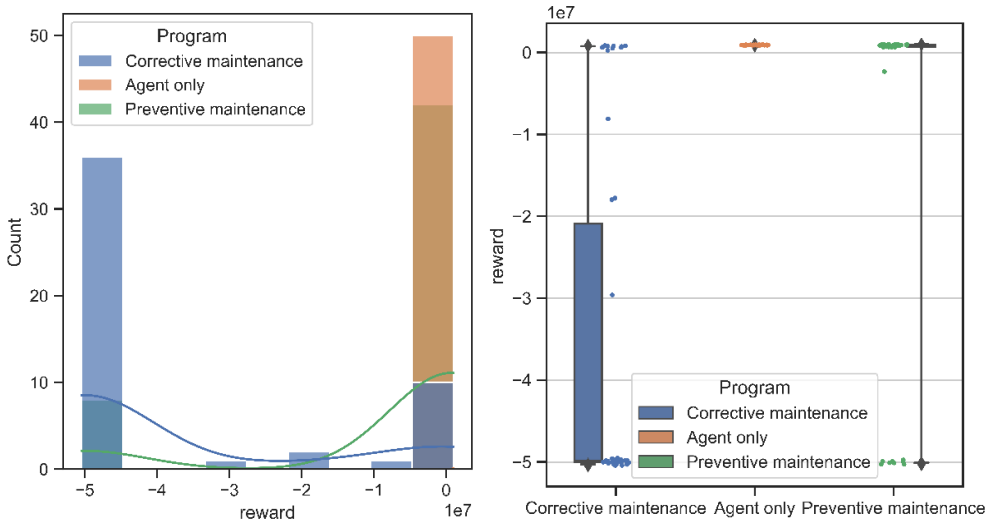
1. A primeira é quanto a instabilidade do agente, já apresentada no tópico 2.3, e que aqui em alguns episódios insiste em repetir ações, ainda que não haja resultado prático.
2. E em segundo lugar, se bloquear as intervenções seguintes adulteraria os resultados dos experimentos, ou até mesmo do aprendizado. Para todos os casos, preferiu-se não remover os erros do agente.

6.4 VISUALIZAÇÃO E ANÁLISE DE FORMA

Desde que as medianas são similares para os três programas, para melhor investigar a política de manutenção proposta pelo agente, os resultados das Tabela 5 e Tabela 6 serão plotados a seguir na forma de histograma com *kernel density estimation* (KDE) e gráfico de caixa. O objetivo do KDE é aumentar a legibilidade dos histogramas, mostrando a tendencia de distribuição quando há sobreposição de plotagem. Em todas as figuras, o eixo horizontal contabiliza o valor e o vertical a contagem de ocorrência nos cinquenta episódios. Sendo a cor azul o programa corretivo, verde o preventivo e dourado, o preditivo realizado pelo DDQN.

As recompensas, Figura 19, mostram o DDQN cumpriu seu papel de maximizar as recompensas com êxito, não permitindo a deterioração e por consequência a ocorrência de penalidades. Todos os cinquenta episódios encontram-se distribuídos entre os de melhor performance. Também é possível visualizar uma tendencia no comportamento que não ocorre nas outras soluções, uma vez que os resultados do DDQN, conforme mostrado pelo coeficiente de variação, encontram-se agrupados em uma estreita faixa. Isso mostra também que há um *tradeoff* quanto ao quanto se deve intervir no sistema para que os resultados sejam relevantes, e que o agente foi capaz de resolvê-lo de maneira mais elegante que a solução preventiva.

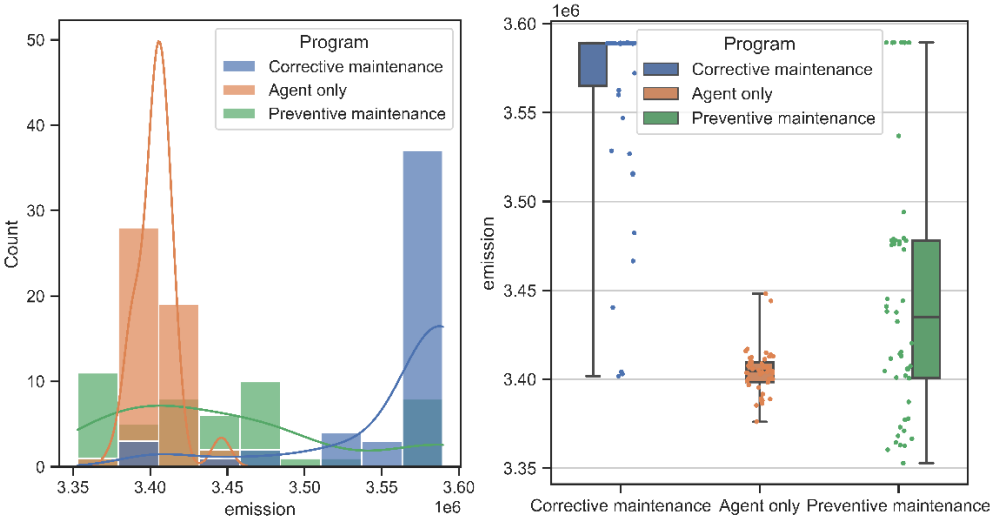
Figura 19. Histograma com KDE e box-plot das recompensas em cinquenta episódio do agente e dos programas corretivo e preventivo.



Fonte: O autor (2022).

Reiterando, esse êxito também se observa no fluxo de emissões, Figura 20. Enquanto para o programa corretivo há um agrupamento no que seria o máximo possível dentro das condições de operação naturais quebrado por valores atípicos, no preventivo há uma grande faixa de valores possíveis. Todavia, a manutenção baseada em condição realizada pelo agente provê previsibilidade maior nas emissões, já que é capaz de intervir no sistema para manter um padrão típico de performance de maneira mais efetiva.

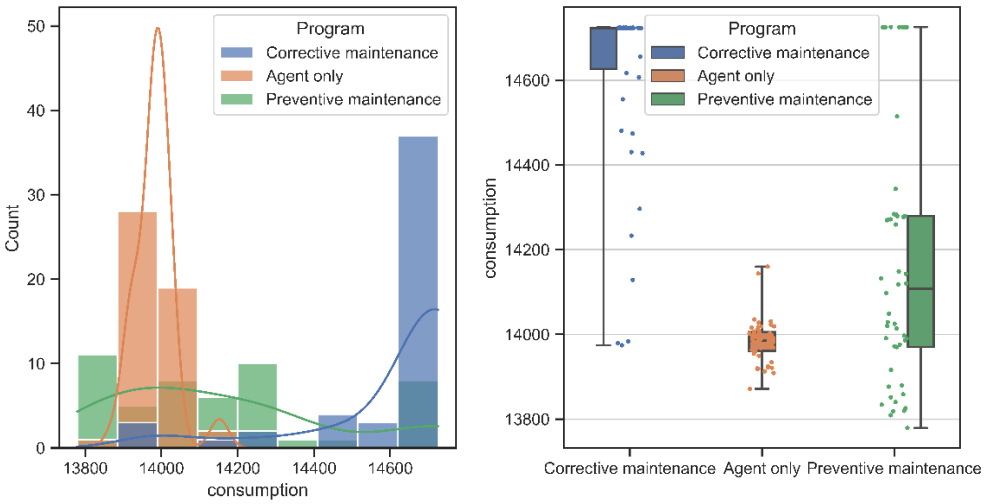
Figura 20. Histograma com KDE e box-plot das emissões em cinquenta episódio do agente e dos programas corretivo e preventivo.



Fonte: O autor (2022).

De forma análoga, o mesmo comportamento se repete no consumo energético, Figura 21.

