UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE ENERGIAS ALTERNATIVAS E RENOVÁVEIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Metodologia para a Extração Automática de Regras para a Modelagem de um Sistema de Refrigeração de Água utilizando Abordagem Digital Twin

Carlos Antonio Alves de Araújo Junior

João Pessoa
Fevereiro – 2021

CARLOS ANTONIO ALVES DE ARAÚJO JUNIOR

METODOLOGIA PARA A EXTRAÇÃO AUTOMÁTICA DE REGRAS PARA A MODELAGEM DE UM SISTEMA DE REFRIGERAÇÃO DE ÁGUA UTILIZANDO ABORDAGEM DIGITAL TWIN

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica - PPGEE, da Universidade Federal da Paraíba - UFPB, como requisito parcial para a obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Dr. Juan Moises Maurício Villanueva

JOÃO PESSOA

2021

Catalogação na publicação
Seção de Catalogação e Classificação

A658m Araújo Junior, Carlos Antonio Alves de.
Metodologia para a extração automática de regras para a modelagem de um sistema de refrigeração de água utilizando abordagem digital twin / Carlos Antonio Alves de Araújo Junior. - João Pessoa, 2021.
88 f. : il.

Orientação: Juan Moises Maurício Villanueva.
Dissertação (Mestrado) - UFPB/CEAR.

1. Inteligência artificial. 2. Setor elétrico. 3. Gêmeo digital. 4. Planta térmica. 5. Lógica fuzzy. I. Villanueva, Juan Moises Maurício. II. Título.

UFPB/BC CDU 004.8(043)

Elaborado por Gracilene Barbosa Figueiredo - CRB-15/794

UNIVERSIDADE FEDERAL DA PARAÍBA – UFPB
CENTRO DE ENERGIAS ALTERNATIVAS E RENOVÁVEIS – CEAR
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA - PPGEE

A Comissão Examinadora, abaixo assinada, aprova a Dissertação

METODOLOGIA PARA A EXTRAÇÃO AUTOMÁTICA DE REGRAS PARA A MODELAGEM DE UM SISTEMA DE REFRIGERAÇÃO DE ÁGUA UTILIZANDO ABORDAGEM DIGITAL TWIN

Elaborada por

CARLOS ANTÔNIO ALVES DE ARAÚJO JUNIOR

como requisito parcial para obtenção do grau de
Mestre em Engenharia Elétrica.

COMISSÃO EXAMINADORA

PROF. DR. JUAN MOISES MAURICIO VILLANUEVA
Orientador – UFPB

PROF. DR. FABIANO SALVADORI
Examinador Interno – UFPB

PROF. DR. RICARDO TANSCHEIT
Examinador Externo – PUC-RIO

João Pessoa/PB, 26 de fevereiro de 2021

SUMÁRIO

LISTA DE ILUSTRAÇÕES IV
LISTA DE TABELAS VI
AGRADECIMENTOS VIII
RESUMO IX
ABSTRACT X
1 INTRODUÇÃO 12
1.1 OBJETIVOS 15
1.2 ESTRUTURA DO TRABALHO 15
2 FUNDAMENTAÇÃO TEÓRICA 17
2.1 UTE TERMOPARAÍBA E UTE TERMONORDESTE 17
2.2 SISTEMAS AUXILIARES 20
2.2.1 Sistema de tratamento de óleo 20
2.2.2 Sistema de alimentação de combustível 21
2.2.3 Sistema de compressão de ar 22
2.2.4 Sistema de geração de vapor 23
2.2.5 Sistema de resfriamento de água 23
2.3 LÓGICA FUZZY 24
2.4 SISTEMA DE INFERÊNCIA FUZZY 26
2.5 EXTRAÇÃO AUTOMÁTICA DE REGRAS 27
2.6 ESTRATÉGIAS EVOLUTIVAS 30
3 PROPOSTA DO TRABALHO 38
3.1 SISTEMA DE REFRIGERAÇÃO DO MOTOR 38
3.2 METODOLOGIA PARA MODELO DA PLANTA USANDO EXTRAÇÃO DE REGRAS AUTOMÁTICA 41
3.2.1 Coleta de dados 43
3.2.2 Pré-processamento de dados 43
3.2.3 Seleção das Variáveis usando PCA 45
3.2.4 Extração automática de regras usando Lógica Fuzzy 48
3.2.5 Atualização da base de conhecimento de forma dinâmica 52
3.2.6 Otimização utilizando estratégias evolutivas 55
3.2.6.1 Inicialização da população 56
3.2.6.2 Avaliação da função fitness 58

3.2.6.3 Método de Seleção 59
3.2.6.4 Cruzamento 59
3.2.6.5 Mutação 60
4 RESULTADOS 62
4.1 FUNCIONAMENTO DO MODELO DURANTE PERÍODO DE REGIME PERMANENTE 66
4.2 FUNCIONAMENTO DO MODELO DURANTE A RAMPA DE PARADA E RAMPA DE PARTIDA DA PLANTA 69
4.3 FUNCIONAMENTO DO MODELO DURANTE PERÍODO COM PERTURBAÇÃO GERADO POR UM PROBLEMA NA MALHA DE CONTROLE DO SISTEMA DE REFRIGERAÇÃO 72
4.4 OTIMIZAÇÃO DO PROCESSO A PARTIR DO GÊMEO DIGITAL 74
4.5 CONSIDERAÇÕES E DISCUSSÕES 78
5 CONCLUSÃO 81
REFERÊNCIAS 83

LISTA DE ILUSTRAÇÕES

FIGURA 1 - FOTO AEREA DAS USINAS TERMOPARAÍBA E TERMONORDESTE 17
FIGURA 2 - MAPA DE LINHAS DE TRANSMISSÃO DA REGIÃO LESTE DO SIN. 18
FIGURA 3 - FOTO AÉREA DA PLANTA TÉRMICA DA EPASA. 20
FIGURA 4 - MÓDULO BOOSTER. 22
FIGURA 5 - REPRESENTAÇÃO DE UM CONJUNTO FUZZY. 25
FIGURA 6 - ESQUEMA DE UM SISTEMA DE INFERÊNCIA FUZZY 26
FIGURA 7 – MAPEAMENTO DAS ENTRADAS E SAÍDAS DE UM SISTEMA GENÉRICO. 28
FIGURA 8 - ESTRUTURA BÁSICA DE UM ALGORITMO EVOLUTIVO. 32
FIGURA 9 - CRUZAMENTO EM UM PONTO K=3 EM UM VETOR DE ALELOS REAIS. 36
FIGURA 10 - (A) OPERAÇÃO DE MUTAÇÃO INVERSÃO DE BIT E (B) MUTAÇÃO CREEP. 36
FIGURA 11 - SISTEMA DE REFRIGERAÇÃO DE ÁGUA PARA UM LADO DO RADIADOR. 39
FIGURA 12 - CIRCUITO DE REFRIGERAÇÃO DE UM MOTOR. 40
FIGURA 13 – A) COLETA DE DADOS, B) CONSTRUÇÃO DO MODELO INICIAL, C) ATUALIZAÇÃO DO MODELO DURANTE A OPERAÇÃO 42
FIGURA 14 - VISUALIZAÇÃO DE DADOS APÓS REMOÇÃO DE OUTLIERS - VARIÁVEL TEMPERATURA DE ENTRADA DE HT 44
FIGURA 15 - TEMPERATURA DE ENTRADA E SAIDA NO RADIADOR DURANTE PRÉ E PÓS PARTIDA DA PLANTA. 45
FIGURA 16 - MODELO INICIAL PROPOSTO DE DIGITAL TWIN DO RADIADOR. 46
FIGURA 17 - VARIAÇÃO EXPLICADA POR NÚMERO DE COMPONENTES UTILIZADO. 47
FIGURA 18 - MODELO PROPOSTO APÓS REDUÇÃO DE VARIÁVEIS USANDO PCA. 47
FIGURA 19 - CONJUNTOS FUZZY DOS ESPAÇOS TEMPERATURA TENTHT E TSAIDAHT. 49
FIGURA 20 - CONJUNTOS FUZZY DOS ESPAÇOS PCA 1, 2, 3, 4. 50
FIGURA 21 - CONJUNTOS FUZZY DO ESPAÇO NÚMERO DE VENTILADORES LIGADOS. 50
FIGURA 22 - VISUALIZAÇÃO DO BANCO DE DADOS DE TREINAMENTO, DADOS NA FORMA ORIGINAL. 51
FIGURA 23 - VISUALIZAÇÃO DO BANCO DE DADOS DE TREINAMENTO, DADOS NA FORMA REDUZIDA ATRAVÉS DE PCA. 52
FIGURA 24 - FLUXOGRAMA DO ALGORITMO DE ATUALIZAÇÃO DE REGRAS. 55
FIGURA 25 - ESTRUTURA DA POPULAÇÃO INICIAL DO AG. 58
FIGURA 26 - VARIAÇÃO DO ERRO QUADRÁTICO E DO NÚMERO DE REGRAS EM FUNÇÃO DO NÚMERO DE CONJUNTO FUZZY 63
FIGURA 27 - RESULTADO DO ALGORITMO GENÉTICO EM FUNÇÃO DO NÚMERO DE CONJUNTOS FUZZY 64
FIGURA 28 - ENTRADAS DO PROCESSO NO CENÁRIO 1, (A) POTÊNCIAS DAS UGD'S, (B) TENTHT E TAMB E (C) ENTRADAS REDUZIDAS E NUMVENTON. 68
FIGURA 29 - RESULTADO MODELO VERSUS VALOR REAL TSAIDAHT E ERRO PERCENTUAL INSTANTÂNEO DO CENÁRIO 1 69
FIGURA 30 - ENTRADAS DO PROCESSO NO CENÁRIO 2, (A) POTÊNCIAS DAS UGD'S, (B) TENTHT E TAMB E (C) ENTRADAS REDUZIDAS E NUMVENTON. 70
FIGURA 31 - RESULTADO MODELO VERSUS VALOR REAL TSAIDAHT E ERRO PERCENTUAL INSTANTÂNEO DO CENÁRIO 2 71
FIGURA 32 - RESULTADO DA SIMULAÇÃO DA AUTO VALIDAÇÃO DO MODELO VERSUS VALOR REAL DE SAÍDA DE TEMPERATURA DO SISTEMA E GRÁFICO DO ERRO PERCENTUAL INSTANTÂNEO, NO CENÁRIO 2. 72
FIGURA 33 - ENTRADAS DO PROCESSO NO CENÁRIO 3, (A) POTÊNCIAS DAS UGD'S, (B) TENTHT E TAMB E (C) ENTRADAS REDUZIDAS E NUMVENTON. 73

FIGURA 34 - RESULTADO MODELO X VALOR REAL TSAIDAHT E ERRO PERCENTUAL INSTANTÂNEO DO CENÁRIO 3. 74
FIGURA 35 - PLANTA DO SISTEMA UTILIZADO PARA OTIMIZAR O NÚMERO DE VENTILADORES DO PROCESSO. 75
FIGURA 36 - SUPERFÍCIE 3D DO SISTEMA DE INFERÊNCIA DO OTIMIZADOR. 76
FIGURA 37 - CONJUNTOS FUZZY DO CONTROLADOR OTIMIZADOR DE VENTILADORES. 76
FIGURA 38 - RESULTADO DO OTIMIZADOR DE VENTILADORES PARA O CENÁRIO 1 77

LISTA DE TABELAS

TABELA 1 - PSEUDO-CÓDIGO DO ALGORITMO DE SELEÇÃO DA ROLETA. 35
TABELA 2 - PSEUDO-CÓDIGO ALGORITMO PARA CRIAÇÃO DA POPULAÇÃO INICIAL. 57
TABELA 3 - BANCO DE REGRAS FUZZY DO OTIMIZADOR. 75
TABELA 4 - LEGENDA DAS REGRAS FUZZY DO OTIMIZADOR. 75
TABELA 5 - COMPARAÇÃO DOS ERROS ENTRE OS CENÁRIOS 79

A meu pai Carlos, minha mãe Ana Maria, meu irmão Aristóteles e minha esposa Ligia, por sempre me apoiarem e motivarem a buscar meus sonhos.

AGRADECIMENTOS

A Deus pelo dom da vida, por toda saúde e proteção proporcionados a mim e às pessoas que amo.

Ao meu orientador Professor Doutor Juan Moises Maurício Villanueva por toda dedicação e atenção dadas a mim, por sempre estar disponível para fazer reuniões de orientação, mesmo que tarde da noite, e pela grande contribuição no meu crescimento na área de Indústria 4.0 e Data Science durante esses dois anos de pesquisa.

À minha família, por serem meu apoio e me motivarem a ser uma pessoa melhor como também um profissional melhor. A meu pai Carlos e minha mãe Ana, por todo investimento feito na minha formação pessoal e acadêmica, por muitas vezes terem sido privados de desejos próprios para investir nos meus estudos. Ao meu irmão Aristóteles pela parceria diária e discursões técnicas que contribuíram diretamente na pesquisa e no meu crescimento intelectual. À minha esposa Ligia pelo companheirismo e amor, por me apoiar em todos os momentos e me ajudar cada dia a crescer como ser humano.

Aos meus amigos de longa data Rafael Angelo, Rodrigo Massao, Diego Augusto, Pedro Paiva, Hugo Lucena e Elvin Linhares, pelos momentos de diversão, jogatinas e conversas aleatórias que fazem com que eu renove minhas forças para continuar lutando no dia a dia

À EPASA que me permitiu a participação no Projeto de P&D intitulado “Otimização Operacional utilizando Digital Twins”, vinculado ao Programa de Pesquisa e Desenvolvimento do Setor Elétrico regulado pela Agência Nacional de Energia Elétrica – ANEEL, ao Diretor-presidente Senhor José Ferreira Abdal Neto e ao Diretor de O&M Senhor Franklin de Araújo Neto por apoiar e motivar o crescimento individual dos seus colaboradores e disponibilizar os dados utilizados neste trabalho. Aos meus colegas de trabalho e amigos Rodrigo Almeida e Isaac Emmanuel pela grande parceria na EPASA e pelas discussões técnicas sobre Indústria 4.0 e tecnologia que tanto contribuíram para meu crescimento profissional.

RESUMO

Na busca pelo aumento da produtividade, a indústria desenvolveu estratégias tecnológicas para atingir este objetivo: Indústria 4.0. No setor elétrico, plantas termelétricas buscam ferramentas capazes de otimizar a produção de energia e diminuir o intervalo de tempo de parada não programada. Este trabalho propôs realizar a modelagem do sistema de refrigeração de água a partir da abordagem "*Digital Twin*" (Gêmeo Digital), utilizando técnicas de inteligência artificial como a lógica fuzzy e extração automática de regras fuzzy. O sistema de refrigeração de água possui muitos ventiladores para realizar a troca térmica entre a água e o ar ambiente, porém, a forte característica de não-linearidade do sistema faz com que os controladores PID convencionais apresentem erro não nulo com o tempo, induzindo o operador a acionar todos os ventiladores para garantir a temperatura da água abaixo do necessário e, consequentemente, aumentando o gasto energético sem necessidade. Nesta abordagem, utilizaram-se os dados de uma planta termelétrica localizada na capital da Paraíba, João Pessoa, para criar um banco de conhecimento através do histórico do sistema e, assim, criar um gêmeo digital capaz de auxiliar na otimização do consumo de energia desse sistema, que chega a cerca de 3% de toda energia gerada na planta. Também foi proposto um algoritmo de atualização de regras do modelo em tempo real. A atualização tem o objetivo de garantir que o sistema tenha diminua o erro percentual conforme novas regras são aprendidas, como também adquirir novos conhecimentos de situações para os quais o modelo não foi treinado. Foi utilizado um algoritmo de otimização utilizando estratégias evolutivas a fim de buscar um menor número de conjuntos fuzzy, atingindo um menor esforço computacional sem maiores impactos no resultado. Os resultados apresentados pelo modelo a partir de testes, em três cenários diferentes, mostraram a robustez do modelo proposto, e em todos os cenários o erro percentual médio foi abaixo de 5% e erro absoluto médio abaixo de 3°C.

Palavra-chave: Setor Elétrico; Planta Térmica; Gêmeo Digital; Inteligência Artificial; Lógica Fuzzy; Extração Automática de Regras.

ABSTRACT

In the search of higher productivity, industry has developed technological strategies to achieve this goal, Industry 4.0. In the energy sector, thermal power plants are looking for tools capable of optimizing energy production and reducing unscheduled downtime. This work proposed to model the water cooling system from the Digital Twin approach, using artificial intelligence techniques such as fuzzy logic and automatic extraction of fuzzy rules. The water cooling system is a system that has many fans to perform the thermal exchange between water and ambient air, however, the strong non-linearity characteristic of the system means that conventional PID controllers do not have a good performance, inducing the operator to activate all the fans to guarantee the water temperature below the necessary and, consequently, increasing the energy expenditure unnecessarily. This approach uses data from a thermal power plant located in the capital of Paraíba, João Pessoa, to create a knowledge database through the history data of the system and, thus, create a digital twin capable of helping to optimize the energy consumption of this system, which arrives about 3% of all energy generated at the plant. An algorithm for online updating the model rules was also proposed. The update aims to ensure that the system has a low instantaneous percentage error, as well as to acquire new knowledge of situations in which the model has not been trained. An optimization algorithm was used, using evolutionary strategies to search for a smaller number of fuzzy sets, achieving less computational effort without major impacts on the result. The results presented by the model from tests, in three different scenarios, showed the robustness of the proposed model, and in all scenarios, the average percentage error was below 5% and the average absolute error below 3°C.

Keyword: Energy Sector; Thermo Powerplant; Digital Twin; Artificial Intelligence; Fuzzy Logic; Automatic Rule Extraction.

1 INTRODUÇÃO

1 INTRODUÇÃO

O Brasil possui um sistema elétrico de potência com características singulares, as quais demandam um grande planejamento e organização. Dentre essas características, podem-se citar:

- Dimensões continentais;
- Uma matriz energética diversificada, composta principalmente por fontes hidrelétrica, solar, eólica, biomassa e termelétrica;
- Alta capacidade de intercâmbio de energia entre as regiões, divididas em quatro subsistemas: Sul, Sudeste, Nordeste e Centro-Oeste [TOLMASQUIM, 2016];

A fonte termoelétrica tem sido usada para aumentar a segurança energética no sistema interligado nacional (SIN), assim como, também é usada para o suprimento energético em sistemas isolados (DE ALMEIDA, 2017). Nos últimos cinco anos, as usinas termelétricas foram responsáveis, em média, por 19% da energia gerada no Brasil (ONS), como consequência do aumento de geração de fontes renováveis intermitentes, solar e eólica, agindo de forma a estabilizar a variabilidade na geração em curto prazo. Além disso, as usinas termelétricas atuam de forma significativa durante períodos de secas, tendo em vista que a matriz energética brasileira é majoritariamente hídrica (EPE).