Figura 21. Histograma com KDE e box-plot do consumo em cinquenta episódio do agente e dos programas corretivo e preventivo.



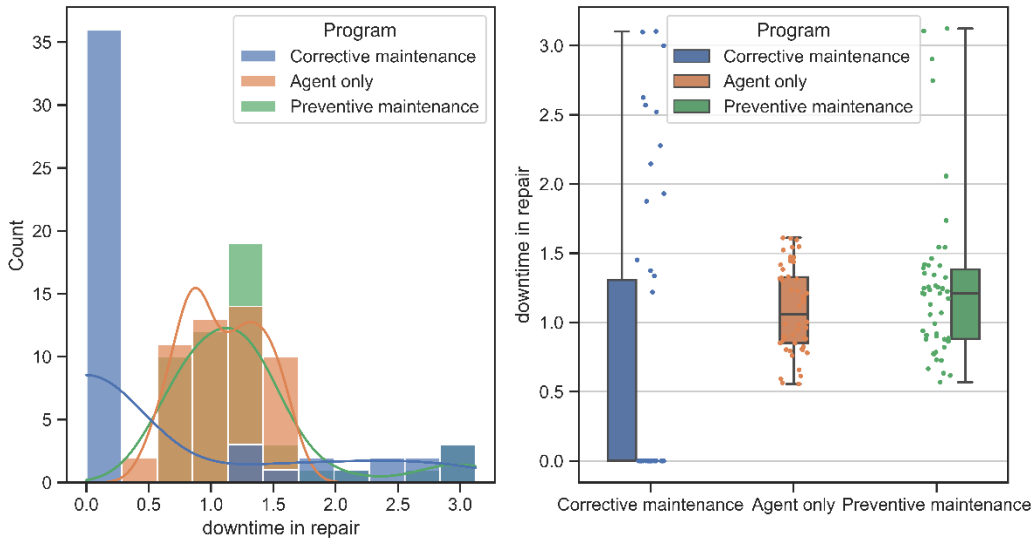
Fonte: O autor (2022).

No *tradeoff* que relaciona custo-recompensa, o tempo em manutenção também é afetado, já que o agente evita que o sistema chegue a níveis muito altos de degradação, o que implicaria em custos maiores e mais tempo parado. Similarmente,

vemos uma tendência que aumenta previsibilidade e reduz a ocorrência de situações atípicas, ou extremas.

Também se observa que do contrário do programa corretivo, o agente, mesmo quando não há transição de estado no ambiente, busca intervencionar o sistema em pelo ou menos uma vez ao longo da vida útil, para que seja recuperada parte da performance perdida pela degradação natural do sistema. Assim como não permite que quando o sistema está em falha, isto é $j = 1$, alcance níveis de degradação que exijam a substituição do compressor.

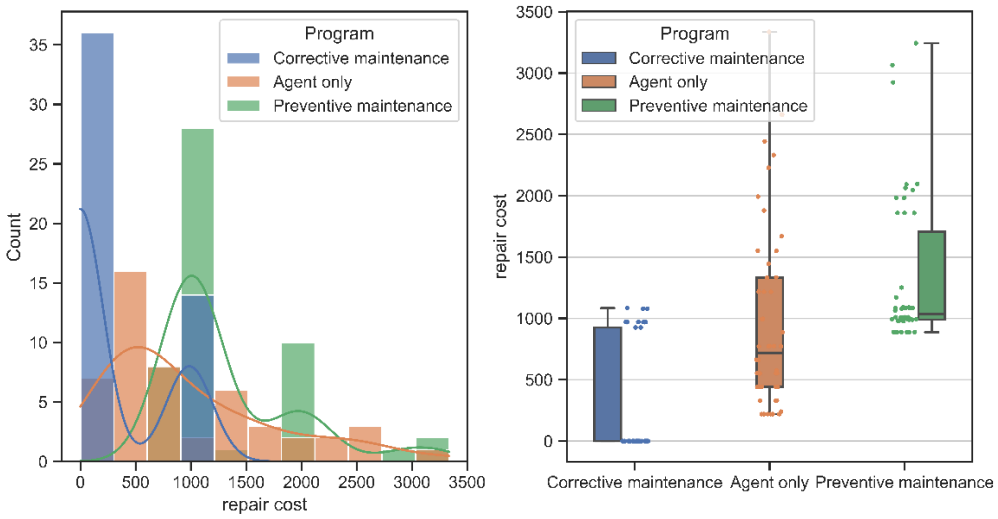
Figura 22. Histograma com KDE e box-plot do tempo de manutenção em cinquenta episódio do agente e dos programas corretivo e preventivo.



Fonte: O autor (2022).

Isso não se deve dizer, porém, do custo de manutenção pelos motivos já levantados na seção 6.3. Todavia, para 66% dos episódios, o custo de manutenção ainda ficou abaixo da mediana do programa preventivo.

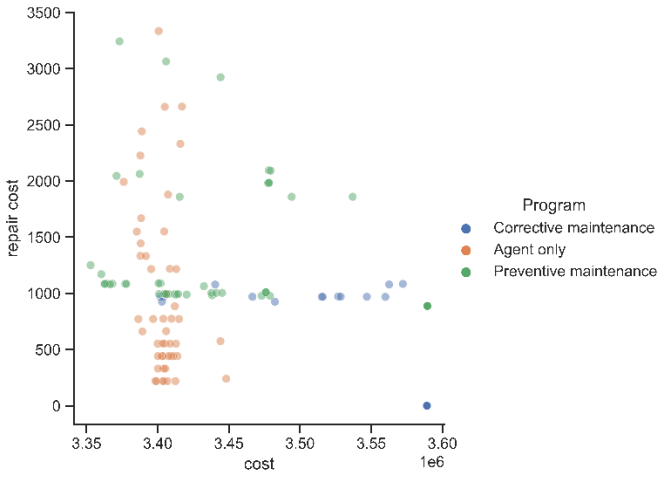
Figura 23. Histograma com KDE e box-plot do custo de manutenção em cinquenta episódio do agente e dos programas corretivo e preventivo.



Fonte: O autor (2022).

Realizando a plotagem do custo de reparo x custo de energia em moeda corrente, encontra-se a Figura 24. Nela, o obtido nas Figura 22 e Figura 23 está na forma de nuvem de pontos.

Figura 24. Custo de reparo x custo energético em moeda corrente.



Fonte: O autor (2022).

O agente consegue minimizar ambos os custos de maneira mais eficiente que o programa preventivo e o corretivo, que possuem dispersões muito maiores. Isso é positivo porque garante previsibilidade nos custos do ciclo de vida do aparelho, e contribui para a saúde financeira de negócios que operam com margens estreitas.

A formação da política de manutenção por parte do agente, ainda que pareça simples, é preciso lembrar, parte da ideia de que para os três programas de manutenção, o momento correto de intervenção é desconhecido até que se tenha o sistema em funcionamento. Além de que o conhecimento do agente acerca dos detalhes do modelo é nulo. E desse ponto de vista, considera-se relevante que o DDQN, através das observações selecionadas e da função de reforço desenhada, alcance performance iguais ou superiores aos outros programas do ponto de vista das emissões, do consumo, dos custos etc.

7 CONCLUSÕES

Esse trabalho apresentou um *test bench* e uma metodologia para otimizar o processo de manutenção de pequenos refrigeradores utilizando o framework da aprendizagem por reforço. Mesmo com os sucessivos avanços no campo da inteligência artificial e das aplicações de internet das coisas, a aplicação de aprendizagem por reforço ao campo da manutenção e de refrigeradores, por extensão ainda é incipiente e limitada.

O *test bench* proposto, principalmente nos componentes 4.4, 4.5 e 4.6, é um esforço para preencher essa falta apresentando uma base para o desenvolvimento de metodologias de aprendizagem por reforço aplicadas a refrigeradores, e as seções 4.3 e 4.4 são esforços para simplificando o problema, gerar um ponto de partida sem as barreiras e os riscos aplicáveis a integridade de um sistema real.

O *test bench* é baseado em um sistema real, é capaz de simular as leituras que seriam obtidas do sistema em intervalos de cinco minutos, e é capaz de interagir com o agente nas necessidades do paradigma a AR. Por consequência, ele provê uma plataforma capaz de entregar diferentes políticas de manutenção, não apenas limitadas à inteligência artificial, mas nomeadamente a corretiva e a preventiva. E com isso, investigar as possíveis consequências e dificuldades na operação do ciclo de vida e posterior transposição do *framework* a dispositivos reais.

O algoritmo proposto foi treinado e testado em um *test bench* e os resultados apresentaram diversas vantagens, como redução de 6% no consumo de energia e emissões em relação à manutenção corretiva, e redução de 33% nos custos de manutenção em relação à manutenção preventiva, além de melhorar disponibilidade do sistema, reduzindo o número e o tempo de paradas de manutenção. Não houve falha completa durante o gerenciamento do agente. Todos esses benefícios se refletem no aumento da performance do refrigerador e na previsibilidade de custos, relevantes para cargas perecíveis e comércios.

Embora o framework proposto seja uma abordagem promissora, e já esteja registrado no Instituto Nacional da Propriedade Industrial sob o título ‘REMRL - Programa de análise de falha e programação de manutenção em freezer baseado em

Reinforcement Learning', mais validação experimental é necessária para avaliar suas vantagens em um sistema real. Assim, na continuação desta pesquisa, os autores estão trabalhando no avanço do agente de AR para superar os resultados atuais em termos de desempenho e estabilidade para, então, aumentar o número de componentes degradantes e falhas simultâneas. Consequentemente, os autores estão trabalhando no desenvolvimento de uma solução AR confiável e eficiente que requer entradas de dados mínimas quando aplicada a um freezer real.

Há algumas implementações, refinamentos e formulações que são válidas de se explorar. Como contribuição a trabalhos futuros, o grau de sofisticação do algoritmo e dos resultados pode ser aumentado da seguinte maneira:

1. Além do compressor, habilitar a degradação dos demais componentes. Com isso, se pode observar a habilidade do agente de manusear diferentes problemas ocorrendo ao mesmo tempo;
2. Nesse trabalho, as ações foram discriminadas como $(0,1,2)$, o que não é um problema já que apenas o compressor está sendo observado. Quando se tem mais de um componente degradando, em processos paralelos e interdependentes, é valido discriminar as ações em vetores em que o número de elementos do vetor é o número de elementos reparáveis. Assim, o agente poderá decidir que componente intervencionar a cada momento;
3. A modelagem do freezer é simples, usa linearizações e limita o número de características observáveis. Questiona-se quais características também poderiam ser relevantes para a formação da política. Para isso, sugere-se aumentar a complexidade do modelo fazendo uso de bibliotecas como Coolprop™ e PyroMat™ por exemplo, para obter valores intermediários dos processos termodinâmicos, estimar medidas invasivas e aumentar a robustez dos resultados;
4. A quantidade de episódios de avaliação foi limitada pelo tempo e recursos computacionais disponíveis. Sugere-se investir na paralelização do treinamento e avaliação do agente para fazer melhor uso dos recursos computacionais e aumentar a quantidade de resultados possíveis. Isso requer alterações na estrutura do ambiente, mas deve possibilitar a exploração de outros algoritmos de

aprendizado, que aqui não foram explorados em virtude do tempo disponível.

Obviamente, as implementações acima aumentam a complexidade e a carga computacional do modelo, mas podem gerar resultados relevantes para a exploração das capacidades do algoritmo de aprendizagem por reforço aplicados a manutenção de refrigeradores por compressão mecânica.

REFERÊNCIAS

PAVLOV, I. Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex. **Annals of Neurosciences**, 1 jun. 2010. v. 17, n. 3, p. 136. Disponível em: </pmc/articles/PMC4116985/>. Acesso em: 2 jan. 2023.

ADSULE, A.; KULKARNI, M.; TEWARI, A. Reinforcement learning for optimal policy learning in condition-based maintenance. **IET Collaborative Intelligent Manufacturing**, 1 dez. 2020. v. 2, n. 4, p. 182–188. . Acesso em: 8 jan. 2023.

AHMAD, R.; KAMARUDDIN, S. An overview of time-based and condition-based maintenance in industrial application. **Computers & Industrial Engineering**, 1 ago. 2012. v. 63, n. 1, p. 135–149. . Acesso em: 22 set. 2022.

AMARI, S. V; MCLAUGHLIN, L.; PHAM, H. **Cost-Effective Condition-Based Maintenance Using Markov Decision Processes**. [S.l.]: [s.n.], 2006.

ANDRIOTIS, C. P.; PAPAKONSTANTINOU, K. G. Managing engineering systems with large state and action spaces through deep reinforcement learning. **Reliability Engineering and System Safety**, 1 nov. 2019. v. 191. . Acesso em: 8 jan. 2023.

_____; _____. Deep reinforcement learning driven inspection and maintenance planning under incomplete information and constraints. **Reliability Engineering and System Safety**, 1 ago. 2021. v. 212. . Acesso em: 8 jan. 2023.

ASSAWAMARTBUNLUE, K.; BRANDEMUEHL, M. J. Refrigerant leakage detection and diagnosis for a distributed refrigeration system. **HVAC and R Research**, 2006. v. 12, n. 3, p. 389–405. . Acesso em: 22 set. 2022.

BANSAL, T. *et al.* AI based Diagnostic Service for IOT enabled Smart Refrigerators. **Proceedings - 2021 International Conference on Future Internet of Things and Cloud, FiCloud 2021**, 1 ago. 2021. p. 163–168. . Acesso em: 5 jan. 2023.

BARDE, S. R. A.; YACOUT, S.; SHIN, H. Optimal preventive maintenance policy based on reinforcement learning of a fleet of military trucks. **Journal of Intelligent Manufacturing**, 31 jan. 2019. v. 30, n. 1, p. 147–161.

BARRETT, E.; LINDER, S. Autonomous hvac control, a reinforcement learning approach. [S.l.]: Springer Verlag, 2015. V. 9286, p. 3–19.

BEGHI, A. *et al.* Data-driven Fault Detection and Diagnosis for HVAC water chillers. **Control Engineering Practice**, 1 ago. 2016. v. 53, p. 79–91.

BEGHI, Alessandro; RAMPAZZO, Mirco; ZORZI, S. Reinforcement Learning Control of Transcritical Carbon Dioxide Supermarket Refrigeration Systems. **IFAC-PapersOnLine**, 1 jul. 2017. v. 50, n. 1, p. 13754–13759. . Acesso em: 31 jul. 2022.

BEHFAR, A.; YUILL, D.; YU, Y. Automated fault detection and diagnosis methods for supermarket equipment (RP-1615). **Science and Technology for the Built Environment**, 17 nov. 2017. v. 23, n. 8, p. 1253–1266.

BONVINI, M. *et al.* Robust on-line fault detection diagnosis for HVAC components based on nonlinear state estimation techniques. **Applied Energy**, 1 jul. 2014. v. 124, p. 156–166.

CAPOZZOLI, A.; LAURO, F.; KHAN, I. Fault detection analysis using data mining techniques for a cluster of smart office buildings. **Expert Systems with Applications**, 1 jun. 2015. v. 42, n. 9, p. 4324–4338. . Acesso em: 6 jan. 2023.