Para atender o sistema interligado, as Usinas Termelétricas precisam manter a alta disponibilidade e confiabilidade, além da preocupação de manter o alto desempenho operacional, a fim de reduzir custos no uso do combustível e, consequentemente, numa menor emissão de carbono para atmosfera (SAHA 2014). É possível mensurar o desempenho de uma planta termelétrica a partir de alguns indicadores, tais como consumo específico e eficiência térmica (SAHA 2014), mas existe outro ponto que impacta o desempenho - o consumo de energia nos equipamentos auxiliares. Tais equipamentos têm como função garantir que os parâmetros dos equipamentos principais (caldeiras, motores, geradores e turbinas), responsáveis para que a produção de energia esteja controlada e, para tal, tais equipamentos necessitam consumir energia elétrica da própria produção para manter seu funcionamento (KRUHOL 2019).

Normalmente, os equipamentos auxiliares são bombas, ventiladores e compressores de ar, tendo muitas vezes como configuração de acionamento dos motores elétricos a partida direta, ou seja, sem o controle fino de consumo de energia. Usualmente, esses equipamentos são dimensionados para garantir o funcionamento da planta em sua capacidade nominal, porém, muitas vezes a usina é demandada a gerar um valor menor que sua capacidade nominal, fazendo com que os equipamentos auxiliares operem em zonas consideradas ineficientes pelo fabricante (KRUHOL 2019).

Neste contexto, o sistema de refrigeração de água geralmente é o sistema com o maior consumo de uma usina termelétrica. Por exemplo, em uma usina de grande porte com geração de 340 MW, que possui 40 motores a combustão interna, bi-combustível sendo oléo pesado OCB1 como principal e oléo Diesel secundário, o consumo de um sistema de refrigeração chega a 2% da potência bruta gerada somente com esse sistema, o que é um valor bastante expressivo quando convertido em valor financeiro, que pode equivaler a um custo de aproximadamente R$ 90.000,00 reais por dia de geração (CCEE).

O sistema de refrigeração é composto por um circuito fechado de água com vários ventiladores, os quais são responsáveis pela retirada de calor na água do circuito utilizada para refrigerar a unidade geradora. Em algumas usinas, o sistema de refrigeração é compartilhado entre várias unidades geradoras, e os ventiladores utilizam partida direta, não sendo possível realizar o controle da temperatura a partir da variação de rotação deles. O compartilhamento da refrigeração torna o sistema extremamente não linear e dependente de vários parâmetros associadas às diferentes unidades geradoras.

Para minimizar a operação ineficiente, projetos de controladores clássicos, como proporcional-integral-derivativo (PID), normalmente são implementados em usinas termoelétricas mas esses controladores acabam não mantendo a eficiência do processo da maneira mais apropriada, perdendo a sintonia durante a operação e implicando na perda da otimização das características dinâmicas da resposta desejada, tais como amortecimento, sobre-sinal, tempo de acomodação e erro em regime permanente. Portanto, em muitas das instalações industriais, os controladores PID não estão apropriadamente sintonizados e, em alguns casos, são encontrados operando em modo manual devido à necessidade de constantes reajustes e sintonias.

Na busca da otimização operacional, as indústrias buscam por tecnologias que possam aumentar a eficiência a partir de ferramentas inovadoras, e é nesse contexto de inovação que aconteceu a quarta revolução industrial, a revolução da Indústria 4.0.

O termo Indústria 4.0 surgiu pela primeira vez em um artigo publicado pelo o governo alemão, em 2011, o qual definia a estratégia para pesquisa e desenvolvimento de tecnologias de ponta (ZHOU, 2015). Atualmente, esse tema vem sendo discutido, e hoje é um dos temas de maior destaque na indústria. Baseia-se na integração do ambiente de tecnologia da informação (TI) e do ambiente da tecnologia da automação (TA) industrial, a fim de tornar o ambiente mais iterativo entre pessoas, máquinas e produtos, em tempo real, a partir de um sistema ciber-físico (CPS - *Cyber-Physical System*) (ZHOU, 2015).

O conceito do termo Indústria 4.0 é um termo guarda-chuva para novos paradigmas industriais que abrangem o desenvolvimento de ferramentas que utilizam a análise de grandes volumes de dados (*Big Data Analytics)*, Internet das coisas industrial (*Industrial Internet of Things – IIoT)*, Robótica, Manufatura na nuvem e uso de ferramentas de realidade aumentada (PEREIRA, 2017). Essa revolução impulsionou o uso de técnicas que utilizam inteligência artificial para tratamento e análise de dados com diversas finalidades, tais como predição de falhas usando *machine learning* e a modelagem de processos e serviços utilizando gêmeos digitais (do inglês, *Digital Twins*).

Um gêmeo digital é um dos principais conceitos associados à Indústria 4.0, e apresenta diversos conceitos na indústria e no meio científico. Um desses conceitos é dado como um sistema virtual e computadorizado, capaz de simular um sistema físico para diversas finalidades através da exploração de dados em tempo real, ou não, dos sensores em campo (NEGRI, 2017).

Tal conceito associa o uso dos dados gerados pelos sensores em campo, em combinação com a criação de modelos computacionais, para detectar anomalias no processo de produção ou prever falhas em componentes, sem que haja a necessidade de parar a produção ou otimizar o planejamento da manutenção (ARMENDIA, 2019).

Em busca de ferramentas que possam melhorar o desempenho dos equipamentos de um processo de produção de energia de uma usina termelétrica, esta pesquisa teve como objetivo o desenvolvimento de um *digital twins* de um sistema

de refrigeração de água utilizando um sistema de inferência fuzzy baseado no algoritmo de extração automática de regras oriundas de dados dos sensores de campo. A construção deste modelo teve a finalidade de melhorar o desempenho e eficiência energética deste equipamento.

1.1 OBJETIVOS

- Objetivo geral

  Desenvolvimento de uma metodologia para a construção de um gêmeo digital de um sistema de refrigeração de uma usina termelétrica de forma a possibilitar seu uso para otimização do processo.

- Objetivos específicos
  1. Verificar a eficiência do uso da técnica de extração automática de regras fuzzy para compor o conhecimento do modelo.
  2. Verificar a possibilidade de utilizar um conjunto reduzido de dados (de um a dez dias de histórico do banco de dados) para a identificação do sistema.
  3. Criar uma estratégia de atualização do banco de conhecimento do modelo sem a necessidade do re-treino.
  4. Verificar o uso de estratégias evolutivas para otimização computacional do sistema de inferência fuzzy.

1.2 ESTRUTURA DO TRABALHO

Além do capítulo introdutório, este trabalho está composto por mais quatro capítulos.

No Capítulo 2, serão discutidas a fundamentação teórica envolvendo o sistema de refrigeração da usina, conceitos fundamentais sobre lógica fuzzy, técnicas de extração automática de regras e algoritmos evolutivos. No Capítulo 3, será apresentado a proposta do trabalho, detalhando os procedimentos metodológicos para a construção do *digital twins*. No Capítulo 4, serão apresentados os resultados da pesquisa. No Capítulo 5, serão apresentadas as conclusões desta pesquisa e considerações para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

2 FUNDAMENTAÇÃO TEÓRICA

Este Capítulo contemplará as principais características de uma usina termelétrica à combustão interna, seus principais equipamentos auxiliares, fundamentos de lógica fuzzy e procedimentos para a extração automática de regras. Será apresentada uma breve descrição das usinas da EPASA, UTE Termoparaíba e UTE Termonordeste, as quais cederam os dados do sistema de refrigeração para a realização dessa pesquisa.

2.1 UTE TERMOPARAÍBA E UTE TERMONORDESTE

As usinas Termoparaíba e Termonordeste estão localizadas no distrito industrial na cidade de João Pessoa, no Estado da Paraíba. A concessão para a operação destas duas usinas foi ganha pela EPASA – Centrais Elétricas da Paraíba S.A., vencedora do 4° Leilão de Energia Nova em julho de 2007. As obras foram iniciadas em outubro de 2009 e finalizadas em dezembro 2010, entrando em operação logo em seguida (Figura 1).

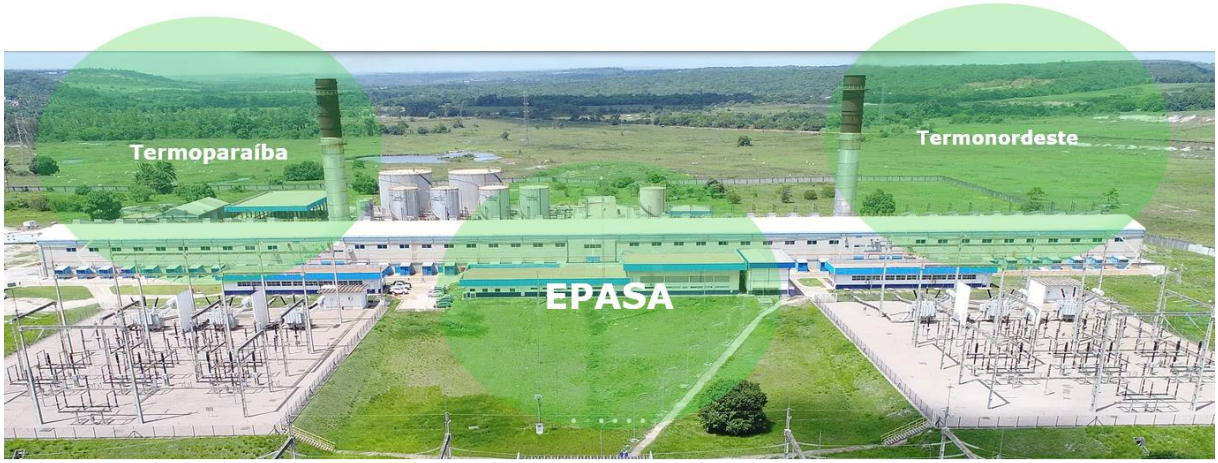


FIGURA 1 - FOTO AEREA DAS USINAS TERMOPARAÍBA E TERMONORDESTE

Cada usina possui potência instalada de 170,752MW distribuída em 20 unidades geradoras a motor de combustão interna, sendo 19 motores de modelo STX/MAN 18V32/40 de potência 12.287 cavalos e um motor de modelo STX/MAN 9L32/40 de potência 6.144 cavalos, que utilizam o óleo combustível pesado do tipo

OCB1 (oléo combustível de baixo teor de enxofre). Cada motor é acoplado a um gerador Hyundai modelo HSGD 1011K10, e este conjunto forma a unidade geradora. Nos motores 18V32/40, o gerador possui potência nominal de 10,954MVA e no motor 9L32/40, 5,444MVA. A potência total instalada a ser entregue ao SIN pela a EPASA é de 341,504MW através de 2 subestações elevadoras contendo três transformadores elevadores 13,8/230kV da WEG cada, com potências nominais de 66MVA, 72MVA e 77MVA, cada qual refere-se ao Grupo Gerador 1, 2 e 3, respectivamente, compostos por seis unidades geradoras no Grupo 1, sete unidades geradoras no Grupo 2 e sete unidades Geradoras no Grupo 3.

Cada subestação elevadora é conectada à subestação de manobra, da qual sai uma linha de transmissão de aproximadamente 5km que se conecta ao Sistema Inteligado Nacional, na Subestação SE Mussuré II, da CHESF, como ilustrado na Figura 2.

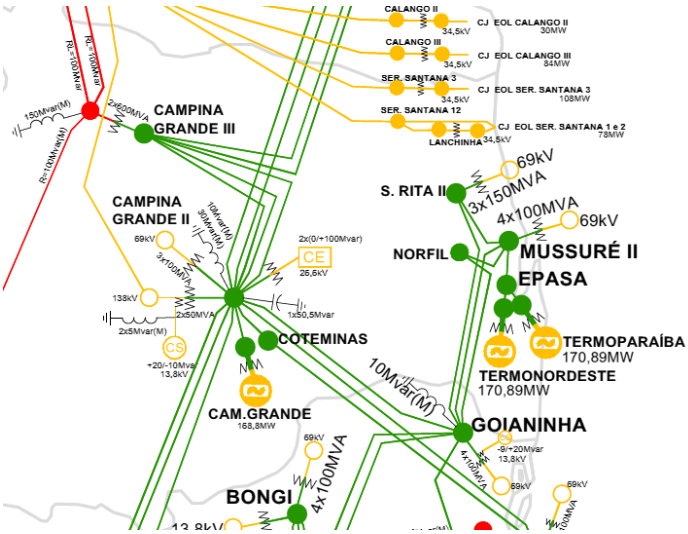


FIGURA 2 - MAPA DE LINHAS DE TRANSMISSÃO DA REGIÃO LESTE DO SIN.

Uma usina termelétrica à combustão interna é composta por diversas áreas e sistemas auxiliares que permitem o funcionamento em harmonia necessário para o processo de geração de energia elétrica. Na Figura 3, são identificados as áreas e alguns sistemas auxiliares contidos na planta:

- Parque de recebimento: nessa área é onde o óleo combustível e lubrificante é transferido dos caminhões para os tanques de armazenamento de combustível e óleo lubrificante;

- Parque de tanques (Tancagem): área onde ficam armazenados os insumos necessários para a geração, tais como óleo OCB1, óleo Diesel, óleo lubrificante. Também é armazenada toda a borra gerada pelo processo de geração de energia para posteriormente ser tratada. O parque possui dez tanques, sendo dois tanques de armazenamento e dois tanques de uso diário de OCB1, dois tanques de óleo combustível diesel, dois tanques de óleo de lubrificante e dois tanques de borra;
- Casa de bombas: área onde ficam as bombas de alimentação de óleo e as centrífugas de tratamento de óleo combustível para os motores;
- Casa de caldeiras auxiliares: área onde ficam as caldeiras auxiliares, as quais operam durante o tempo que a usina está parada, têm a funcionalidade de gerar vapor para aquecer os tanques de óleo combustíveis;
- Radiadores: localizados na área externa da usina, esse sistema é responsável por refrigerar a água utilizada nas unidades geradoras para estabilizar a temperatura do processo;
- Casa de máquinas: área onde se localizam as unidades geradoras (motor e gerador) e alguns equipamentos auxiliares necessários para o bom funcionamento da unidade geradora.
- Prédio administrativo: área onde ficam localizados os colaboradores da EPASA e a oficina eletromecânica.

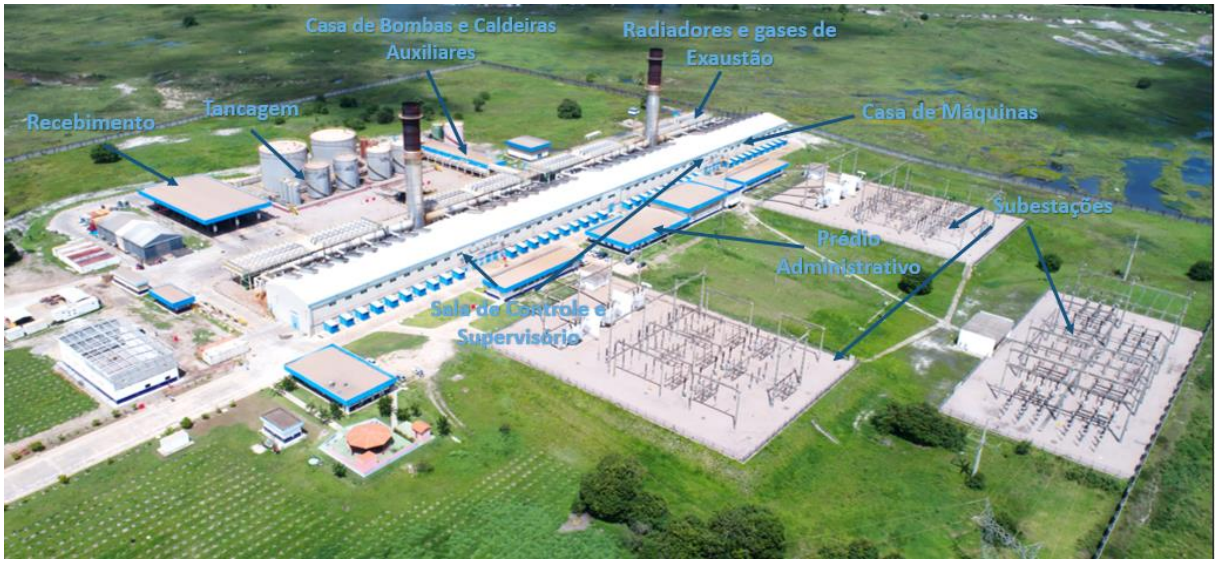


FIGURA 3 - FOTO AÉREA DA PLANTA TÉRMICA DA EPASA.

2.2 SISTEMAS AUXILIARES

Para manter os parâmetros de temperatura e pressão da unidade geradora, é necessária uma série de equipamentos que auxiliam o funcionamento normal do moto e compõem o sistema de serviços auxiliares.

Em uma usina termelétrica à combustão interna, os seguintes sistemas estão presentes em sua operação:

- Sistema de tratamento de óleo;
- Sistema de alimentação de combustível;
- Sistema de compressão de ar;
- Sistema de geração de vapor;
- Sistema de resfriamento de água;

2.2.1 Sistema de tratamento de óleo

O sistema de tratamento de óleo é composto por centrífugas que fazem a filtragem do insumo para ser utilizado no processo. Existem dois tipos de centrífugas usadas na usina: a centrífuga de óleo pesado e a centrífuga de óleo lubricante. A

centrífuga de óleo pesado tem como responsabilidade remover particulados sólidos e diminuir a presença de água no óleo combustível antes de ser utilizado no processo. Após tratamento, o óleo é armazenado em um outro tanque, chamado tanque de uso diário, contendo somente óleo tratado.