CHANG, J. *et al.* Intelligent Prediction of Refrigerant Amounts Based on Internet of Things. **Complexity**, 2020a. v. v 2020. Disponível em: <https://www.engineeringvillage.com/share/document.url?mid=cpx_77e69a741709cce67c8M791310178163146&database=cpx&view=detailed>. Acesso em: 29 jul. 2022.

_____ *et al.* Intelligent Prediction of Refrigerant Amounts Based on Internet of Things. **Complexity**, 2020b. v. 2020. . Acesso em: 5 jan. 2023.

CHEN, Jing; CHEN, Jia; ZHANG, Hongke. DRL-QOR: Deep Reinforcement Learning-Based QoS/QoE-Aware Adaptive Online Orchestration in NFV-Enabled Networks. **IEEE Transactions on Network and Service Management**, 1 jun. 2021. v. 18, n. 2, p. 1758–1774. . Acesso em: 8 jan. 2023.

CHENG, J. *et al.* Optimum condition-based maintenance policy with dynamic inspections based on reinforcement learning. **Ocean Engineering**, 1 out. 2022. v. 261. . Acesso em: 8 jan. 2023.

CHENG, M.; FRANGOPOL, D. M. A Decision-Making Framework for Load Rating Planning of Aging Bridges Using Deep Reinforcement Learning. **Journal of Computing in Civil Engineering**, nov. 2021. v. 35, n. 6. . Acesso em: 8 jan. 2023.

CORREA-JULLIAN, C.; LÓPEZ DROGUETT, E.; CARDEMIL, J. M. Operation scheduling in a solar thermal system: A reinforcement learning-based framework. **Applied Energy**, 15 jun. 2020. v. 268.

COTRUFO, N.; ZMEUREANU, R. PCA-based method of soft fault detection and identification for the ongoing commissioning of chillers. **Energy and Buildings**, 15 out. 2016. v. 130, p. 443–452. . Acesso em: 6 jan. 2023.

CRACIUN, I.; LECOQ, F.; DIGAVALLI, S. Condition Based Maintenance for Oil and Gas Industry Based on Data Reconciliation Techniques. **Society of Petroleum Engineers - Abu Dhabi International Petroleum Exhibition and Conference 2019, ADIP 2019**, 11 nov. 2019. Disponível em: </SPEADIP/proceedings-abstract/19ADIP/3-19ADIP/217033>. Acesso em: 8 jan. 2023.

CUI, P. *et al.* Predictive maintenance decision-making for serial production lines based on deep reinforcement learning. **Jisuanji Jicheng Zhizao Xitong/Computer Integrated Manufacturing Systems, CIMS**, 1 dez. 2021. v. 27, n. 12, p. 3416–3428. . Acesso em: 8 jan. 2023.

DAI, S. *et al.* Fault Diagnosis of Data-Driven Photovoltaic Power Generation System Based on Deep Reinforcement Learning. **Mathematical Problems in Engineering**, 2021. v. 2021. . Acesso em: 8 jan. 2023.

DAVENPORT, M. L.; QI, D.; ROE, B. E. Food-related routines, product characteristics, and household food waste in the United States: A refrigerator-based pilot study. **Resources, Conservation and Recycling**, 1 nov. 2019. v. 150, p. 104440. . Acesso em: 14 set. 2022.

DEY, M.; RANA, S. P.; DUDLEY, S. Semi-supervised learning techniques for automated fault detection and diagnosis of HVAC systems. [S.l.]: IEEE Computer Society, 2018. V. 2018-November, p. 872–877.

_____; _____; _____. Smart building creation in large scale HVAC environments through automated fault detection and diagnosis. **Future Generation Computer Systems**, 1 jul. 2020. v. 108, p. 950–966. . Acesso em: 7 jan. 2023.

DING, T. L.; SUBIANTORO, A.; NORRIS, S. Reinforcement Learning Control for Vapor Compression Reinforcement Learning Control for Vapor Compression Refrigeration Cycle. 2021. Disponível em: <https://docs.lib.purdue.edu/iracc>. Acesso em: 31 jul. 2022.

DONG, Z. *et al.* Artificial Intelligence Enabled Smart Refrigeration Management System Using Internet of Things Framework. **ACM International Conference Proceeding Series**, 24 abr. 2020. p. 65–70. . Acesso em: 27 jul. 2022.

DOYEN, L.; GAUDOIN, O. Classes of imperfect repair models based on reduction of failure intensity or virtual age. **Reliability Engineering & System Safety**, 1 abr. 2004. v. 84, n. 1, p. 45–56. . Acesso em: 8 set. 2022.

DU, A.; GHAVIDEL, A. Parameterized deep reinforcement learning-enabled maintenance decision-support and life-cycle risk assessment for highway bridge portfolios. **Structural Safety**, 1 jul. 2022. v. 97. . Acesso em: 8 jan. 2023.

DU, Z. *et al.* Fault detection and diagnosis for buildings and HVAC systems using combined neural networks and subtractive clustering analysis. **Building and Environment**, mar. 2014. v. 73, p. 1–11.

DUMONT, O.; QUOILIN, S.; LEMORT, V. Importance of the reconciliation method to handle experimental data in refrigeration and power cycle: application to a reversible heat pump/organic Rankine cycle unit integrated in a positive energy building. **International Journal of Energy and Environmental Engineering**, 1 jun. 2016. v. 7, n. 2, p. 137–143. . Acesso em: 8 jan. 2023.

EUROSTAT. Development of electricity prices for household consumers, EU, 2008-2021 (EUR per kWh) v2.png - Statistics Explained. [s.d.]. Disponível em: <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Development_of_electricity_prices_for_household_consumers,_EU,_2008-2021_(EUR_per_kWh)_v2.png>. Acesso em: 22 ago. 2022.

FAN, C.; XIAO, F.; ZHAO, Yang. A short-term building cooling load prediction method using deep learning algorithms. **Applied Energy**, 1 jun. 2017. v. 195, p. 222–233. . Acesso em: 6 jan. 2023.

FÉLIX JÚNIOR, W. S. **Aplicação de aprendizado por reforço em navegação de robôs**. Belo Horizonte: Universidade Federal de Minas Gerais, 2022. Essay.

FENG, M.; LI, Y. Predictive Maintenance Decision Making Based on Reinforcement Learning in Multistage Production Systems. **IEEE Access**, 2022. v. 10, p. 18910–18921. . Acesso em: 8 jan. 2023.

FMI. FMI | Grocery Store Chains Net Profit. 2022. Disponível em: <https://www.fmi.org/our-research/supermarket-facts/grocery-store-chains-net-profit>. Acesso em: 22 set. 2022.

FRANZESE, M.; IULIANO, A. Hidden Markov Models. **Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics**, 1 jan. 2019. v. 1–3, p. 753–762. . Acesso em: 5 out. 2022.

FRICON. Manual de instruções - 07.0501REV.Z (29/11/2021). **Fricon**, 29 nov. 2021. Disponível em: <https://friconadm-my.sharepoint.com/personal/patricia_ribeiro_fricon_com_br/_layouts/15/onedrive.aspx?id=%2Fpersonal%2Fpatricia%5Fribeiro%5Ffricon%5Fcom%5Fbr%2FDocuments%2FManuais%20de%20produtos%2F07%2E0501%2D%20MANUAL%20DE%20INSTRU%C3%87%C3%95ES%20HORIZONTAL%5FRev%2EZ%2Epdf&parent=%2Fpersonal%2Fpatricia%5Fribeiro%5Ffricon%5Fcom%5Fbr%2FDocuments%2FManuais%20de%20produtos&ga=1>. Acesso em: 6 set. 2022.

GAO, P. *et al.* Dynamic scheduling method of distributed photovoltaic operation and maintenance resources based on reinforcement learning. **Jisuanji Jicheng Zhizao Xitong/Computer Integrated Manufacturing Systems, CIMS**, 1 fev. 2022. v. 28, n. 2, p. 552–563. . Acesso em: 8 jan. 2023.

GINER, J. *et al.* Demonstrating Reinforcement Learning for Maintenance Scheduling in a Production Environment. **IEEE International Conference on Emerging Technologies and Factory Automation, ETFA**, 2021. v. 2021-September. . Acesso em: 16 nov. 2022.

GYM TEAM. Gym Documentation. 2022. Disponível em: <https://www.gymlibrary.dev/#>. Acesso em: 5 out. 2022.

HAN, H. *et al.* Study on a hybrid SVM model for chiller FDD applications. **Applied Thermal Engineering**, 1 mar. 2011. v. 31, n. 4, p. 582–592. . Acesso em: 6 jan. 2023.

HASSELT, H. VAN; GUEZ, A.; SILVER, D. Deep Reinforcement Learning with Double Q-learning. **30th AAAI Conference on Artificial Intelligence, AAAI 2016**, 22 set. 2015. p. 2094–2100. Disponível em: <https://arxiv.org/abs/1509.06461v3>. Acesso em: 31 jul. 2022.

HE, S. *et al.* Fault detection and diagnosis of chiller using Bayesian network classifier with probabilistic boundary. **Applied Thermal Engineering**, 25 ago. 2016. v. 107, p. 37–47. . Acesso em: 6 jan. 2023.

HONG, T. *et al.* State-of-the-art on research and applications of machine learning in the building life cycle. **Energy and Buildings**, 2020a. v. 212, n. April.

_____ *et al.* **State-of-the-art on research and applications of machine learning in the building life cycle**. **Energy and Buildings**. Elsevier Ltd.

HU, J. *et al.* Optimal maintenance scheduling under uncertainties using Linear Programming-enhanced Reinforcement Learning. **Engineering Applications of Artificial Intelligence**, 1 mar. 2022a. v. 109, p. 104655. . Acesso em: 16 nov. 2022.

_____ *et al.* Optimal maintenance scheduling under uncertainties using Linear Programming-enhanced Reinforcement Learning. **Engineering Applications of Artificial Intelligence**, 1 mar. 2022b. v. 109. . Acesso em: 8 jan. 2023.

IEA. Average CO2 emissions intensity of hourly electricity supply in the European Union, 2018 and 2040 by scenario and average electricity demand in 2018 – Charts – Data & Statistics - IEA. [s.d.]. Disponível em: <https://www.iea.org/data-and-statistics/charts/average-co2-emissions-intensity-of-hourly-electricity-supply-in-the-european-union-2018-and-2040-by-scenario-and-average-electricity-demand-in-2018>. Acesso em: 22 ago. 2022.

JANG, D. *et al.* Offline-online reinforcement learning for energy pricing in office demand response: Lowering energy and data costs. **BuildSys 2021 - Proceedings of the 2021 ACM International Conference on Systems for Energy-Efficient Built Environments**, 17 nov. 2021. p. 131–139. Disponível em: <https://doi.org/10.1145/3486611.3486668>. Acesso em: 30 out. 2022.

KATIPAMULA, S.; BRAMBLEY, M. R. Review article: Methods for fault detection, diagnostics, and prognostics for building systems—A review, part I. **HVAC and R Research**, 2005. v. 11, n. 1, p. 3–25. . Acesso em: 22 set. 2022.

KIM, S. *et al.* Vision-Based Deep Q-Learning Network Models to Predict Particulate Matter Concentration Levels Using Temporal Digital Image Data. 2019. Disponível em: <https://doi.org/10.1155/2019/9673047>. Acesso em: 7 out. 2022.

KIM, W.; KATIPAMULA, S. A review of fault detection and diagnostics methods for building systems. **Science and Technology for the Built Environment**, 2018a. v. 24, n. 1, p. 3–21.

_____; _____. A review of fault detection and diagnostics methods for building systems. **Science and Technology for the Built Environment**, 2 jan. 2018b. v. 24, n. 1, p. 3–21.

KNOWLES, M.; BAGLEE, D. The role of maintenance in energy saving in commercial refrigeration. **Journal of Quality in Maintenance Engineering**, 2012. v. 18, n. 3, p. 282–294. . Acesso em: 29 ago. 2022.

_____; _____; WERMTER, S. Reinforcement learning for scheduling of maintenance. [S.l.]: Springer London, 2011. p. 409–422.

KONGKIJPIPAT, P. *et al.* Wet Gas Pipeline Maintenance Process Using Reinforcement Learning. **2022 19th International Joint Conference on Computer Science and Software Engineering, JCSSE 2022**, 2022. . Acesso em: 16 nov. 2022.

KOPRINKOVA-HRISTOVA, P. Reinforcement Learning for Predictive Maintenance of Industrial Plants. **Information Technologies and Control**, 23 jun. 2014. v. 11, n. 1, p. 21–28.

KUŽNAR, D. *et al.* An intelligent system to monitor refrigeration devices. **Expert Systems**, 1 out. 2017. v. 34, n. 5.

KUZNETSOVA, E. *et al.* Reinforcement learning for microgrid energy management. **Energy**, 15 set. 2013. v. 59, p. 133–146.

LAMPRECHT, R.; WURST, F.; HUBER, M. F. Reinforcement Learning based Condition-oriented Maintenance Scheduling for Flow Line Systems. **IEEE**

International Conference on Industrial Informatics (INDIN), 27 ago. 2021. v. 2021-July. Disponível em: <https://arxiv.org/abs/2108.12298v1>. Acesso em: 16 nov. 2022.

LEE, D.; CHEN, M.-H.; LAI, G.-W. Achieving energy savings through artificial-intelligence-assisted fault detection and diagnosis: Case study on refrigeration systems. **Case Studies in Thermal Engineering**, 1 dez. 2022. v. 40, p. 102499. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2214157X22007365>. Acesso em: 13 nov. 2022.

LEI, Y. *et al.* **Applications of machine learning to machine fault diagnosis: A review and roadmap**. **Mechanical Systems and Signal Processing**. Academic Press.

LI, D. *et al.* Fault detection and diagnosis for building cooling system with a tree-structured learning method. **Energy and Buildings**, 1 set. 2016. v. 127, p. 540–551. . Acesso em: 6 jan. 2023.

_____; HU, G.; SPANOS, C. J. A data-driven strategy for detection and diagnosis of building chiller faults using linear discriminant analysis. **Energy and Buildings**, 15 set. 2016. v. 128, p. 519–529. . Acesso em: 6 jan. 2023.

LI, G. *et al.* An improved fault detection method for incipient centrifugal chiller faults using the PCA-R-SVDD algorithm. **Energy and Buildings**, 15 mar. 2016. v. 116, p. 104–113. . Acesso em: 6 jan. 2023.

_____ *et al.* Identification and isolation of outdoor fouling faults using only built-in sensors in variable refrigerant flow system: A data mining approach. **Energy and Buildings**, 1 jul. 2017. v. 146, p. 257–270. . Acesso em: 6 jan. 2023.

_____ *et al.* An improved decision tree-based fault diagnosis method for practical variable refrigerant flow system using virtual sensor-based fault indicators. **Applied Thermal Engineering**, 25 jan. 2018. v. 129, p. 1292–1303. . Acesso em: 6 jan. 2023.