2.2.2 Sistema de alimentação de combustível

O sistema de alimentação de combustível é composto por bombas que alimentam as unidades geradoras. Esse sistema é composto por dois módulos, o módulo de bombas alimentadoras (Feeder) e o módulo de bombas elevadoras de pressão (Booster). O módulo Feeder é composto por quatro conjuntos de bombas, sendo dois conjuntos por usina, que são responsáveis por bombear o combustível do tanque de uso diário para a casa de máquinas, onde é recebido no módulo Booster. O módulo Booster é composto por bombas, tanque de mistura de combustível e um trocador de calor com válvula de controle de vazão de vapor para regular a temperatura do combustível de forma a adequar a viscosidade do combustível OCB1, sendo necessário o seu aquecimento à 140°C para alcançar o patamar de 12cSt (centistokes). Além disso, o módulo Booster é responsável por elevar a pressão da linha de combustível para entrada deste nos bicos injetores das unidades geradoras (ver Figura 4).

Cada conjunto de bombas do módulo Feeder alimenta uma casa de máquinas, ou seja, dez unidades geradoras. Já os módulos Boosters apresentam em toda a planta oito módulos, sendo quatro por usina, e cada módulo é responsável por atuar no processo de geração de cinco unidades geradoras em paralelo.

FIGURA 4 - MÓDULO BOOSTER.

2.2.3 Sistema de compressão de ar

O sistema de compressão de ar é composto por 14 compressores de ar, sendo que dez deles trabalham no nível de pressão final de 30bar e quatro trabalham no nível de pressão final de 6bar. Esse sistema tem como objetivo fornecer ar comprimido para os motores e os instrumentos pneumáticos de toda planta. A linha de ar com pressão de 30bar é utilizada nos motores para auxiliar na partida e parada das unidades geradoras. Já a linha de ar com pressão de 6bar é a linha de ar chamada de ar de instrumentação, utilizado para acionar, predominantemente, válvulas pneumáticas e outros instrumentos/equipamentos que necessitam de ar comprimido para executar sua função.

2.2.4 Sistema de geração de vapor

O sistema de geração de vapor é composto por 16 caldeiras de recuperação, sendo oito de cada usina, que estão instalados em série com a saída dos gases de exaustão de algumas unidades geradoras. Também faz parte deste sistema as duas caldeiras auxiliares localizadas na casa de bombas. As caldeiras auxiliares operam majoritariamente durante o perído em que as usinas estão desligadas, suprindo vapor para aquecimento dos tanques de combustíveis, como também podem operar como auxílio às caldeiras de recuperação durante operação da planta para normalizar a pressão de vapor na linha ocasionado por alguma anomalia. Já as caldeiras de recuperação operam somente em caso de despacho da planta, tendo como princípio de operação o uso do gás de exaustão dos motores para transformar a água em estado líquido para estado gasoso, realizando o processo de geração de vapor por cogeração.

2.2.5 Sistema de resfriamento de água

O sistema de resfriamento de água é um dos maiores sistemas de uma usina termelétrica. Normalmente, cada unidade geradora tem seu sistema de refrigeração totalmente independente, mas existem casos onde, devido ao grande número de unidades geradoras na planta, este sistema se torna compartilhado com outras unidades, como encontrado nas usinas Termoparaíba e Termonordeste.

O sistema de resfriamento de água é composto por radiadores, dispositivos utilizados comumente para realizar a troca de calor entre a atmosfera e a água através de ventiladores, no qual cada radiador estabiliza a temperatura da água de dez unidades geradoras, totalizando quatro radiadores para as 40 unidades. Cada radiador é dividido em dois lados, nomeados como radiador lado “A” e lado “B”, cada lado possui de 50 a 55 ventiladores e o circuito de água de cada lado é compartilhado com cinco unidades geradoras em paralelo. Além dos radiadores, o sistema conta com bombas e válvulas elétricas, que são responsáveis por transportar a água quente dos motores até o radiador e, logo depois da troca térmica, devolvê-la novamento para os motores, compondo um circuito fechado de refrigeração de água.

Este sistema é a maior carga própria da usina, possui na sua totalidade cerca de 430 motores elétricos com potência de 15kVA e fator de potência 0,8, ou seja, totalizando cerca de 5,16MW de potência consumida.

2.3 LÓGICA FUZZY

O ser humano tem a capacidade de tomar decisões a partir de informações imprecisas ou aproximadas. Tal estratégia geralmente permite a expressão da decisão através de termos linguísticos. A partir da teoria de conjuntos fuzzy e conceitos de lógica fuzzy, é possível converter a linguagem natural imprecisa em informações no formato numérico e de regras, de simples manipulação (TANSCHEIT, 2004). Devido a isso, tais conceitos vêm sendo amplamente utilizados na conversão do conhecimento dos especialistas de sistemas (plantas) para maximizar a operação de processos industriais ou controladores.

A definição de conjuntos fuzzy foi apresentada pela primeira vez, em 1965, por Lofti Asker Zadeh, como uma forma matemática de representar termos linguísticos.

A teoria de lógica fuzzy é baseado no princípio de que o pensamento humano é estruturado em classes cuja transição entre pertencer ou não a um determinado conjunto é gradual e não abrupta. Assim, a lógica fuzzy nasceu a partir da modificação da lógica tradicional a fim de se adaptar conforme princípio humano. Tal lógica realiza operações sobre conjuntos denominados fuzzy, cujos elementos pertencentes a esse conjunto são caracterizados a partir de um grau de pertinência no intervalo de [0,1]. O grau de pertinência ou função de pertinência é a função que descreve a característica de pertinência do elemento e tal função pode assumir valores reais não negativos entre os extremos 0, elemento não pertence ao conjunto, e 1, elemento pertence ao conjunto (SARAIVA, 2006).

Um conjunto fuzzy pode ser denotado por um conjunto de pares ordenados onde o primeiro elemento do par é o elemento do universo em questão e o segundo é o grau de pertinência (SARAIVA, 2006). Pode-se definir o conjunto Universo como um conjunto onde pertencem todos os valores possíveis do elemento.

Outro conceito importante é o conceito do conjunto suporte, o conjunto suporte é o subconjunto que contém todos os elementos do conjunto fuzzy que apresentam o grau de pertinência maior que zero.

Exemplificando os conceitos acima, dado uma variável T, um conjunto representando a temperatura de um ambiente, pode ser definido por T = {x_1, ... ,x_N}, onde x_i é um elemento de T. Assim, um conjunto fuzzy U no universo de T pode ser definido pelo seguinte par:

$$U = \{ (x, \mu_U\ (x)) \mid x \in T \},$$

onde: U é o conjunto fuzzy, x é o elemento do conjunto e $\mu_U(x)$ é a função de pertinência do elemento x em U (LEE, 1996). A Figura X ilustra a representação de um conjunto fuzzy, especificando o suporte e o universo do conjunto.

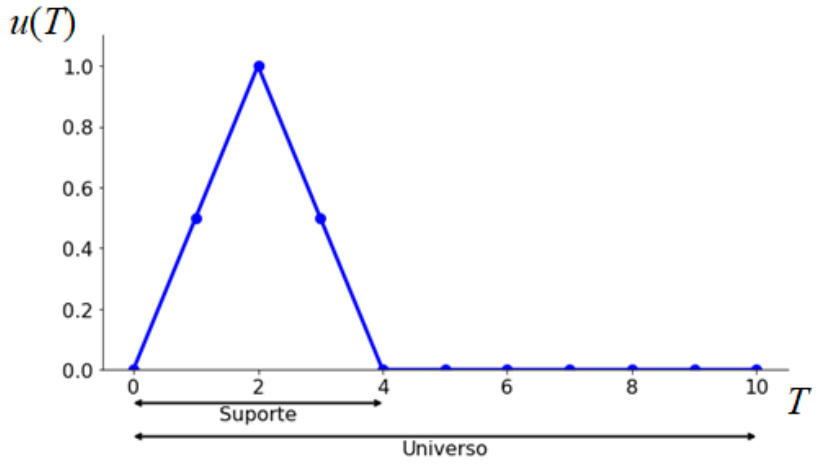


FIGURA 5 - REPRESENTAÇÃO DE UM CONJUNTO FUZZY.

Os sistemas fuzzy vêm sendo utilizados atualmente em diversas aplicações, tais como classificadores, modelos de previsão de séries temporais, planejamento e otimização, muitas vezes de forma híbrida com redes neurais e programação evolutiva.

O uso de sistemas que utilizam lógica fuzzy neste trabalho se dá a possibilidade de utilizar um banco de conhecimento que pode ser criado a partir de poucas amostras de dados, tendo em vista que, para treinar uma rede neural capaz de prever o comportamento de algum processo é necessário uma grande quantidade de dados, caso este nem sempre possível de se atingir a curto/médio prazo.

2.4 SISTEMA DE INFERÊNCIA FUZZY

Sistemas de inferência fuzzy são sistemas que utilizam regras linguísticas para realizar um processo de tomada de decisão. Tais sistemas são baseados em regras do tipo “se...então” e são compostos por quatro blocos principais:

- Bloco de Fuzzificação;
- Motor de inferência;
- Banco de regras fuzzy;
- Bloco de defuzzificação;

A Figura 6 apresenta o esquema básico de um sistema de inferência.

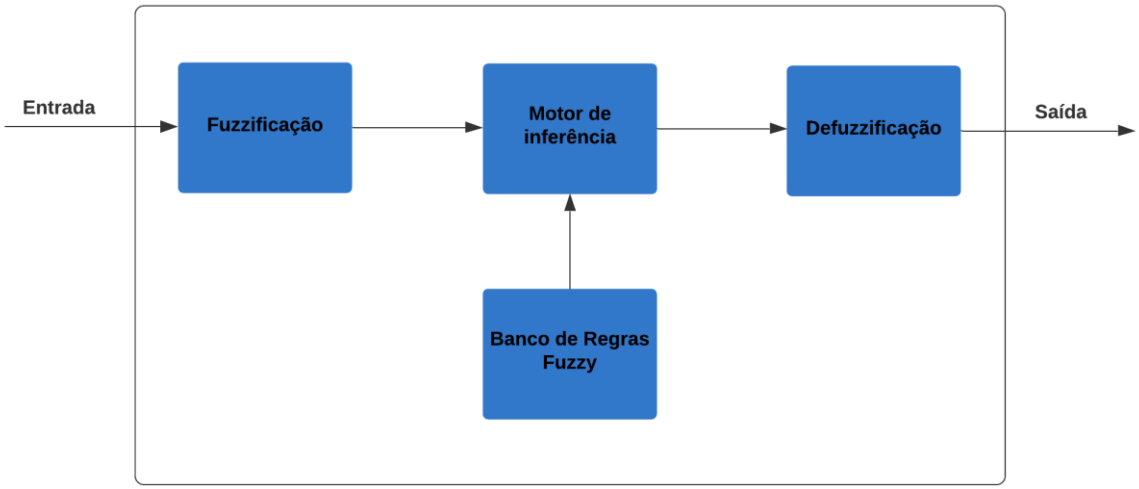


FIGURA 6 - ESQUEMA DE UM SISTEMA DE INFERÊNCIA FUZZY

Fuzzificação - responsável por converter a entrada analógica ou discreta não fuzzy proveniente de medições para conjuntos fuzzy. Neste estágio, deverão ser mapeadas os conjuntos fuzzy para os diversos intervalos em termos linguísticos. É permitido o uso de qualquer tipo de função (triangular, trapezoidal ou outras) como também a mesclagem dessas funções. O importante é mapear os termos de forma adequada.

Banco de regras fuzzy - responsável por armazenar o conhecimento do sistema à base de proposições, conhecidas como regras do sistema. As regras podem ser criadas a partir do conhecimento de um especialista, em forma de sentenças linguisticas, com o intuito de replicar para o sistema o resultado de uma tomada de

decisão que teria sido feito pelo especialista em determinadas situações. A boa composição do banco de regras é fundamental para se ter um bom desempenho do sistema.

As proposições podem ser das seguintes formas:

- Proposições condicionais:
  "Se **temperatura** é **alta** então **ar condicionado** é **potência alta**".
  "Se **temperatura** é **baixa** então **ar condicionado** é **potência baixa".**
- Proposições não condicionais:
  "**Ar condicionado** é **potência média**".

Existem outras formas de definir regras além da extração a partir do conhecimento do especialista, e existem métodos de extração de regras a partir dos dados históricos armazenadas em banco de dados. Na seção posterior, será discutido esse tópico.

Motor de inferência - responsável por realizar as operações entre os conjuntos fuzzy de entrada/saída a partir do conhecimento das regras, de forma a interpretar os valores de entrada para associar os valores de saída do sistema.

Bloco de defuzzificação - responsável por interpretar os resultados obtidos pelo motor de inferência. Para se utilizar na planta do processo, é necessário que a saída seja convertida para valores precisos (analógicos ou discretos). Existem vários métodos de defuzzificação bem difundidos na literatura, os mais comuns são o cálculo do centroide da área do resultado do motor de inferência e o cálculo da média dos máximos.

2.5 EXTRAÇÃO AUTOMÁTICA DE REGRAS

O banco de regra do sistema fuzzy é o bloco de maior impacto no desempenho do sistema, isso devido ao fato de que todo o conhecimento do sistema está contido nele. Quando se tem plantas complexas ou fortemente não lineares, a construção do banco de conhecimento fuzzy se torna mais difícil e trabalhosa, tornando, às vezes, inviável a utilização de um sistema fuzzy para realização do controle (WANG e MENDEL, 1992).

A partir dessa motivação, foi criada uma abordagem capaz de extrair regras automáticas de dados históricos dos sensores desenvolvido por Wang-Mendel (1992).

Essa abordagem Wang-Mendel apresenta cinco etapas gerais para extrair regras fuzzy de dados numéricos. Os passos são:

1° passo - Dividir as entradas e saídas em regiões: Esta etapa deverá realizar a divisão do domínio das entradas e saídas em regiões fuzzificadas. Cada variável deverá ser dividida em 2N+1 regiões. Não é necessário que todas as entradas tenham o mesmo N, e o formato de cada função de pertiência poderá ser de qualquer forma. Cada região deverá ser classificada da seguinte forma: PN (Pequeno N), ..., P1 (Pequeno 1), Centro, G1 (Grande 1), ..., Grande (N). Para exemplificar, dado um sistema genérico com duas entradas (x1 e x2) e uma saída (y), assume-se N = 2 para a x1, x2 e y e dado os nomes das regiões de Pequeno 1, Centro e Grande 1 em todas as variáveis. A Figura 7 apresenta o resultado desse mapeamento.

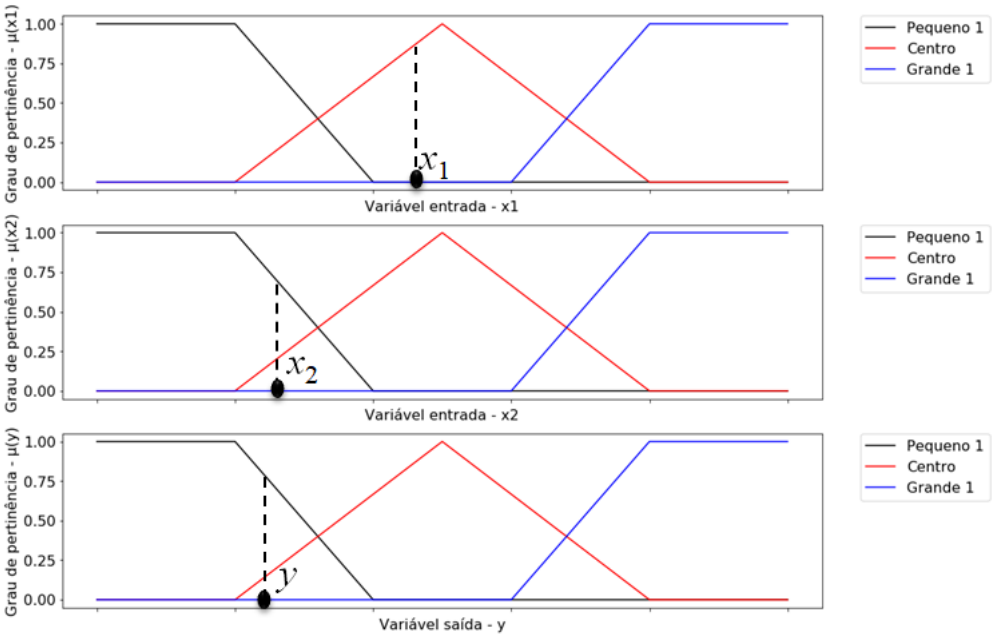


FIGURA 7 – MAPEAMENTO DAS ENTRADAS E SAÍDAS DE UM SISTEMA GENÉRICO.

2° passo - Gerar regras a partir das entradas e saídas dos dados históricos: Nesta etapa ocorrerá a criação das regras a partir dos dados de acordo com os conjuntos fuzzy criados na etapa anterior. Utilizando o exemplo do sistema genérico acima, se para um determinado par de dados onde x1 assume a região “Centro”, x2 a região “Pequeno 1” e y a região “Pequeno 1”, pode-se extrair a seguinte regra:

$$Se\ \boldsymbol{x1}\ \text{é}\ \boldsymbol{Centro}\ e\ \boldsymbol{x2}\ \text{é}\ \boldsymbol{Pequeno\ 1}\ então\ \boldsymbol{y}\ \text{é}\ \boldsymbol{Pequeno\ 1}.$$

Neste caso, as regras foram criadas a partir somente do operador “AND” mas é possível utilizar outros operadores para criar regras mais adequadas para o processo. Deverá ser feito este passo para todo o banco de dados que será usado no processo de aprendizagem.

3° passo – Atribuir um peso para as regras criadas: Após gerar regras a partir do banco de dados de aprendizagem, será necessário atribuir pesos para as regras, pois em alguns processos podem existir regras onde os antecedentes sejam iguais, mas os consequentes das regras sejam diferentes, o que pode trazer conflito durante a inferência das regras (WANG e MENDEL, 1992). Uma das formas de resolver esse tipo de conflito é atribuir peso para as regras criadas a partir de dados mais expressivos. Wang propõe utilizar o grau de pertinência das entradas e saídas, multiplicá-los entre si, se tornando o peso da regra, como mostrado a equação a seguir para o sistema mostrado no 2° passo:

$$Peso(regra) = \mu(x1) * \mu(x2) * \mu(y)$$

4° passo – Montar um banco de dados de regras: Neste passo deverá ser tratado o banco de dados extraído dos dados. As regras em duplicidade deverão ser eliminadas, e as regras ao qual possuem conflitos deverão ser selecionadas com base no maior peso entre as regras em conflito.