_____; HU, Y. Improved sensor fault detection, diagnosis and estimation for screw chillers using density-based clustering and principal component analysis. **Energy and Buildings**, 15 ago. 2018. v. 173, p. 502–515. . Acesso em: 6 jan. 2023.

LI, S.; WEN, J. A model-based fault detection and diagnostic methodology based on PCA method and wavelet transform. **Energy and Buildings**, 1 jan. 2014. v. 68, n. PARTA, p. 63–71. . Acesso em: 6 jan. 2023.

LI, Z.; ZHONG, S.; LIN, L. An aero-engine life-cycle maintenance policy optimization algorithm: Reinforcement learning approach. **Chinese Journal of Aeronautics**, 1 set. 2019. v. 32, n. 9, p. 2133–2150. . Acesso em: 8 jan. 2023.

LIU, Jiahui *et al.* Abnormal energy identification of variable refrigerant flow air-conditioning systems based on data mining techniques. **Applied Thermal Engineering**, 5 mar. 2019a. v. 150, p. 398–411. . Acesso em: 6 jan. 2023.

_____ *et al.* Abnormal energy identification of variable refrigerant flow air-conditioning systems based on data mining techniques. **Applied Thermal Engineering**, 5 mar. 2019b. v. 150, p. 398–411. . Acesso em: 6 jan. 2023.

LIU, Jiangyan *et al.* Energy diagnosis of variable refrigerant flow (VRF) systems: Data mining technique and statistical quality control approach. **Energy and Buildings**, 15 set. 2018. v. 175, p. 148–162. . Acesso em: 6 jan. 2023.

LIU, Tao *et al.* A novel deep reinforcement learning based methodology for short-term HVAC system energy consumption prediction. **International Journal of Refrigeration**, 1 nov. 2019a. v. 107, p. 39–51. . Acesso em: 31 jul. 2022.

_____ *et al.* A novel deep reinforcement learning based methodology for short-term HVAC system energy consumption prediction. **International Journal of Refrigeration**, 1 nov. 2019b. v. 107, p. 39–51.

LIU, Yu; CHEN, Yiming; JIANG, T. Dynamic selective maintenance optimization for multi-state systems over a finite horizon: A deep reinforcement learning approach. **European Journal of Operational Research**, 16 maio. 2020. v. 283, n. 1, p. 166–181. . Acesso em: 8 jan. 2023.

LUO, Y. Application of Reinforcement Learning Algorithm Model in Gas Path Fault Intelligent Diagnosis of Gas Turbine. **Computational Intelligence and Neuroscience**, 2021. v. 2021. . Acesso em: 8 jan. 2023.

MAHMOODZADEH, Z. *et al.* Condition-based maintenance with reinforcement learning for dry gas pipeline subject to internal corrosion. **Sensors (Switzerland)**, 1 out. 2020. v. 20, n. 19, p. 1–26.

MARTÍNEZ-MARADIAGA, D.; BRUNO, J. C.; CORONAS, A. Steady-state data reconciliation for absorption refrigeration systems. **Applied Thermal Engineering**, 2013. v. 51, n. 1–2, p. 1170–1180. . Acesso em: 8 jan. 2023.

MARUYAMA, S.; MORIYA, S. Newton's Law of Cooling: Follow up and exploration. **International Journal of Heat and Mass Transfer**, 1 jan. 2021. v. 164, p. 120544. . Acesso em: 8 out. 2022.

MENDES, T. *et al.* Uma revisão sobre técnicas de diagnóstico termodinâmico em geral, e aplicados a sistemas de refrigeração e ar condicionado por compressão de vapor. **CBIM**, 2011. v. 10, p. 1151–1168.

MENG, F.; BAI, Y.; JIN, J. An advanced real-time dispatching strategy for a distributed energy system based on the reinforcement learning algorithm. **Renewable Energy**, 1 nov. 2021. v. 178, p. 13–24. . Acesso em: 8 jan. 2023.

MNIH, V. *et al.* Playing Atari with Deep Reinforcement Learning. 19 dez. 2013. Disponível em: <https://arxiv.org/abs/1312.5602v1>. Acesso em: 12 jul. 2022.

MOBLEY, R. K. An introduction to predictive maintenance. 2002. p. 438. . Acesso em: 14 set. 2022.

MOHAMMADI, R.; HE, Q. A deep reinforcement learning approach for rail renewal and maintenance planning. **Reliability Engineering and System Safety**, 1 set. 2022. v. 225. . Acesso em: 8 jan. 2023.

MUNGUBA, C. F. L. *et al.* Integração Entre Geradores Fotovoltaicos e Retrofit Energético Em Edifícios. **Revista de Engenharia e Pesquisa Aplicada**, 16 maio. 2020. v. 5, n. 3, p. 28–39. Disponível em: <http://revistas.poli.br/~anais/index.php/repa/article/view/1268/636>. Acesso em: 9 out. 2022.

NAJAFI, M. *et al.* A statistical pattern analysis framework for rooftop unit diagnostics. **HVAC and R Research**, 6 maio. 2012. v. 18, n. 3, p. 406–416. . Acesso em: 6 jan. 2023.

NASCIMENTO, A. S. B. D. S.; FLESCH, R. C. C.; FLESCH, C. A. Data-driven soft sensor for the estimation of sound power levels of refrigeration compressors through vibration measurements. **IEEE Transactions on Industrial Electronics**, 1 ago. 2020. v. 67, n. 8, p. 7065–7072.

NGUYEN, V. T. *et al.* Artificial-intelligence-based maintenance decision-making and optimization for multi-state component systems. **Reliability Engineering and System Safety**, 1 dez. 2022. v. 228. . Acesso em: 8 jan. 2023.

NOVAES PIRES LEITE, G. DE. Métodos inteligentes para detecção de falha em sistemas de refrigeração. Recife: IFPE, [s.d.].

ONG, K. S. H.; WANG, W.; NIYATO, D.; *et al.* Deep-Reinforcement-Learning-Based Predictive Maintenance Model for Effective Resource Management in Industrial IoT. **IEEE Internet of Things Journal**, 1 abr. 2022. v. 9, n. 7, p. 5173–5188. . Acesso em: 8 jan. 2023.

_____; _____; HIEU, N. Q.; *et al.* Predictive Maintenance Model for IIoT-Based Manufacturing: A Transferable Deep Reinforcement Learning Approach. **IEEE Internet of Things Journal**, 1 set. 2022. v. 9, n. 17, p. 15725–15741. . Acesso em: 8 jan. 2023.

PARASCHOS, P. D.; KOULINAS, G. K.; KOULOURIOTIS, D. E. Reinforcement learning for combined production-maintenance and quality control of a manufacturing system with deterioration failures. **Journal of Manufacturing Systems**, 1 jul. 2020. v. 56, p. 470–483. . Acesso em: 8 jan. 2023.

PENG, S.; FENG, Q. (May). Reinforcement learning with Gaussian processes for condition-based maintenance. **Computers and Industrial Engineering**, 1 ago. 2021. v. 158. . Acesso em: 8 jan. 2023.

PENG, Y.; DONG, M.; ZUO, M. J. Current status of machine prognostics in condition-based maintenance: a review. **The International Journal of Advanced Manufacturing Technology 2009 50:1**, 6 jan. 2010. v. 50, n. 1, p. 297–313. Disponível em: <https://link.springer.com/article/10.1007/s00170-009-2482-0>. Acesso em: 22 set. 2022.

PINCIROLI, L. *et al.* Deep reinforcement learning based on proximal policy optimization for the maintenance of a wind farm with multiple crews. **Energies**, 1 out. 2021. v. 14, n. 20. . Acesso em: 8 jan. 2023.