5° passo – Determinar um mapeamento baseado nas regras combinadas: A etapa final deverá determinar a estratégia de defuzzificação do sistema, o qual deverá converter as entradas atuais, combinado com o banco de regras extraído para encontrar a saída atual. Wang-Mendel chama esse 5° passo em “mapeamento” pois se considerar os cinco passos como uma “caixa preta”, pode-se dizer que a função desse bloco é mapear o espaço das entradas no espaço da saída.

Antonelli (2016) propõe uma melhoria no método de extração de regras a partir da otimização da seleção de variáveis. É proposta uma nova abordagem de seleção de variáveis baseada no critério de mínimo-redundância-máxima-relevância. A nova abordagem tem como principais objetivos o aumento na velocidade do treinamento e aumento na interpretabilidade do modelo final.

2.6 ESTRATÉGIAS EVOLUTIVAS

A computação evolutiva foi criada inspirada na evolução natural, normalmente utilizada em sistemas para resolução de problemas complexos a partir da otimização de diversos parâmetros na busca das melhores soluções. Surgiu nos anos 1950, a partir dos trabalhos de Bremermann, Friedberg, Box e outros, e até hoje continua sendo utilizada em problemas nas diversas áreas de estudo (BACK, 1997).

Algoritmos evolutivos são métodos de busca estocástica inspirados na evolução a partir da seleção natural e normalmente são compostos por procedimentos padrões que podem ser usados em uma ampla gama de problemas e em diversa áreas do conhecimento (VILLANUEVA, 2015).

A Figura 8 exibe a estrutura básica padrão utilizados pelos diversos algoritmos evolutivos, basicamente apresenta o seguinte funcionamento:

- Criação da primeira geração de indivíduos: É criada a geração de indivíduos de forma aleatória onde cada indivíduo da população é uma possível solução do problema.
- Avaliação da solução de cada indivíduo: Cada indivíduo da geração é testado a parte de uma função objetiva comumente chamada de função fitness. O fitness pode ser entendido como a qualidade da solução entregue por aquele indivíduo.
- Verificar o critério de otimização: Esta etapa consiste em verificar a qualidade do fitness de todos os indivíduos da geração. A verificação é feita a partir do critério de otimização, podendo ser escolhida a partir medidores de erro ou idade da geração. Caso o critério seja atendido, o algoritmo entenderá que encontrou a melhor solução para o problema caso contrário uma nova geração será criada a partir dos passos de seleção, cruzamento e mutação para que seja feita uma nova avaliação do fitness.
- Seleção: O primeiro passo na criação da nova geração é selecionar os indivíduos, baseado na qualidade da solução e da geração atual, para que possam sofrer o processo de evolução nos próximos passos. A seleção consiste em escolher indivíduos de forma probabilística, porém, indivíduos com melhores desempenhos do fitness têm maior possibilidade de serem escolhidos do que os indivíduos com os piores desempenho do fitness.

- Cruzamento ou recombinação: Na etapa do cruzamento é criado um indivíduo a partir da combinação de características de dois indivíduos selecionados, indivíduos estes escolhidos de forma aleatória.
- Mutação: A mutação é uma operação que causa a variação de uma característica do indivíduo filho (após o cruzamento, ou, no caso de métodos que não apresentam a etapa de mutação, após a seleção). Esta operação ocorre de forma estocástica podendo ser mais frequente (em casos de algoritmos onde a única forma de evolução é a mutação) ou menos frequente. Após esta etapa, será criada a nova geração, retornando ao loop de avaliação da solução de cada novo indivíduo.
- Melhores indivíduos encontrados: Quando o critério de otimização é atendido, a solução otimizada é encontrada e, assim, o algoritmo é finalizado com tal solução. Porém, muitas vezes, não é possível alcançar o critério de otimização em problemas do mundo real, desta forma, é comum adotar critérios de paradas diferentes do critério de otimização tais como máximo de gerações ou quando o valor do fitness se mantém constante por um certo número de gerações.

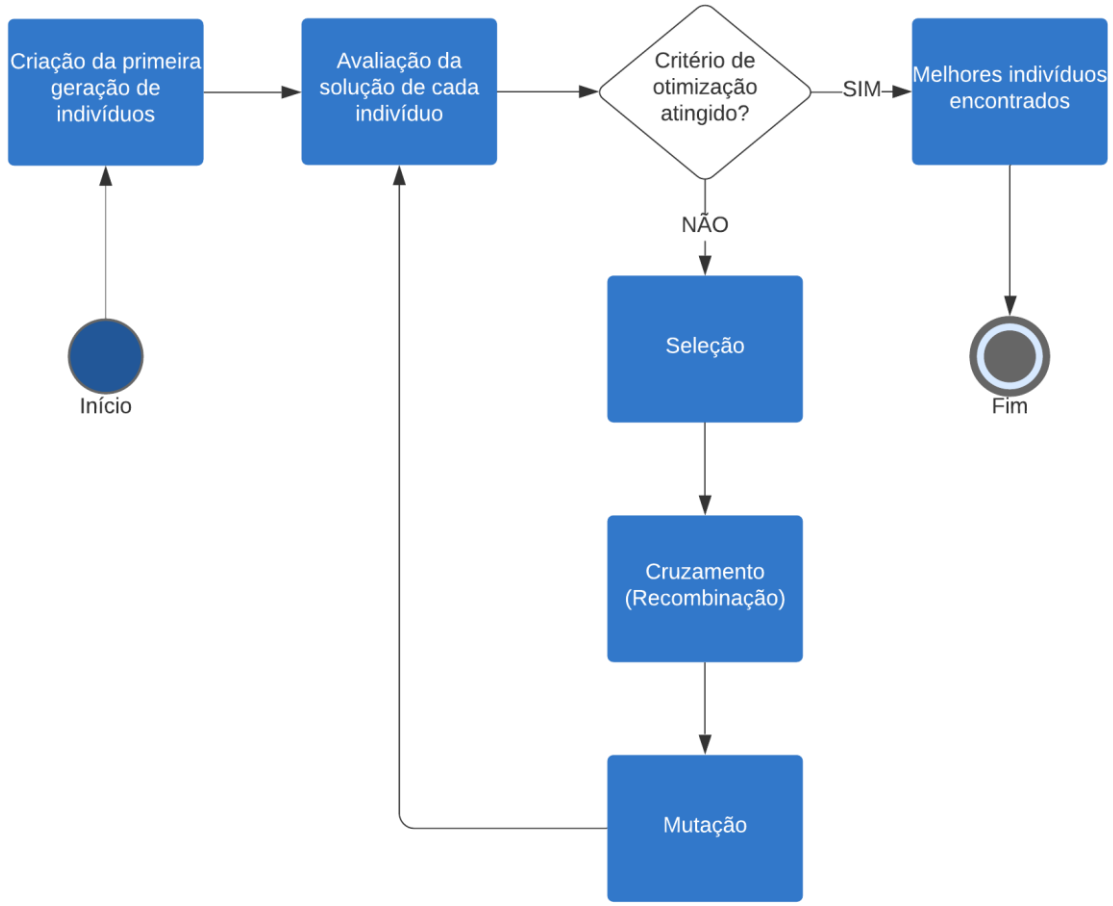


FIGURA 8 - ESTRUTURA BÁSICA DE UM ALGORITMO EVOLUTIVO.

A maioria dos algoritmos evolutivos descende de três abordagens principais: programação evolutiva, estratégias evolutivas e algoritmos genéticos. Neste trabalho, será mostrado o uso do algoritmo de estratégias evolutivas.

A estratégia evolutiva (EE) foi desenvolvida por volta dos anos 1960, mesma época no qual apareceu a programação evolutiva, também desenvolvida por Rechenberg e Schwefel. Os primeiros algoritmos de estratégia evolutiva foram chamados de estratégia evolutiva (1+1) e (1,1). Ainda não havia um conceito de população, ou seja, só havia um indivíduo pai e a mutação era o único operador de evolução. Na estratégia (1+1), após mutação ocorre a seleção do melhor indivíduo (competição entre pai e filho) e o vencedor permanece para a próxima geração. Na estratégia (1,1), após mutação não há seleção, ou seja, o pai é eliminado permanecendo somente o filho para próxima geração. A mutação é feita a partir do pai da seguinte forma:

Seja $v_i = (x_I, \sigma)$, onde v_i é o vetor que representa o indivíduo da geração i, x_i é o ponto no espaço de busca da geração i e σ o desvio padrão utilizado para atualização de x. A mutação é realizada atualizando x da seguinte forma:

$$x_{i+1} = x_i + N(0, \sigma)$$

Onde $N(0, \sigma)$ significa o operador de mutação chamado de mutação gaussiana (vetor composto por número gerado aleatoriamente por uma função gaussiana de média zero e desvio σ).

Uma das principais descobertas durante pesquisas na estratégia evolutiva foi a criação do mecanismo chamado “regra de sucesso 1/5” que atualizava o desvio durante as gerações. A regra consiste na razão entre as mutações que geram um filho melhor que o pai (φ), chamadas de mutações positivas, em relação a todas deve ser 1/5. Se φ for menor que 1/5 o desvio deve ser diminuído, caso seja maior, aumentado.

Com o tempo, o algoritmo de estratégia evolutiva foi sendo aprimorado de forma a permitir cruzamento discreto (seleciona-se dois pais e, randomicamente, com chances iguais, escolhe-se somente um pai para o filho herdar todas as características) ou cruzamento intermediário (seleciona-se dois pais e o filho herda a média da soma dos dois pais). A etapa de seleção foi aprimorada, tornando-se elitista determinística (μ,λ) ou (μ+λ), ou seja, população com mais de um indivíduo e mais de uma descendência e característica de autoadaptação dos parâmetros da estratégia.

O algoritmo genético (AG) é o algoritmo evolutivo mais conhecido (EIBEN & SMITH, 2015). Criado inicialmente por Holland, o AG foi inspirado pela genética utilizando-se da estrutura genética para representar, analogamente, a estrutura do algoritmo, na qual a população é conhecida como cromossomo que contém as entradas (genótipo) de uma possível solução do problema (fenótipo) (VILLANUEVA, 2015).

Um AG simples possui as seguintes características:

- Indivíduos representados como estrutura, lista de valores ou vetores;
- A seleção é feita a partir do algoritmo da roleta, onde, conceitualmente, os espaços da roleta representam cada cromossomo daquela geração e o tamanho do espaço é a probabilidade de ser escolhido. A probabilidade é afetada a partir da qualidade do fitness do cromossomo, podendo ser maior caso a avaliação do cromossomo tenha sido alta ou menor caso tenha sido baixa.

- Utiliza o cruzamento de um ponto do genótipo, ou seja, os pais escolhidos trocam parte do seu genótipo para formar novos cromossomos;
- Mutação binária a partir de *bit-flipping* (inversão do bit) em um número inteiro a partir da probabilidade p_m;
- No final da geração, toda a população é substituída pela nova geração criada, apesar de que é possível um indivíduo passar de uma geração para outra sem que tenha mutação ou cruzado tendo em vista que cada passo é estocástico;

Apesar de possuir uma característica probabilística, o AG não é um algoritmo puramente aleatório, ele possui mecanismos que ajudam a direcionar a procura, tal como o método da seleção da roleta, que é amplamente utilizado. Esse método consiste em selecionar μ pais de λ membros de uma população. Para isso, é calculada uma função distribuição cumulativa de probabilidade de todos os possíveis pais, utilizando a qualidade do fitness de cada indivíduo, na qual a probabilidade de ser selecionado de cada indivíduo é dada por:

$$\mathrm{p}_{\mathrm{sel}}(i) = \frac{fitness(i)}{\sum_{t=1}^{\lambda} fitness(t)} \qquad (1)$$

E a função distribuição cumulativa de probabilidade (a_λ) é dada por:

$$a_\lambda = \sum_{x=1}^{\lambda} p_{sel}(x) = 1 \qquad (2)$$

É possível que o mesmo indivíduo seja selecionado mais de uma vez. Abaixo segue um exemplo de pseudocódigo do algoritmo de roleta:

TABELA 1 - PSEUDO-CÓDIGO DO ALGORITMO DE SELEÇÃO DA ROLETA.

```
Programa Roleta
Início
        t ← 0
        μ = número de pais selecionados
        b = soma dos fitness de todos os indivíduos
        /* cria vetor com a probabilidade de cada indivíduo ser escolhido
        rank (i) = cada elemento do vetor fitness(i) / b
        enquanto iteração <= μ fazer
        início
                rand = número randômico gerado nessa iteração (0,1)
                x = 1
                enquanto a_λ(x) < rand fazer (onde a_λ(x) = ∑_{x=1}^{i} p_sel(x))
                        início
                        x = x + 1
                        fim
                seleciona indivíduo x
                iteração = iteração +1
        fim
fim
```

O cruzamento é feito em pares, onde cada indivíduo vai compartilhar parte do seu alelo com o outro formando novos indivíduos. Para alelos com representação em números reais é utilizado o cruzamento aritmético, que é o caso deste trabalho.

No cruzamento aritmético de um ponto os alelos de dois indivíduos diferentes sofrem uma operação de soma e logo após é dividido pelo número de indivíduos para se obter a média, essa operação ocorre somente a partir de uma posição k de cada individuo onde antes da posição k os alelos se mantém e após a posição k os alelos são substituídos pela a média da soma. A Figura 9 exemplifica a operação de cruzamento aritmético de um ponto.

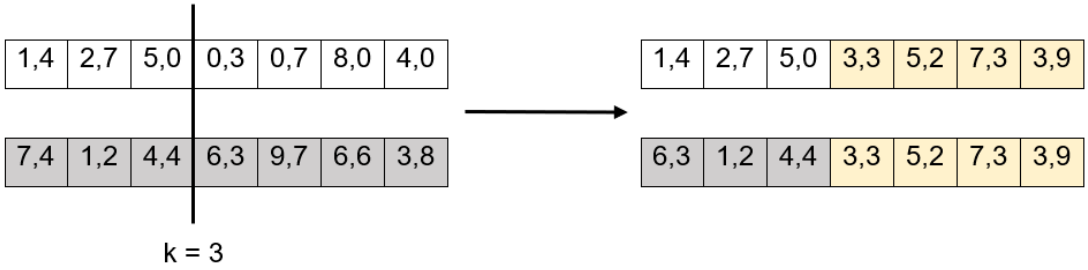


FIGURA 9 - CRUZAMENTO EM UM PONTO K=3 EM UM VETOR DE ALELOS REAIS.

A mutação ocorre de forma probabilística nos novos indivíduos após o cruzamento. As mutações mais simples utilizadas em AGs são a mutação de inversão de bits e mutação creep. A mutação de inversão de bits consiste em, quando os alelos do cromossomo são representados por binários, inverter aleatoriamente o bit de cada alelo com probabilidade de ocorrer p_m. A mutação creep é utilizado quando os alelos são representados por números inteiros ou reais e consiste em adicionar um pequeno valor, negativo ou positivo, a cada alelo com probabilidade de ocorrer p_m. A Figura 10 exemplifica a duas operações.

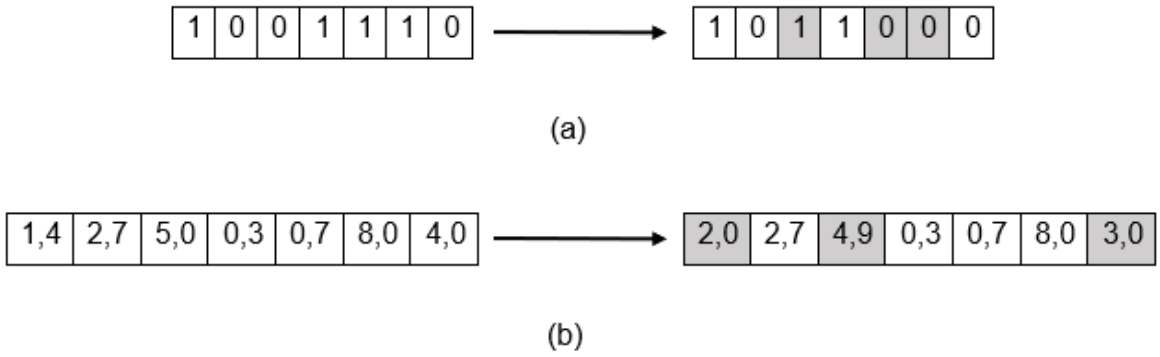


FIGURA 10 - (A) OPERAÇÃO DE MUTAÇÃO INVERSÃO DE BIT E (B) MUTAÇÃO CREEP.

Atualmente, já existem diversos outros métodos de cruzamento e mutação aplicáveis à AGs, porém, os métodos mais simples ainda são muito utilizados nos mais diversos otimizadores de problemas.

3 PROPOSTA DO TRABALHO

3 PROPOSTA DO TRABALHO

Neste capítulo, será apresentada a abordagem utilizada para criação de um *digital twins* do sistema de refrigeração de uma usina termelétrica a partir dos dados das usinas Termoparaíba e Termonordeste. É utilizado novo método de aprendizagem do modelo, seguindo os conceitos fundamentais de *digital twins*, em que o modelo estará em constante atualização quando deparado com situações nunca vistas no treinamento. Além disso, conta com um método de seleção de variáveis utilizando Análise de Componentes Principais a fim de otimizar o esforço computacional do modelo, como também, manter a alta interpretabilidade do modelo mesmo após redução de variável. É utilizado um algoritmo genético a fim de otimizar o modelo, visando otimizar o erro médio e o esforço computacional do modelo, buscando novas formas de conjuntos fuzzy nas variáveis de entrada e saída do modelo.

3.1 SISTEMA DE REFRIGERAÇÃO DO MOTOR

A planta termelétrica estudada neste trabalho apresenta um sistema bem particular, ela utiliza 55 ventiladores para compor um radiador que é utilizado para o controle da temperatura da água de refrigeração de cinco motores simultaneamente, como se ilustra na Figura 11. Além disso, os ventiladores são acionados usando uma configuração de partida direta, ou seja, não há a possibilidade de variação de velocidade dos ventiladores, dificultando o controle fino de temperatura do processo.

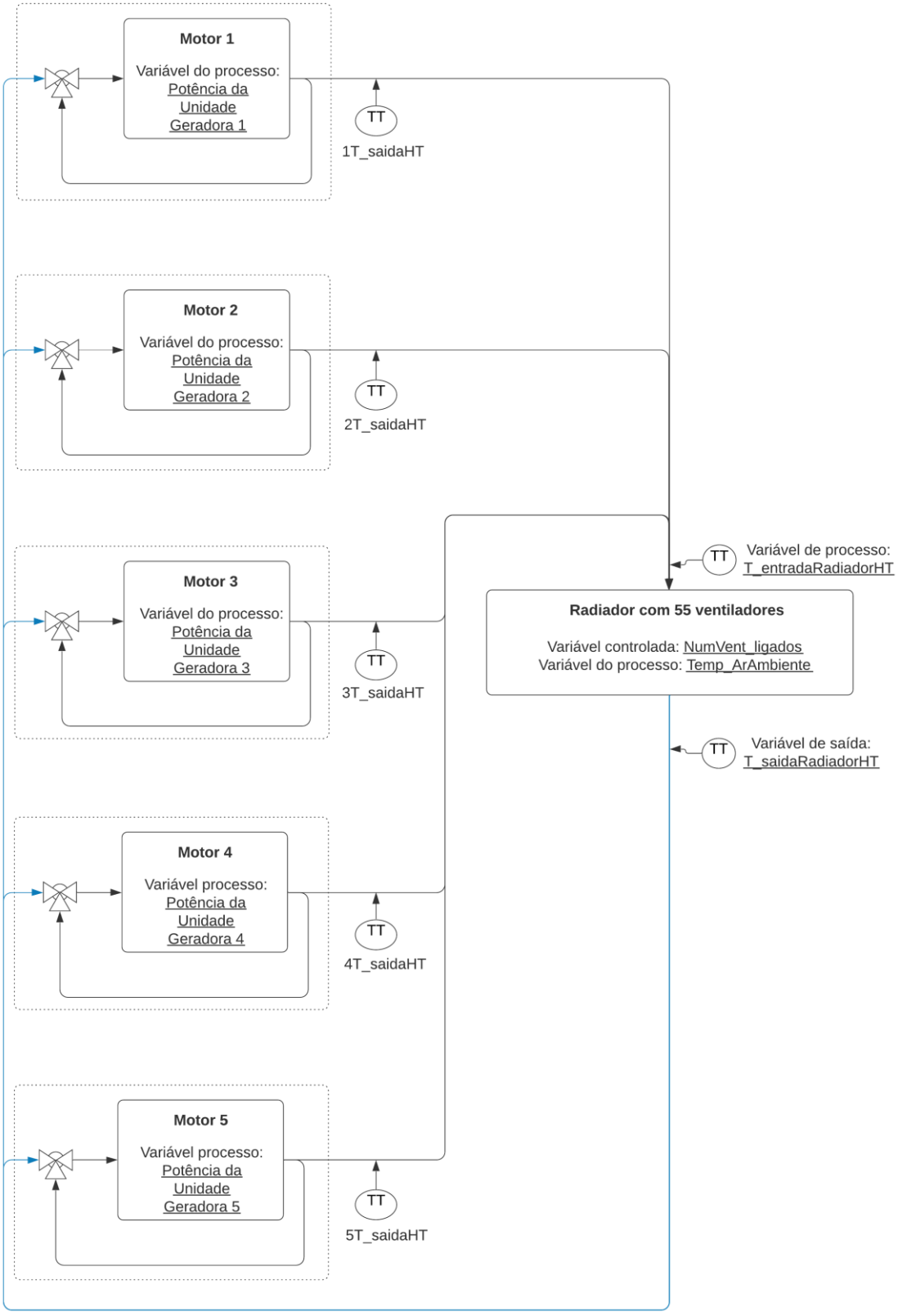


FIGURA 11 - SISTEMA DE REFRIGERAÇÃO DE ÁGUA PARA UM LADO DO RADIADOR.

O compartilhamento desse sistema torna o processo não-linear, tendo em vista que os motores operaram em potências diferentes, apresentam problemas diferentes que impactam na sua temperatura, além da possibilidade de vazamentos no circuito e, consequentemente, entrada de ar que prejudica a troca de calor entre o radiador e a água. Com isso, o uso de controladores comumente utilizado na indústria, tais como controladores do tipo proporcional (P) e proporcional-integral (PI), mostram-se ineficientes no controle de tal processo, isso devido à característica não linear, sendo necessária a sintonização dos parâmetros do controlador com certa constância, o que é normalmente inviável para a equipe de manutenção.

Para o controle da temperatura da água, atualmente é utilizado um controlador proporcional-integral para cada motor, o qual fornece como resultado uma temperatura controlada T_entHT, como se ilustra na Figura 12. Nesta figura, observa-se que a variável manipulada é o ângulo da válvula de três vias, cuja atuação leva a variável controlada à temperatura de *set-point*. A definição do ângulo da válvula definirá a transferência de energia (em forma de calor) entre a temperatura da água fornecida pelo radiador e a temperatura da água proveniente da saída do motor (T_saidaHT).

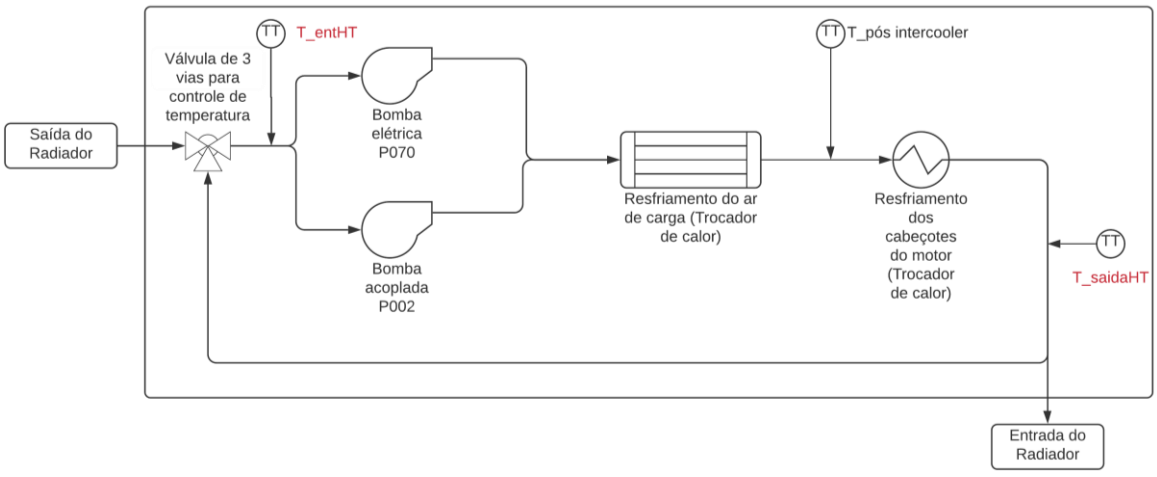


FIGURA 12 - CIRCUITO DE REFRIGERAÇÃO DE UM MOTOR.

A estratégia de controle apresenta algumas limitações, tais como: a) não se garante o erro zero em regime estacionário quando os parâmetros não estão

sintonizados; b) as não linearidades do sistemas levam a uma perda da sintonização constante do controlador; c) o controle da temperatura apresenta um baixo desempenho devido aos efeitos externos, principalmente nos períodos de chuva e à noite, pois a diminuição da temperatura ambiente do ar não é considerada para diminuir o número de ventiladores ligados (usados pelo radiador), levando o sistema de refrigeração da água a um alto consumo de energia.

Deste modo, é apontada a necessidade de desenvolver um modelo baseado em dados, que incorpore sinergicamente em um único modelo, as variáveis de entrada:

- Temperatura da água de alta temperatura (HT) da entrada no radiador (T_entHT);
- Temperatura Ambiente (Tamb);
- Número de ventiladores ligados (NumVentOn);
- Potência da Unidade Geradora 1 (P1);
- Potência da Unidade Geradora 2 (P2);
- Potência da Unidade Geradora 3 (P3);
- Potência da Unidade Geradora 4 (P4);
- Potência da Unidade Geradora 5 (P5);

E a variável de saída:

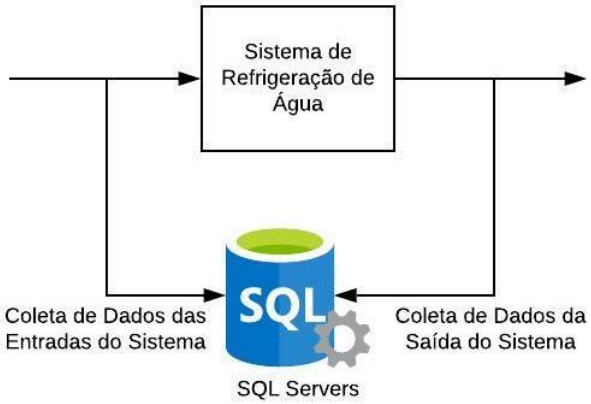
- Temperatura da água de alta temperatura (HT) de saída no radiador (T_saidaHT).

3.2 METODOLOGIA PARA MODELO DA PLANTA USANDO EXTRAÇÃO DE REGRAS AUTOMÁTICA

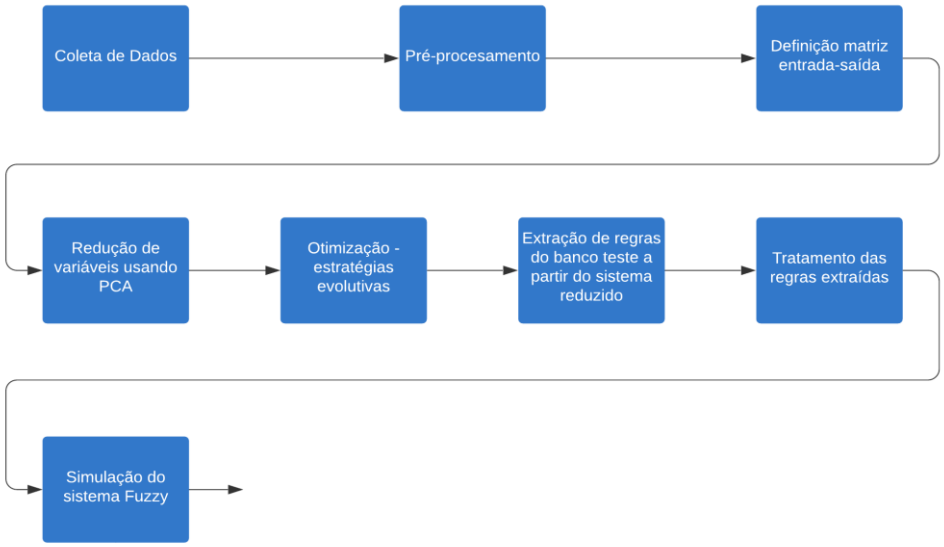
A abordagem de extração automática de regras fuzzy foi escolhido devido à possibilidade de extrair regras com características do sistema a partir de um banco de dados de treinamento com período relativamente pequeno (dois dias de dados com amostragem de um segundo), diferentemente de abordagens utilizando redes neurais, nas quais, muitas vezes, se faz necessário um banco de treinamento maior para se extrairem características do sistema. Outro ponto é a facilidade de se extrair novas regras e atualizar o sistema fuzzy a partir de situações desconhecidas do conjunto de

regras ou caso o erro do modelo ultrapasse patamares indesejados. A Figura 13 mostra o diagrama para modelagem do sistema de forma sequencial, cujos módulos serão explicados logo em seguida.

A)



B)



C)

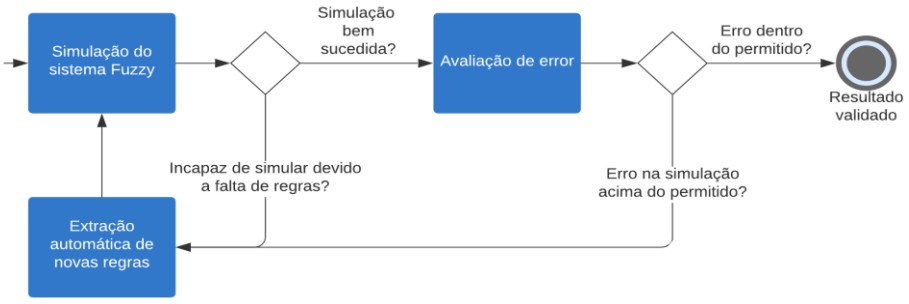


FIGURA 13 – A) COLETA DE DADOS, B) CONSTRUÇÃO DO MODELO INICIAL, C) ATUALIZAÇÃO DO MODELO DURANTE A OPERAÇÃO

3.2.1 Coleta de dados

Devido à alta intermitência no despacho de uma usina termelétrica, as medições armazenadas no banco de dados contêm uma grande quantidade de dados que caracterizam a planta desligada, entretanto, muitas destas informações não apresentam utilidade para o modelo. Assim, buscou-se no banco de dados intervalos de operação com maior tempo de despacho ininterrupto da planta no ano de 2019, para que contivesse o mínimo de informações nos dados de planta desligada.

Foi extraído dados com 1.123.200 amostras das variáveis T_entHT, Tamb, NumVentOn, P1, P2, P3, P4, P5 e T_saidaHT do banco de dados da planta, dados estes equivalentes ao despacho ocorrido entre 14 de maio de 2019 a 25 de maio de 2019, com tempo de amostragem de um segundo para todas as variáveis.

3.2.2 Pré-processamento de dados

Para garantir que não sejam extraídas características incoerentes para as regras do sistema, foi necessário tratar os dados brutos a fim de eliminar os dados espúrios, que podem ser gerados a partir da falha de conexão entre sensor e controlador, rompimento do cabo do sensor ou curto-circuito na malha de alimentação dos sensores.

Foi utilizado um algoritmo para detectar os dados considerados outliers a partir da média-móvel de uma janela dos dados e substituir esses pontos pelos valores limites considerados não espúrios. A Figura 14 mostra o resultado obtido após o filtro dos dados espúrios na variável T_entHT. No geral, a massa de dados extraída do banco de dados da planta possuía poucos *outliers*, pois os dados de todos os sensores são armazenados no banco a partir do software PIMS (Plant Information Management Systems, em inglês) que possui um algoritmo de armazenamento que trata os dados a fim de gravá-los da forma mais eficiente tanto em tamanho de disco rígido quanto em quantidade de espúrios.

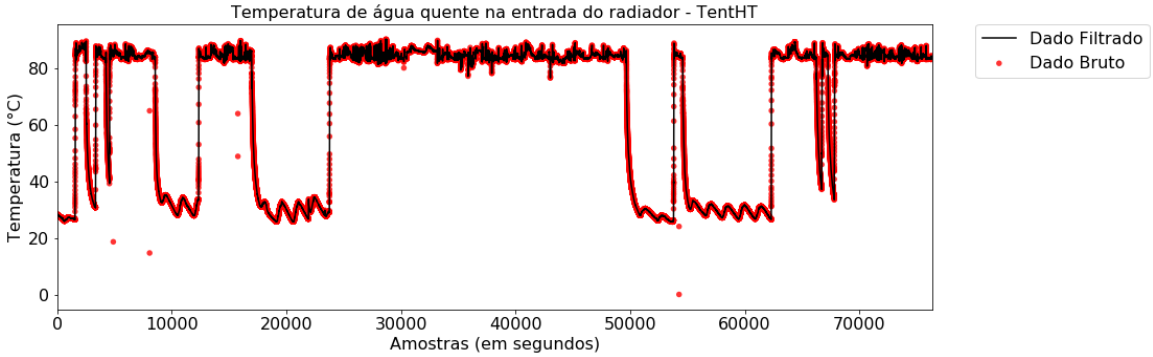


FIGURA 14 - VISUALIZAÇÃO DE DADOS APÓS REMOÇÃO DE OUTLIERS - VARIÁVEL TEMPERATURA DE ENTRADA DE HT

Mesmo extraindo-se o período com maior tempo de despacho contínuo da planta, os dados coletados continham características de quatro estados diferentes da planta: usina desligada, usina em partida, usina em rampa (rampa de subida ou rampa de descida) e usina em regime permanente. A Figura 15 ilustra esses quatro estados a partir dos dados em um momento de início do despacho. Os dados relativos à usina parada não agregam valor ao modelo, tendo em vista que não há a necessidade de controlar a temperatura do sistema. Os dados relativos à usina em partida e em rampa contém características relacionado ao transitório da planta, na qual o controle da temperatura é mais complexo devido à grande variação de temperatura no sistema, cuja característica não-linear é mais expressiva. Os dados relativos à usina em regime permanente é o estado da usina no qual as variáveis de entrada e saída estão mais estáveis e o sistema apresenta uma atenuação na sua característica não-linear.

A última etapa do pré-processamento dos dados é aplicar uma normalização nos dados das entradas, de forma que a média dos dados normalizados seja a mais próximo de zero e o desvio padrão o mais próximo de um, para que estejam prontos para aplicar a análise de componentes principais. A normalização foi realizada seguindo a equação:

$$z = \frac{x_i - \mu_x}{\sigma_x} \qquad (3)$$

onde μ_x e σ_x são o valor médio e desvio padrão da variável x, cujos valores estão armazenados no banco de dados de treinamento.

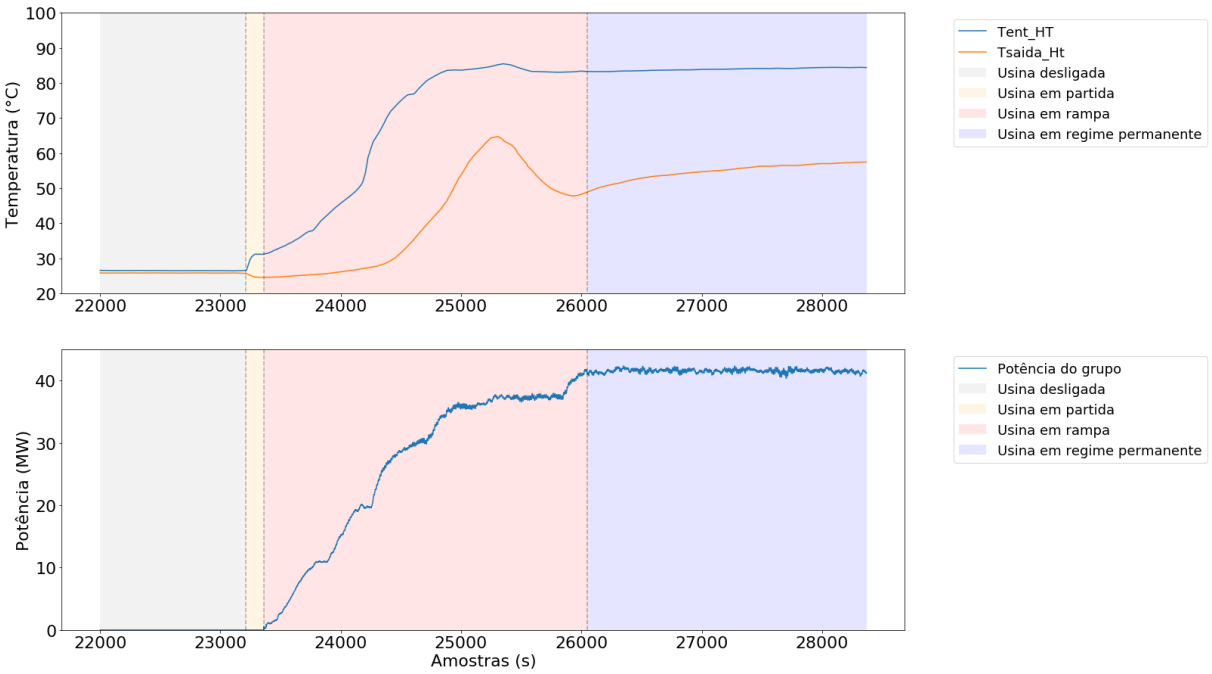


FIGURA 15 - TEMPERATURA DE ENTRADA E SAIDA NO RADIADOR DURANTE PRÉ E PÓS PARTIDA DA PLANTA.