REN, N. *et al.* Fault diagnosis strategy for incompletely described samples and its application to refrigeration system. **Mechanical Systems and Signal Processing**, 1 fev. 2008. v. 22, n. 2, p. 436–450. . Acesso em: 6 jan. 2023.

RENARD, S.; CORBETT, B.; SWEI, O. Minimizing the global warming impact of pavement infrastructure through reinforcement learning. **Resources, Conservation and Recycling**, 1 abr. 2021. v. 167. . Acesso em: 8 jan. 2023.

RIBEIRO, C. H. C. **A Tutorial on Reinforcement Learning Techniques**. **ITA**. Instituto de Tecnologia da Aeronautica.

ROCCHETTA, R. *et al.* A reinforcement learning framework for optimal operation and maintenance of power grids. **Applied Energy**, 1 maio. 2019. v. 241, p. 291–301.

RODRIGUES, J. G. Das N. **Aprendizagem Automática Aplicada à Condução de um Veículo com Direção Ackermann**. Aveiro: [s.n.], 2018. Disponível em: <https://ria.ua.pt/bitstream/10773/25129/1/Documento.pdf>. Acesso em: 28 ago. 2022.

RUAN, J. H. *et al.* A reinforcement learning-based algorithm for the aircraft maintenance routing problem. **Expert Systems with Applications**, 1 maio. 2021. v. 169. . Acesso em: 8 jan. 2023.

RUIZ RODRÍGUEZ, M. L. *et al.* Multi-agent deep reinforcement learning based Predictive Maintenance on parallel machines. **Robotics and Computer-Integrated Manufacturing**, 1 dez. 2022. v. 78. . Acesso em: 8 jan. 2023.

SANUSI, I. *et al.* Reinforcement learning for condition-based control of gas turbine engines. **2019 18th European Control Conference, ECC 2019**, 1 jun. 2019. p. 3928–3933. . Acesso em: 22 set. 2022.

SCHWINDEN LEAL, E. **Predição de falhas em equipamentos de refrigeração com técnicas de aprendizado de máquina**. Florianópolis: [s.n.], 2019.

SCIPY TEAM. scipy.stats.weibull_min — SciPy v1.9.2 Manual. 2022. Disponível em: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.weibull_min.html#scipy.stats.weibull_min>. Acesso em: 9 out. 2022.

SEYR, H.; MUSKULUS, M. Use of Markov decision processes in the evaluation of corrective maintenance scheduling policies for offshore wind farms. **Energies**, 3 ago. 2019. v. 14, n. 15.

SHI, S. *et al.* An efficient VRF system fault diagnosis strategy for refrigerant charge amount based on PCA and dual neural network model. **Applied Thermal Engineering**, 25 jan. 2018. v. 129, p. 1252–1262. . Acesso em: 6 jan. 2023.

SHI, Z. A Probabilistic Distributed Fault Detection, Diagnostics and Evaluation Framework for Building Systems. Ottawa, Ontario: 2018. Disponível em: <https://curve.carleton.ca/e949251b-ff18-41f2-bf9a-aa55ea35fdbb>. Acesso em: 6 jan. 2023.

SOLTANI, Z. *et al.* Fault detection and diagnosis in refrigeration systems using machine learning algorithms. **International Journal of Refrigeration**, 1 dez. 2022. v. 144, p. 34–45. . Acesso em: 13 nov. 2022.

SUN, J. *et al.* Fault detection of low global warming potential refrigerant supermarket refrigeration system: Experimental investigation. **Case Studies in Thermal Engineering**, 1 ago. 2021. v. 26. . Acesso em: 6 nov. 2022.

SUN, K. *et al.* A novel efficient SVM-based fault diagnosis method for multi-split air conditioning system's refrigerant charge fault amount. **Applied Thermal Engineering**, 5 set. 2016. v. 108, p. 989–998. . Acesso em: 6 jan. 2023.

SUN, S. *et al.* A hybrid ICA-BPNN-based FDD strategy for refrigerant charge faults in variable refrigerant flow system. **Applied Thermal Engineering**, 25 dez. 2017. v. 127, p. 718–728.

SUN, Z. *et al.* Gradual fault early stage diagnosis for air source heat pump system using deep learning techniques. **International Journal of Refrigeration**, 1 nov. 2019. v. 107, p. 63–72.

SUTTON, R. S. Learning to predict by the methods of temporal differences. **Machine Learning 1988 3:1**, ago. 1988. v. 3, n. 1, p. 9–44. Disponível em: <https://link.springer.com/article/10.1007/BF00115009>. Acesso em: 7 out. 2022.

SUTTON, R. S.; BARTO, A. G. **Reinforcement Learning: An Introduction Second edition, in progress**. [S.l.]: [s.n.], 2018.

TANIMOTO, A. Combinatorial Q-Learning for Condition-Based Infrastructure Maintenance. **IEEE Access**, 2021. v. 9, p. 46788–46799. . Acesso em: 8 jan. 2023.

TENSORFORCE TEAM. Tensorforce: a TensorFlow library for applied reinforcement learning — Tensorforce 0.6.5 documentation. 2022. Disponível em: <https://tensorforce.readthedocs.io/en/latest/>. Acesso em: 5 out. 2022.

TRAN, D. A. T. *et al.* A robust online fault detection and diagnosis strategy of centrifugal chiller systems for building energy efficiency. **Energy and Buildings**, 1 dez. 2015. v. 108, p. 441–453. . Acesso em: 6 jan. 2023.

TROTT, A. R. (Albert R.; WELCH, T. Refrigeration and air-conditioning. 2000. p. 377. . Acesso em: 7 jan. 2023.

VALET, A. *et al.* Opportunistic maintenance scheduling with deep reinforcement learning. **Journal of Manufacturing Systems**, 1 jul. 2022a. v. 64, p. 518–534. . Acesso em: 16 nov. 2022.

_____ *et al.* Opportunistic maintenance scheduling with deep reinforcement learning. **Journal of Manufacturing Systems**, 1 jul. 2022b. v. 64, p. 518–534. . Acesso em: 8 jan. 2023.

WANG, H.; YAN, Q.; ZHANG, S. Integrated scheduling and flexible maintenance in deteriorating multi-state single machine system using a reinforcement learning approach. **Advanced Engineering Informatics**, 1 ago. 2021. v. 49. . Acesso em: 8 jan. 2023.

WANG, Jiangyu *et al.* Liquid floodback detection for scroll compressor in a VRF system under heating mode. **Applied Thermal Engineering**, 5 mar. 2017. v. 114, p. 921–930. . Acesso em: 6 jan. 2023.

WANG, Jing; LEI, D.; CAI, J. An adaptive artificial bee colony with reinforcement learning for distributed three-stage assembly scheduling with maintenance. **Applied Soft Computing**, 1 mar. 2022. v. 117. . Acesso em: 8 jan. 2023.

WANG, S. K. (Shan K.; LAVAN, Z. (Zalman); NORTON, P. (Paul M. Air conditioning and refrigeration engineering. 2000. p. 86. . Acesso em: 10 out. 2022.

WANG, Xiao *et al.* Predictive Maintenance and Sensitivity Analysis for Equipment with Multiple Quality States. **Mathematical Problems in Engineering**, 2021. v. 2021. . Acesso em: 8 jan. 2023.

WATKINS, C. J. C. H. **Learning From Delayed Rewards**. London: King's College, 1989. Disponível em: <https://www.cs.rhul.ac.uk/~chrisw/new_thesis.pdf>. Acesso em: 13 set. 2022.

WEI, S.; BAO, Y.; LI, Hui. Optimal policy for structure maintenance: A deep reinforcement learning framework. **Structural Safety**, 1 mar. 2020. v. 83. . Acesso em: 8 jan. 2023.