3.2.3 Seleção das Variáveis usando PCA

Foram analisadas as principais variáveis que contribuem com o comportamento do radiador, a partir do conhecimento dos especialistas da planta, técnicos e engenheiros que estão em contato diariamente com a operação da planta e dos equipamentos, como também da malha de controle existente na planta, a fim de definir a melhor configuração de entradas e saídas para o modelo. Assim, chegou-se à configuração de oito entradas: temperatura da água de entrada no radiador lado alta temperatura - T_entHT, temperatura Ambiente - Tamb, número de ventiladores ligados - NumVentOn, potência da unidade geradora 1 - P1, potência da unidade geradora 2 - P2, potência da unidade geradora 3 - P3, potência da unidade geradora 4 - P4 e potência da unidade geradora 5 - P5, e 1 saída: temperatura da água de saída no radiador lado alta temperatura - T_saidaHT, conforme é ilustrado na Figura 16.

FIGURA 16 - MODELO INICIAL PROPOSTO DE DIGITAL TWIN DO RADIADOR.

Devido ao grande número de variáveis para modelagem do sistema, foi utilizada a técnica de redução de variáveis a fim de diminuir o esforço computacional da ferramenta.

Na aplicação da análise de componentes principais (PCA), levou-se em consideração que a variável NumVentOn não incorporasse as entradas para a análise, pois ela seria usada para otimizar a temperatura do processo a partir de um otimizador, e para isso, seria mais fácil e rápido utilizar a variável sem que a transformasse em componente principal.

A análise de componentes principais visa à redução de dimensionalidade de um sistema, seja para visualização da resposta, ou, como é no caso desta dissertação, para simplificar o sistema sem que haja perda de informações significativa e, consequentemente, reduzir o esforço computacional para construção do sistema. Isso é possível devido ao fato de que, em sistemas com muitas variáveis, há informações redundantes expressadas por variáveis diferentes no mesmo processo e, ao aplicar a análise de componentes principais, é possível identificar e reduzir variáveis com pouca expressividade na saída do sistema.

O número de componentes principais considerados para o novo sistema foi quatro. Este valor foi obtido a partir da análise gráfica utilizando o método do *screeplot* (Figura 17), que é a análise da declividade dos autovalores de cada fator, localizando o fim da declividade e início da estabilização dos autovalores. A partir do PC4 é notadamente perceptível a estabilização dos autovalores [Neisse & Hongyu, 2016]. Com isso, o novo sistema com após redução de variáveis foi definido, conforme mostra a Figura 18.

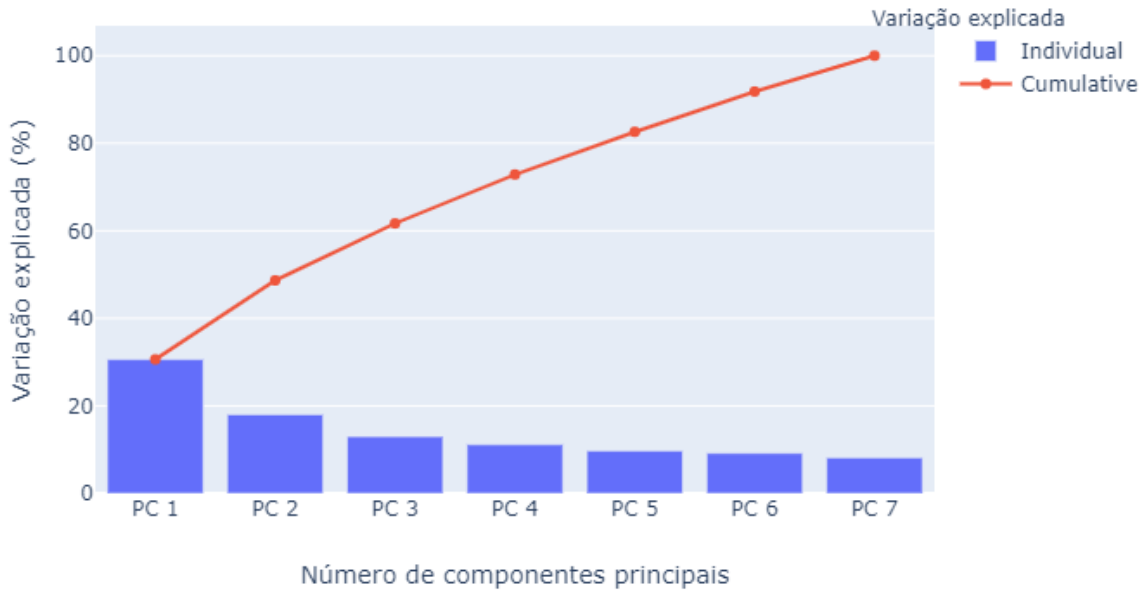


FIGURA 17 - VARIAÇÃO EXPLICADA POR NÚMERO DE COMPONENTES UTILIZADO.

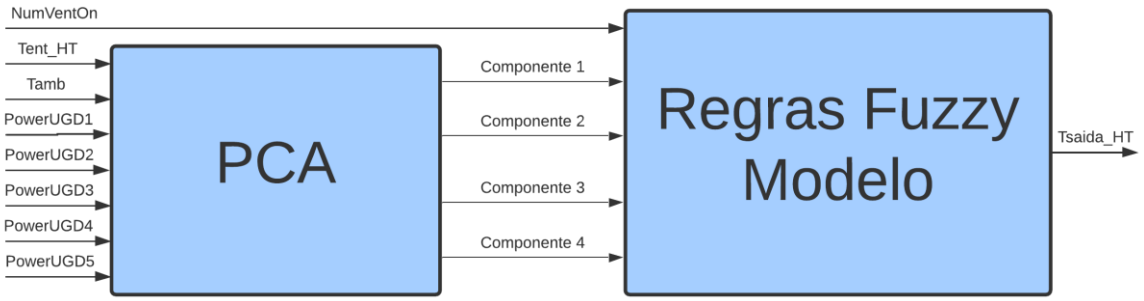


FIGURA 18 - MODELO PROPOSTO APÓS REDUÇÃO DE VARIÁVEIS USANDO PCA.

O sistema apresentado na Figura 16, chamado de sistema convencional, e o sistema apresentado na Figura 18, chamado de sistema reduzido, foram comparados para verificar a eficácia na redução do número de regras fuzzy no banco de regras dos modelos. A redução do número de regras extraídas pelo sistema reduzido em comparação ao sistema convencional é bastante expressiva. A partir do treinamento de um dia de dados, comparando o banco de regras do sistema convencional com o sistema reduzido, chegou-se uma redução de 83,79% na quantidade de regras não redundantes extraídas em um dia de treinamento.

3.2.4 Extração automática de regras usando Lógica Fuzzy

A fim de criar um modelo do sistema de refrigeração a partir de um sistema fuzzy, de acordo com o sistema da Figura 18, foram definidos os limites inferior e superior para cada variável de entrada e saída. Para os componentes principais, foram limitados entre -1 e 1, de acordo com a normalização utilizada para análise dos componentes. Para as variáveis NumVentOn e Tsaida_HT, medições físicas diretas, foram utilizados os limites reais da aplicação, ou seja: o limite inferior e superior de NumVentOn são, respectivamente, 0 e 55 ventiladores ligados, e o limite inferior e superior de TsaidaHT são, respectivamente, 20 e 100 °C, igualmente ao limite de medição do sensor.

Dessa forma, decidiu-se, a partir da comparação de vários testes, utilizar a metodologia da tentativa e erro para identificar o valor satisfatório, visando ao equilíbrio entre o baixo erro com menor esforço computacional e o número de regiões das funções de inferência para cada variável de entrada e saída. Foram realizados testes com as seguintes configurações:

- Utilizou-se um pequeno trecho do banco de dados de treinamento, cerca de 10% deste banco, ou seja, 8640 amostras;
- A partir de 20 conjuntos fuzzy para a variável de temperatura, foi variando de 20 em 20 até atingir 200 conjuntos, mantendo um número de subdivisões constantes;
- Da mesma forma que o anterior, para variáveis de componentes principais utilizou o ponto inicial igual a 20 e foi incrementando de 20 em 20 até 200 mantendo um número de subdivisões constantes;
- E para variável de número de ventiladores, iniciou-se em 15 e foi variando de 15 em 15 até o máximo 56;

Verificou-se que, quanto maior a quantidade de conjuntos fuzzy por variável, menor o erro entre o valor esperado e o valor real, considerando o banco de validação igual ao banco de treinamento adotado para o teste, porém, o número de regras aumentava consideravelmente, necessitando de um esforço computacional cada vez maior. Assim, adotou-se inicialmente a quantidade de funções de inferência da seguinte forma: TentHT e TsaidaHT, 100 funções triangulares distribuída de forma simétrica ao longo do eixo temperatura (20 a 100 °C); PCA1, PCA2, PCA3 e PCA4,

60 funções triangulares distribuída de forma simétrica ao longo do eixo das componentes (-1 a 1); NumVentOn, 56 funções triangulares distribuída de forma simétrica ao longo do eixo número de ventiladores ligados (0 a 55 ventiladores ligados), a fim testar o esforço computacional economizado a partir da redução do sistema ocorrida ao aplicar o PCA, como também o resultado do modelo.

O modelo apresentou o erro percentual abaixo de 5%, porém, o desempenho computacional ainda foi elevado, principalmente em operação por um longo período de amostragem (200000 amostras).

A Figura 19 exibe a variável analógica TsaidaHT fuzzificada com 100 conjuntos fuzzy para a temperatura. A Figura 20 exibe a variável PCA1, como as componentes estão normalizadas, todas as quatro componentes sofreram o mesmo processo de fuzzificação, ou seja, a PCA 2, 3 e 4 são idênticas à PCA1. A Figura 21 exibe a variável NumVentOn.

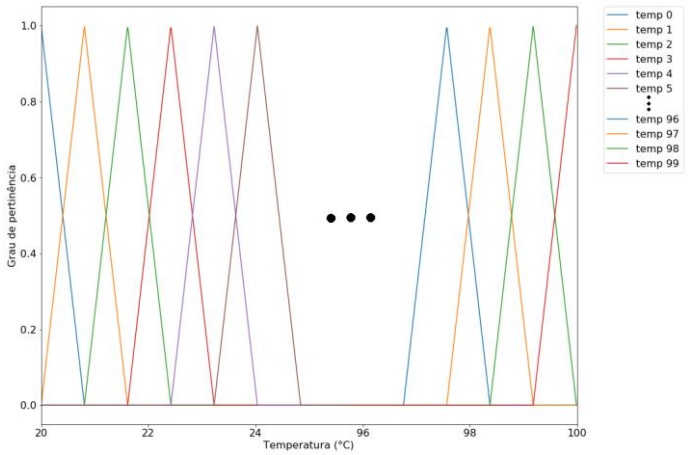


FIGURA 19 - CONJUNTOS FUZZY DOS ESPAÇOS TEMPERATURA TENTHT E TSAIDAHT.

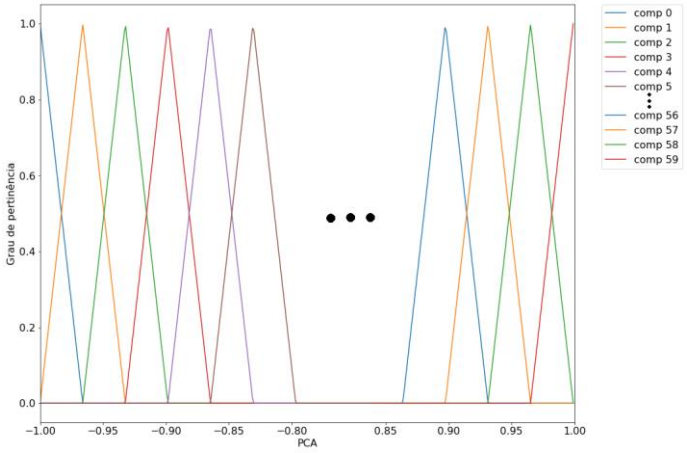


FIGURA 20 - CONJUNTOS FUZZY DOS ESPAÇOS PCA 1, 2, 3, 4.

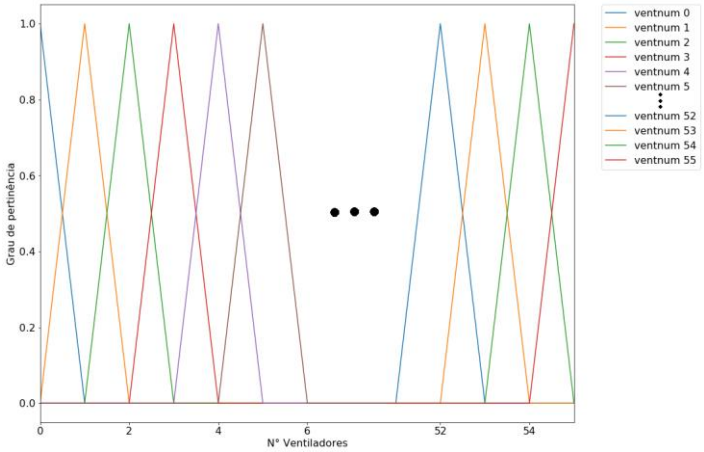


FIGURA 21 - CONJUNTOS FUZZY DO ESPAÇO NÚMERO DE VENTILADORES LIGADOS.

Para construir o banco de regras fuzzy que modela o funcionamento do sistema foi feita uma extração automática de regras a partir dos dados de um dia do banco de dados de treinamento, a uma amostragem por segundo para cada medição, totalizando 86.400 amostragens por variável de entrada/saída do sistema. A Figura 22 exibe os dados brutos, e a Figura 23 exibe os dados utilizados para extração de regras após redução de variáveis.

Na Figura 22, observa-se os dados de treinamento. Os dados utilizados se referem a uma parte da planta totalmente parada para geração em plena carga a qual permanece até o regime permanente estável. Nesta figura, observa que, na amostra número 65000, aproximadamente, ocorre o desligamento da unidade geradora 3, observado a partir da diminuição de potência P3, e logo após a manutenção retorna em operação. É possível perceber que, durante a redução de carga da unidade 3,

houve um distúrbio na temperatura de entrada e saída do radiador, o que é totalmente normal pois o radiador é compartilhado entre cinco unidades geradoras. Apesar disso, esses dados representam a rotina normal de operação da usina.

Na Figura 23, são ilustrados os dados processados da Figura 22, em que foram reduzidos em formas de componentes principais, e é dessa forma que o sistema fuzzy está enxergando o cenário de usina em partida até o regime permanente. É possível identificar distúrbios nos sistemas a partir dos componentes de forma rápida, pois uma ou vários componentes tendem a divergir do zero quando isso acontece, apesar de que não é possível identificar a causa do distúrbio a partir dos componentes.

Para a criação das regras utilizou-se o seguinte padrão: “Se (Entrada1 é x1) e (Entrada2 é x2) e … (EntradaN é xN) então Saída é y”.

A cada amostragem do banco de treinamento é retirada uma regra, totalizando 86400 regras, porém, dessa maneira, acabam existindo regras redundantes, devido a alta periodicidade da amostra. Assim, é feito um tratamento no banco de regras fuzzy a fim de remover regras em duplicidade e diminuir o esforço computacional do modelo. Das 86400 regras extraídas, após tratamento do banco de regras fuzy, chegou-se a 1376 regras únicas.

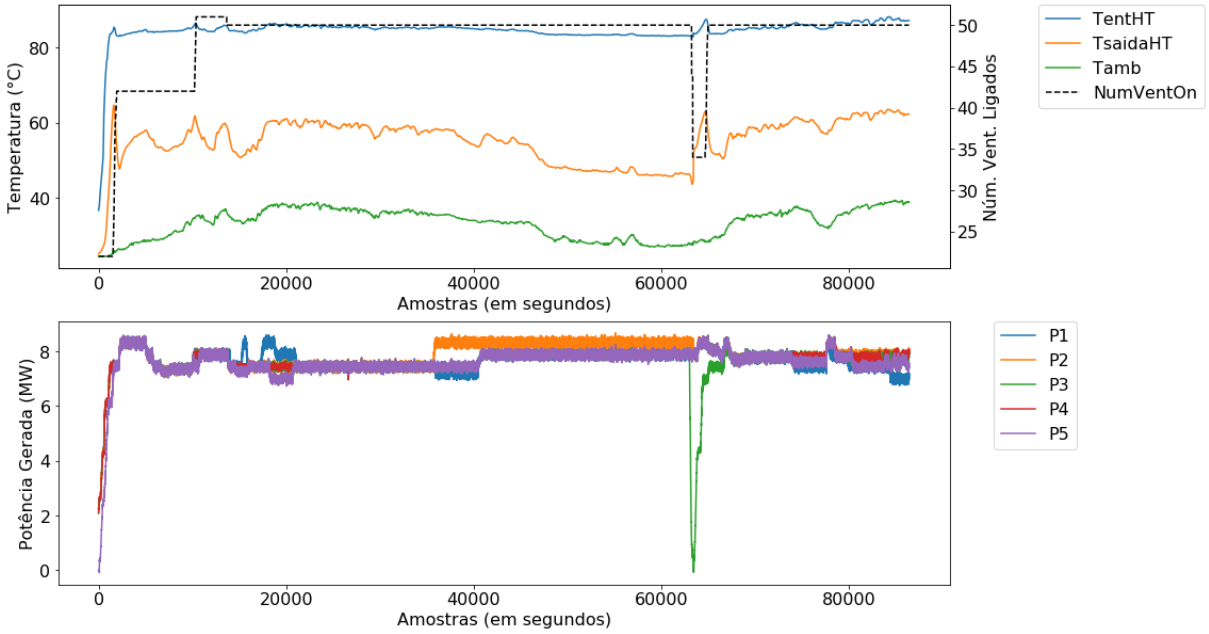


FIGURA 22 - VISUALIZAÇÃO DO BANCO DE DADOS DE TREINAMENTO, DADOS NA FORMA ORIGINAL.

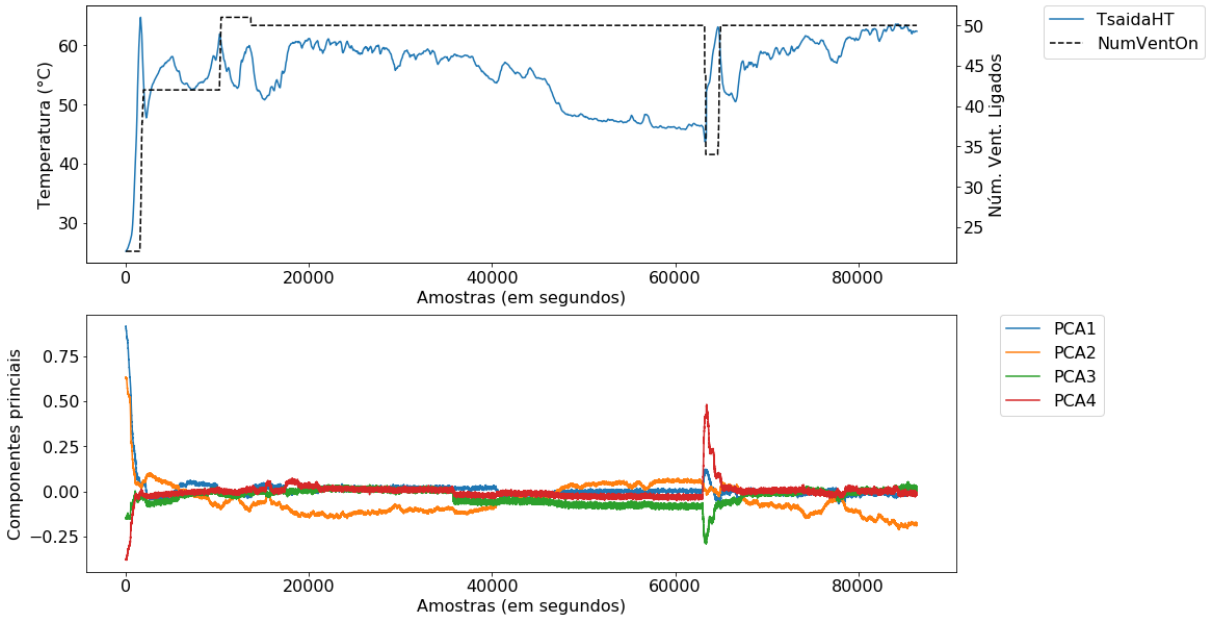


FIGURA 23 - VISUALIZAÇÃO DO BANCO DE DADOS DE TREINAMENTO, DADOS NA FORMA REDUZIDA ATRAVÉS DE PCA.

3.2.5 Atualização da base de conhecimento de forma dinâmica

O conceito de gêmeo digital consiste em construir um modelo virtual que tem a capacidade de simular ativos físicos para fins de previsão de falhas e identificar potenciais problemas que o sistema físico possa vir a ter e, para isso, é preciso que o modelo seja capaz de se adaptar continuamente conforme seu gêmeo físico (FARSI, 2020).

Com base nessa característica do gêmeo digital, propõe-se o desenvolvimento de um método para atualização da base de conhecimento de forma dinâmica a fim de “sintonizar” o modelo conforme o resultado destoa do valor real do processo.

Normalmente, a aprendizagem de modelos a partir de dados é feita utilizando uma quantidade de dados específica buscando patamares abaixo de uma métrica estabelecida, como por exemplo, erro médio quadrático. Entretanto, ao longo do tempo, a planta do processo vai tendo sua característica modificada devido aos desgastes dos equipamentos e às manutenções realizadas nas quais há a troca de peças e, dessa forma, as regras aprendidas ficam obsoletas, principalmente se o

banco de treinamento utilizar dados antigos que não refletem a dinâmica atual do sistema.

Portanto, quando a regra utilizada para simular a saída do sistema começa a ter um erro maior do que o resultado verificado no banco de teste, indica que a regra está obsoleta ou tendendo a isto. Dessa forma, foi criado um algoritmo de atualização de regras em tempo real para aprender a nova dinâmica do sistema, caracterizando um gêmeo digital.

O algoritmo verifica duas condições importantes para o modelo: 1) a saída do bloco de defuzzificação quando zero indica valores inválidos (sem ativação de regras) e/ou 2) garantir um delta-erro dentro do admissível. Quando essas condições apresentam respostas insatisfatórias, é necessária a extração de novas regras a partir do cenário que gerou a condição inválida, e dessa forma, atualizar a aprendizagem do modelo sem a necessidade de ter que reaprender toda a base com novos dados novamente. Outro ponto importante é que o algoritmo permite inserir regras, caso sejam únicas, a qualquer momento, a partir do conhecimento do especialista industrial. A Figura 24 contém o fluxograma do algoritmo implementado. Para melhor entendimento do fluxograma, este será apresentado a partir de passos:

- Primeiro passo: As entradas já reduzidas são carregadas no sistema fuzzy.
- Segundo passo: Inicia-se a simulação do sistema, nessa etapa será executada a sequência básica de um sistema de inferência fuzzy, ou seja, as entradas serão fuzzificadas, depois serão processadas pelo motor de inferência a partir do banco de regras do sistema para realizar a tomada de decisão da saída fuzzificada, e por fim, ocorrerá a defuzzificação do resultado a fim de converter o valor nebuloso em valores reais,
- Terceiro passo: Nesse momento, caso não haja a ativação de regras, o conjunto fuzzy de saída será nula, ou seja, irá gerar um valor inválido na saída do sistema e iniciará o quarto passo, caso haja ativação de regras, o resultado da defuzzificação será diferente de zero e irá para o sexto passo.
- Quarto passo: O algoritmo de atualização de regras será ativado e irá iniciar a coleta dos valores dos sensores de campo de todas as entradas e saídas daquele instante, logo após serão normalizadas e aplicada a análise de componentes principais para redução de variáveis.

- Quinto passo: A partir das variáveis de entradas reduzidas e saída coletadas, é feita a extração automática de regra a partir desses dados, logo em seguida, irá buscar dentro do banco de regras do sistema se há alguma regra igual à nova regra extraída, caso não seja encontrada, a nova regra será incorporada no banco de regras fuzzy e retornará para o segundo passo, caso seja, a nova regra será descartada. Caso persista o acionamento do algoritmo de atualização com a ocorrência de regra redundante, a simulação será encerrada.
- Sexto passo: O resultado do modelo será avaliado quanto ao erro percentual instantâneo, caso esse erro seja menor que o valor considerado coerente, que no caso deste trabalho foi adotado 5%, o resultado será considerado validado e irá para a variável de saída do modelo. Caso o erro seja maior que 5%, irá ativar o algoritmo de atualização de regras para verificar a possibilidade de uma regra melhor do que a ativada no resultado atual, ou seja, irá para o quarto passo.

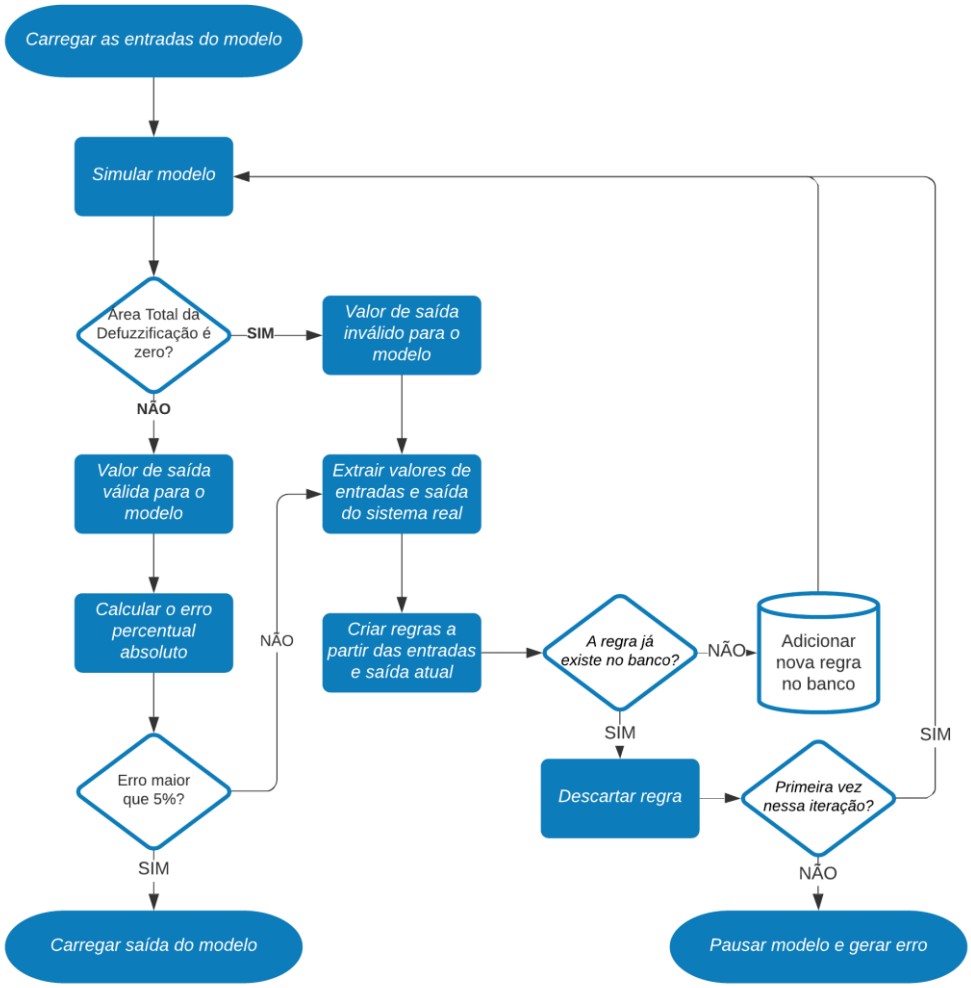


FIGURA 24 - FLUXOGRAMA DO ALGORITMO DE ATUALIZAÇÃO DE REGRAS.

3.2.6 Otimização utilizando estratégias evolutivas

Na busca de diminuir o esforço computacional ainda mais e com o menor impacto possível no resultado do modelo, desenvolveu-se um algoritmo de otimização do tipo AG, cujo objetivo do algoritmo é encontrar a melhor forma dos conjuntos fuzzy, do tipo triangular, que possa diminuir o número de regras a fim de diminuir o esforço computacional do modelo e ao mesmo tempo que o modelo apresentasse um resultado no erro quadrático satisfatório para a aplicação.

A seguir, serão apresentados todos os passos realizados pelo algoritmo.

3.2.6.1 Inicialização da população

O AG irá buscar o formato otimizado dos conjuntos fuzzy do sistema, o primeiro passo foi preparar o vetor cromossomo, no qual cada gene armazenado no cromossomo representa uma possível solução do problema e armazenado nos genes estão os alelos, os parâmetros a serem otimizados. Utilizou-se uma codificação representada por números inteiros dispostos em n vetores-vértice de três números, em que tais números representam os vértices de um triângulo e n é o número de conjuntos fuzzy de uma variável entrada/saída do processo, pertencente a um vetor-variável de quantidade m, onde m é o número de variáveis de entrada/saída do modelo, que no caso deste trabalho é seis. A população inicial do AG é dado pela variável k. A Figura 25 ilustra a estrutura básica da população.

A população inicial é criada de forma aleatória, porém, foi adicionada uma variável guia para que ao longo da iteração os vértices aleatórios gerados no conjunto fuzzy x-1 possuam valores menores que os gerados no conjunto fuzzy x e assim sucessivamente. Também foi aplicado limites (saturador) nos valores gerados para que os vértices dos conjuntos das variáveis de PCA’s se mantivessem entre -1 e 1, os conjuntos da variável NumVent se mantivessem entre 0 e 55 e os conjuntos da variável Tout se mativessem entre 20 e 100. No final da geração dos valores aleatórios de um conjunto fuzzy, é feita uma ordenação a fim de manter o vértice a > b > c.

O pseudocódigo a seguir demonstrará a criação da população inicial.

TABELA 2 - PSEUDO-CÓDIGO ALGORITMO PARA CRIAÇÃO DA POPULAÇÃO INICIAL.

```
Programa População inicial
Início
        k = número de indivíduos na população
        enquanto iteração <= k fazer
        início
        Inicializa o vetor referente aos vértices dos conjuntos para variável PCA1 do indivíduo k
        Inicializa o vetor referente aos vértices dos conjuntos para variável PCA2 do indivíduo k
        Inicializa o vetor referente aos vértices dos conjuntos para variável PCA3 do indivíduo k
        Inicializa o vetor referente aos vértices dos conjuntos para variável PCA4 do indivíduo k
        Inicializa o vetor referente aos vértices dos conjuntos para variável NumVentOn do indivíduo k
        Inicializa o vetor referente aos vértices dos conjuntos para variável Tout do indivíduo k
        Limita todos os vértices de PCA1 entre -1 e 1
        Limita todos os vértices de PCA2 entre -1 e 1
        Limita todos os vértices de PCA3 entre -1 e 1
        Limita todos os vértices de PCA4 entre -1 e 1
        Limita todos os vértices de NumVentOn entre 0 e 55
        Limita todos os vértices de Tout entre 20 e 100
        Ordena os vértices nas variáveis para que a > b > c onde a = menor vértice, b = vértice central e c = maior vértice.
        Adiciona os vetores de vértices na matrix população inicial na ordem criada.
        fim
fim
```

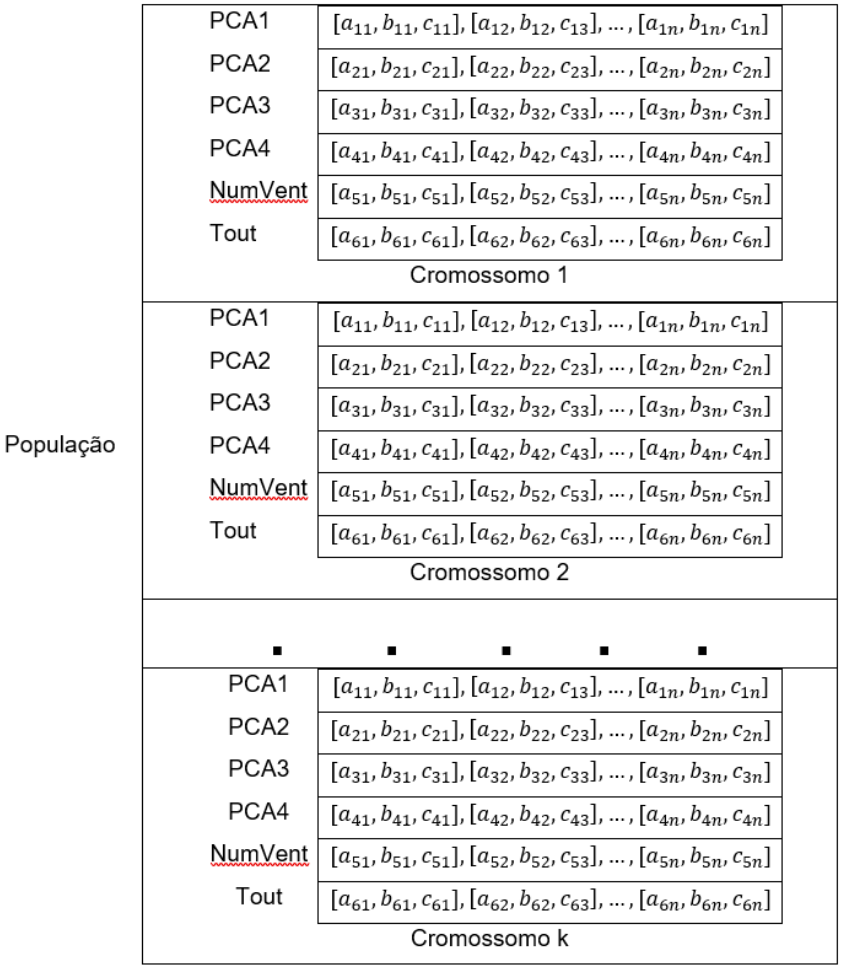


FIGURA 25 - ESTRUTURA DA POPULAÇÃO INICIAL DO AG.

3.2.6.2 Avaliação da função fitness

Para avaliar as possíveis soluções, foi criada uma rotina com um banco de treinamento reduzido (possuindo 10000 segundos, cerca de três horas, extraídos do banco de dados da termelétrica) pois, para avaliar o gêmo digital é preciso simular o modelo e avaliar o erro quadrático. Assim, a rotina recebe a população da geração como entrada e realiza as etapas de extração automática de regras a partir dos conjuntos fuzzy definidos pela população inicial, conforme explicado na seção 3.2.4, e, em seguida, realiza a simulação em cima dos mesmos dados de treinamento com a função de atualização de regras, apresentado na seção 3.2.5, desativado.

O erro quadrático médio da simulação do banco de treinamento reduzido, subtraído de 100, é o resultado da qualidade da função fitness, quanto mais próximo de 100 melhor é a solução.

3.2.6.3 Método de Seleção

A seleção dos cromossomos tem como objetivo escolher os indivíduos que contribuirão com a próxima geração. Desta forma, a partir da geração atual são escolhidos n_{sel} indivíduos que serão separados em pares para o cruzamento. O método de seleção utilizado no algoritmo é o método da roleta, explicado na seção 2.6 deste trabalho. Apesar do método de seleção parecer puramente probabilístico, o método apresenta uma característica que direciona a seleção de forma a escolher os melhores cromossomos da geração, isso ocorre devido ao fato da qualidade do fitness influenciar na probabilidade da escolha.