WEI, T.; WANG, Yanzhi; ZHU, Q. Deep Reinforcement Learning for Building HVAC Control. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2017. V. Part 128280.

WICHMAN, A.; BRAUN, J. E. Fault detection and diagnostics for commercial coolers and freezers. **HVAC and R Research**, 2009. v. 15, n. 1, p. 77–99. . Acesso em: 22 set. 2022.

WU, S.; SUN, J. Q. A top-down strategy with temporal and spatial partition for fault detection and diagnosis of building HVAC systems. **Energy and Buildings**, 1 set. 2011. v. 43, n. 9, p. 2134–2139. . Acesso em: 6 jan. 2023.

XIAO, F. *et al.* Bayesian network based FDD strategy for variable air volume terminals. **Automation in Construction**, 1 maio. 2014. v. 41, p. 106–118. . Acesso em: 6 jan. 2023.

YAN, K. *et al.* ARX model based fault detection and diagnosis for chillers using support vector machines. **Energy and Buildings**, 1 out. 2014. v. 81, p. 287–295. . Acesso em: 6 jan. 2023.

_____ *et al.* Cost-sensitive and sequential feature selection for chiller fault detection and diagnosis. **International Journal of Refrigeration**, 1 fev. 2018. v. 86, p. 401–409. . Acesso em: 6 jan. 2023.

_____ *et al.* Unsupervised learning for fault detection and diagnosis of air handling units. **Energy and Buildings**, 2020. v. 210, p. 109689.

_____; HUA, J. Deep learning technology for chiller faults diagnosis. **Proceedings - IEEE 17th International Conference on Dependable, Autonomic and Secure Computing, IEEE 17th International Conference on Pervasive Intelligence and Computing, IEEE 5th International Conference on Cloud and Big Data Computing, 4th Cyber Scienc**, 2019. p. 72–79.

_____; JI, Z.; SHEN, W. Online fault detection methods for chillers combining extended kalman filter and recursive one-class SVM. **Neurocomputing**, 8 mar. 2017. v. 228, p. 205–212. . Acesso em: 6 jan. 2023.

YANG, A. *et al.* Condition-based maintenance strategy for redundant systems with arbitrary structures using improved reinforcement learning. **Reliability Engineering and System Safety**, 1 set. 2022. v. 225. . Acesso em: 8 jan. 2023.

YANG, D. Y. Adaptive Risk-Based Life-Cycle Management for Large-Scale Structures Using Deep Reinforcement Learning and Surrogate Modeling. **Journal of Engineering Mechanics**, jan. 2022. v. 148, n. 1. . Acesso em: 8 jan. 2023.

YANG, H.; LI, Wenchao; WANG, B. Joint optimization of preventive maintenance and production scheduling for multi-state production systems based on reinforcement learning. **Reliability Engineering and System Safety**, 1 out. 2021. v. 214. . Acesso em: 8 jan. 2023.

YANG, Y.; YAO, L. Optimization Method of Power Equipment Maintenance Plan Decision-Making Based on Deep Reinforcement Learning. **Mathematical Problems in Engineering**, 2021. v. 2021. . Acesso em: 8 jan. 2023.

YOUSEFI, N.; TSIANIKAS, S.; COIT, D. W. Reinforcement learning for dynamic condition-based maintenance of a system with individually repairable components. **Quality Engineering**, 2 jul. 2020. v. 32, n. 3, p. 388–408.

_____; _____; _____. Dynamic maintenance model for a repairable multi-component system using deep reinforcement learning. **Quality Engineering**, 2022. v. 34, n. 1, p. 16–35. . Acesso em: 8 jan. 2023.

ZHANG, D.; GAO, Z. Improvement of Refrigeration Efficiency by Combining Reinforcement Learning with a Coarse Model. **Processes 2019, Vol. 7, Page 967**, 17 dez. 2019. v. 7, n. 12, p. 967. Disponível em: <https://www.mdpi.com/2227-9717/7/12/967/htm>. Acesso em: 31 jul. 2022.

ZHANG, Huidong; DJURDJANOVIC, D. Integrated production and maintenance planning under uncertain demand with concurrent learning of yield rate. **Flexible Services and Manufacturing Journal**, 1 jun. 2022. v. 34, n. 2, p. 429–450. . Acesso em: 8 jan. 2023.

ZHANG, N.; SI, W. Deep reinforcement learning for condition-based maintenance planning of multi-component systems under dependent competing risks. **Reliability Engineering and System Safety**, 1 nov. 2020a. v. 203.

_____; _____. Deep reinforcement learning for condition-based maintenance planning of multi-component systems under dependent competing risks. **Reliability Engineering and System Safety**, 1 nov. 2020b. v. 203. . Acesso em: 8 jan. 2023.

ZHANG, P.; ZHU, X.; XIE, M. A model-based reinforcement learning approach for maintenance optimization of degrading systems in a large state space. **Computers and Industrial Engineering**, 1 nov. 2021. v. 161.

ZHANG, Y. *et al.* Hierarchical Deep Reinforcement Learning for Backscattering Data Collection with Multiple UAVs. **IEEE Internet of Things Journal**, 1 mar. 2021. v. 8, n. 5, p. 3786–3800. . Acesso em: 8 jan. 2023.

ZHANG, Z. *et al.* Novel application of multi-model ensemble learning for fault diagnosis in refrigeration systems. **Applied Thermal Engineering**, 5 jan. 2020. v. 164.

ZHAO, X. Lab test of three fault detection and diagnostic methods' capability of diagnosing multiple simultaneous faults in chillers. **Energy and Buildings**, 1 maio. 2015. v. 94, p. 43–51. . Acesso em: 6 jan. 2023.

ZHAO, Yunfei; SMIDTS, C. Reinforcement learning for adaptive maintenance policy optimization under imperfect knowledge of the system degradation model and partial observability of system states. **Reliability Engineering and System Safety**, 1 ago. 2022. v. 224. . Acesso em: 8 jan. 2023.

ZHONG, C. *et al.* Energy efficiency solutions for buildings: Automated fault diagnosis of air handling units using generative adversarial networks. **Energies**, 7 fev. 2019. v. 12, n. 3.

ZHOU, W. *et al.* A Reinforcement Learning Method for Multiasset Roadway Improvement Scheduling Considering Traffic Impacts. **Journal of Infrastructure Systems**, dez. 2022. v. 28, n. 4. . Acesso em: 8 jan. 2023.

ZHOU, Yifan; LI, B.; LIN, T. R. Maintenance optimisation of multicomponent systems using hierarchical coordinated reinforcement learning. **Reliability Engineering and System Safety**, 1 jan. 2022. v. 217. . Acesso em: 8 jan. 2023.

ZHOU, Z.; LI, G.; *et al.* A comparison study of basic data-driven fault diagnosis methods for variable refrigerant flow system. **Energy and Buildings**, 2020a. v. 224, p. 110232.

_____; WANG, Jiangyu; *et al.* An online compressor liquid floodback fault diagnosis method for variable refrigerant flow air conditioning system. **International Journal of Refrigeration**, 1 mar. 2020. v. 111, p. 9–19.

_____; LI, G.; *et al.* A comparison study of basic data-driven fault diagnosis methods for variable refrigerant flow system. **Energy and Buildings**, 1 out. 2020b. v. 224.

ZOGG, D.; SHAFAI, E.; GEERING, H. P. Fault diagnosis for heat pumps with parameter identification and clustering. **Control Engineering Practice**, 1 dez. 2006. v. 14, n. 12, p. 1435–1444. . Acesso em: 6 jan. 2023.