3.2.6.4 Cruzamento

Após selecionados n_{sel} indivíduos separados em pares, os cromossomos irão cruzar a partir do cruzamento aritmético. Na seção 2.6 foi apresentado o cruzamento aritmético em um ponto, porém, neste trabalho foi aplicado o cruzamento aritmético em todo o cromossomo. O par de cromossomos tem uma probabilidade (p_{cruz}) de ocorrer o cruzamento, caso ocorra, o par irá gerar um novo par descendente onde o primeiro novo indivíduo vai herdar um valor aleatório multiplicado pelo valor do alelo do cromossomo 1 mais o complemento desse valor aleatório multiplicado pelo valor do alelo do cromossomo 2 e o segundo novo indivíduo vai herdar o complementado do número aleatório multiplicado pelo valor do alelo do cromossomo 1 mais o valor aleatório multiplicado pelo valor do alelo do cromossomo 2. Expressando em uma equação, teremos:

$$\alpha = rand(0,1) \tag{4}$$

$$cromossomo1_{g_{(n+1)}} = (\alpha * cromossomo1_{g_n}) + ((1-\alpha) * cromossomo2_{g_n}) \tag{5}$$

$$cromossomo2_{g_{(n+1)}} = ((1-\alpha) * cromossomo1_{g_n}) + (\alpha * cromossomo2_{g_n}) \tag{6}$$

No caso não ocorra o cruzamento, os cromossomos selecionados da geração anterior irão para a próxima geração.

3.2.6.5 Mutação

Após a etapa do cruzamento, os novos cromossomos poderão sofrer mutação. Nesta etapa, todos os alelos dos novos cromossomos serão mutados de acordo com a probabilidade p_{mut}, os alelos mutados sofrerão o método de mutação creep, apresentado na seção 2.6, onde a codificação do alelo terá um valor aleatório entre -1 e 1, $\alpha = rand(-1,1)$, adicionado ao valor do vértice do triangulo contido naquele alelo. Após mutação, os alelos serão testados para verificar que não ultrapassaram os limites das variáveis de entrada/saída, caso ocorra ultrapassagem serão substituídos pelo limite mais próximo.

4 RESULTADOS

4 RESULTADOS

Nessa seção, serão apresentados os resultados obtidos através da aplicação da técnica de modelagem a partir de três cenários:

- Cenário 4.1: desempenho do digital twin operando durante o regime permanente do processo;
- Cenário 4.2: desempenho do digital twin durante transientes no processo (partida e parada de planta);
- Cenário 4.3: desempenho do digital twin durante grande transientes decorridos de falha no controle do sistema;

Os dados utilizados nos três cenários foram retirados do banco de dados da planta. Foi adotado o valor de erro admissível percentual instantâneo máximo de 5%, tal erro representa o gatilho necessário para o algoritmo criar regras no intuito de diminuir o erro. Este valor de erro adotado representa 3°C da temperatura de saída do radiador, valor este considerado aceitável pelo especialista da planta, pois não apresentaria problemas ao sistema porque o controle atual utiliza uma banda de faixa morta de +3°C, ou seja, não ocasionaria acionamentos de alarmes ou desligamentos desnecessários caso fosse utilizado o modelo para auxílio do controle.

Os dados do banco de treinamento são provenientes de equipamentos instalados em uma termelétrica a combustão interna com as seguintes especificações: sistema de refrigeração de água quente com 55 ventiladores acionados remotamente, de forma automatizada ou manual, por operadores da planta, e cinco unidades moto-geradoras compostas por motor MAN/STX 18V32/40 + gerador Hyundai de 10MVA. Tais equipamentos são monitorados a partir de sensores e transmissores de temperatura do tipo pt100 e termopar.

A otimização dos conjuntos fuzzy é uma tarefa bastante complexa pelo fato de que é possível otimizar de várias formas. A primeira forma é a quantidade de conjuntos fuzzy por variável (será considerado que todas as variáveis fuzzy estarão com mesma quantidade) e o segundo parâmetro é a forma de dos conjuntos fuzzy. Avaliando o problema por esses dois horizontes, se tem uma infinidade de soluções otimizadas para esse problema, assim, foi analisado a influência da quantidade de conjuntos fuzzy com formato triangular disperso uniformemente dentro o range da

variável no erro quadrático do modelo, como também na quantidade de regras contida no banco de regras.

Para essa avaliação, foi realizada a simulação de vários modelos contendo de dez a 100 conjuntos fuzzy por variável, se utilizou de um banco de treinamento reduzido (com aproximadamente 8760 amostras de um segundo) e observado o número de regras criado e o resultado de cada modelo para o banco teste igual ao banco de treinamento utilizado. A Figura 26 mostra o resultado obtido.

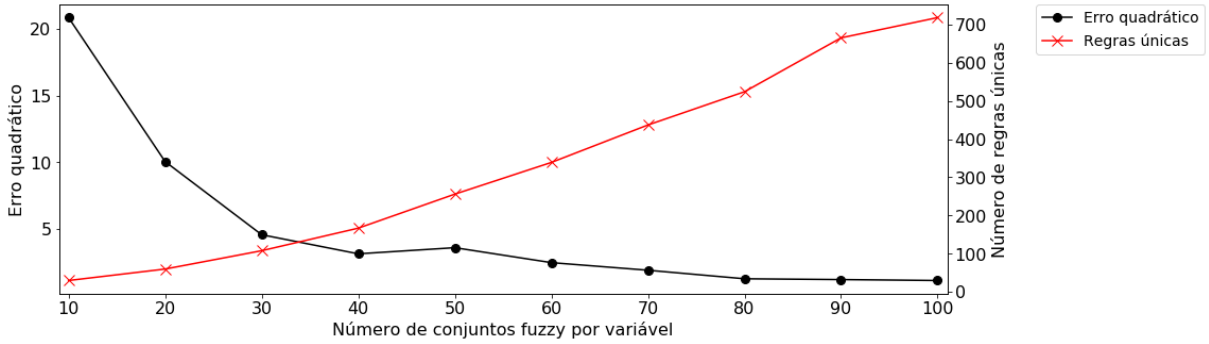


FIGURA 26 - VARIAÇÃO DO ERRO QUADRÁTICO E DO NÚMERO DE REGRAS EM FUNÇÃO DO NÚMERO DE CONJUNTO FUZZY.

O número de conjuntos por variável tem uma forte influência no desempenho do modelo vista pela métrica do erro quadrático, apresenta uma curva característica de decaimento do erro de forma exponencial na qual, a partir de 40 conjuntos, é perceptível uma estabilização no decaimento do erro, não causando uma forte melhoria no desempenho. Em contrapartida do decaimento do erro, pode-se ver que o número de regras tem uma tendência de aumento, e o banco com muitas regras exige mais esforço computacional do hardware para manter o funcionamento do modelo, o que não é desejado.

Desta forma, buscou-se otimizar o modelo de forma a buscar um banco de regras com o menor número possível (pensando em menor esforço computacional) e melhor desempenho, e, tendo em vista que o modelo com conjunto abaixo de 20 tende a erros muito altos. Para isso foi aplicado um algoritmo genético com o intuito de otimizar as formas e disposições dos conjuntos fuzzy de cada variável para a quantidade 20 a 40 conjuntos, incrementando de dois em dois a cada simulação. Os parâmetros utilizados no algoritmo genético foram:

- Número de população inicial: 10;

- Número de cromossomos selecionados para cruzamento e mutação: 6;
- Probabilidade de ocorrer cruzamento: 0,8;
- Probabilidade de ocorrer mutação: 0,02;
- Número de cromossomos substituído por geração: 2;
- Parâmetro de parada do algoritmo: após 50 gerações.

A Figura 27 ilustra o desempenho das melhores soluções encontradas pelo algoritmo genético para cada número de conjuntos fuzzy.

O algoritmo de otimização só teve uma única iteração para cada solução descrita no gráfico, pelo fato de que o tempo médio necessário para a realização dos cálculos em 50 gerações é de 24 horas por solução. Idealmente, para buscar a melhor solução dentre os números dos conjuntos, seria necessária a simulação do algoritmo de otimização por pelo menos 100 vezes por solução, porém, para isso seria necessário investir uma quantidade de tempo inviável para o trabalho ou a aquisição de vários computadores potentes para agilizar tal análise.

Desta forma, dentre as melhores soluções encontradas, a solução com 40 conjuntos apresentou o melhor desempenho e a solução com 30 e 34 conjuntos apresentaram os piores desempenhos. Possivelmente, isso ocorreu devido ao fato das soluções com 30 e 34 conjuntos terem caído em um máximo local durante a procura da melhor solução e não conseguiu encontrar o máximo global. Um dos fatores que pode ter influenciado no resultado ruim foi a inicialização da população aleatória não ter estado muito distante da melhor solução global.

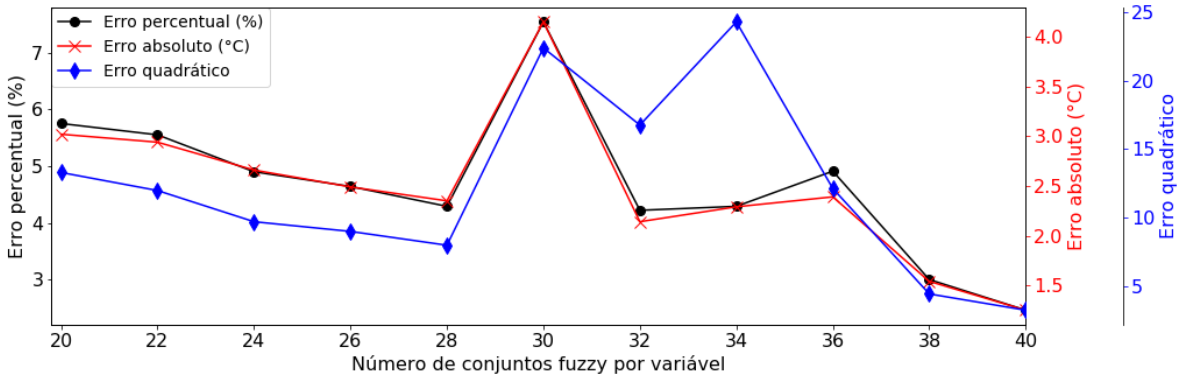


FIGURA 27 - RESULTADO DO ALGORITMO GENÉTICO EM FUNÇÃO DO NÚMERO DE CONJUNTOS FUZZY.

Quando um sistema de inferência fuzzy baseado em regras é utilizado, a interpretabilidade das regras se torna um dos pontos chaves do sistema, tendo em vista que uma das características mais importantes dos sistemas fuzzy é a conversão dos termos linguísticos capaz de facilitar o entendimento de processos não-lineares pelo o especialista, aumentando a capacidade de análise e entendimento dos resultados. Quando se utiliza um sistema fuzzy existem dois objetivos principais, o primeiro é maximizar a acurácia do sistema e o segundo é minimizar a complexidade do sistema e, com isso, aumentar a interpretabilidade. Tais objetivos, porém, acabam sendo contraditórios pois o aumento da acurácia do sistema tende a aumentar a complexidade das regras, logo, diminuindo a interpretabilidade [Ishibashi e Nascimento, 2012]. Jin e Sendhoff em 2003, listaram alguns pontos a serem observados o qual tornam o sistema fuzzy interpretável:

- Entradas e saídas bem distinguíveis com significado físico;
- Número de conjuntos fuzzy limitados;
- Regras bem consistentes, ou seja, sem duplicidade nas regras;
- Minimizar o número de entradas do sistema como também o número de regras no banco;

Não foi possível limitar o número de conjuntos fuzzy a valores capazes de manter a interpretabilidade pois, como visto no resultado da Figura 26, mesmo com 20 conjuntos fuzzy por variável de entrada, o modelo obteve uma acurácia muito baixa, sendo que 20 conjuntos já é um número alto e incapaz do ser humano interpretar regras, e até mesmo criar termos linguísticos para tantos conjuntos.

Utilizou-se, para a construção do *digital twin* neste trabalho, algumas das premissas observadas por Jin e Sendhoff, a minimização do número de regras e entradas do sistema, buscando a otimização computacional do algoritmo, porém, ao aplicar estratégias de redução de dimensionalidade, a interpretabilidade do sistema é afetada. Nesta estratégia, as variáveis de entrada são transformadas em componentes principais que não possuem significado físico em prol da redução do número de entradas e, consequentemente, na redução do número de regras, comprovada a partir dos resultados testados.

Assim, buscou-se, através da utilização do AG, a diminuição dos conjuntos fuzzy para diminuir o esforço computacional, não mais com o objetivo de melhorar a interpretabilidade perdida ao longo dos esforços para melhorar a acurácia e diminuir

o esforço computacional. Para atingir o objetivo do trabalho, foi necessário assumir um compromisso (*tradeoff*), isto é, diminuir a interpretabilidade a fim de conseguir patamares de erro quadrático médio menor e processamento mais baixo do algoritmo através da análise de componentes principais e otimização dos conjuntos utilizando AG.

4.1 FUNCIONAMENTO DO MODELO DURANTE PERÍODO DE REGIME PERMANENTE

Para o cenário a seguir, o modelo foi testado utilizando os dados do banco teste, com amostras de um segundo e período igual a 40000 amostras, equivalente a pouco mais que 11 horas de geração ininterrupta. Os dados são referentes a uma geração em que não houve partida e nem parada de unidade geradora, somente pequenas mudanças de carga. A Figura 28 apresenta os dados de entrada desse período divididos em: (a) dados de potências, (b) dados de temperatura de entrada do radiador e temperatura ambiente, e (c) dados convertidos em componentes principais junto com o número de radiadores.

O modelo apresentou bom resultado durante o regime permanente, apresentando erro médio percentual de 5,1% e erro médio absoluto de 2,76°C. Na Figura 29, é possível ver o resultado da simulação do modelo versus valor real de temperatura do sistema e o erro percentual instantâneo, parâmetro utilizado para o gatilho de aprendizagem de novas regras durante a simulação. A simulação do modelo iniciou-se com o número de regras aprendidas no banco de dados de treinamento, 547 regras, após simulação o modelo detectou oportunidades de aprendizagem e finalizou o período com 591 regras, o motivo da aprendizagem são a falta de regras e pontos que atingiram acima de $\pm$ 5% de erro percentual instantâneo.

Neste cenário, foi possível perceber que a temperatura de saída do sistema de refrigeração, tanto o dado real quanto o modelo, caíram de 60°C para abaixo de 50°C no intervalo de amostragem 25000 a 35000, onde a potência das unidades geradoras UGDs se manteve constante e número de ventiladores ligados constantemente, havendo variação na temperatura de entrada de água quente (Tent_HT) e na temperatura ambiente. Isso acontece pois o sistema real apresenta

uma deficiência no controle de temperatura -não consegue regular a temperatura de saída do sistema quando a temperatura ambiente cai devido a mudança do dia para a noite e/ou acontece pancadas de chuvas. Na sessão seguinte, será mostrada a otimização do controle de temperatura do sistema utilizando o modelo feito neste trabalho, na qual o modelo vai sugerir para o operador ou o próprio sistema de controle o melhor número de ventiladores de forma a otimizar o gasto de energia do sistema e garantir o bom funcionamento dos motores.

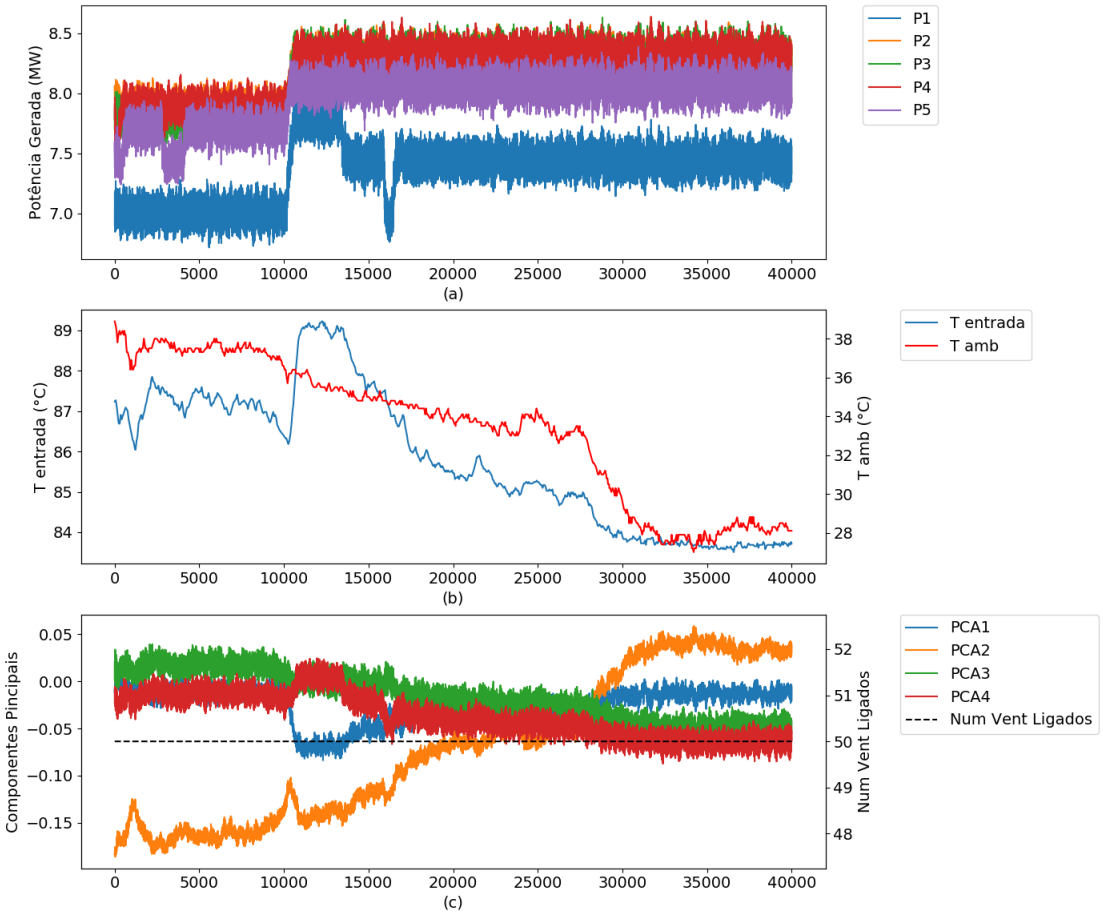


FIGURA 28 - ENTRADAS DO PROCESSO NO CENÁRIO 1, (A) POTÊNCIAS DAS UGD'S, (B) TENTHT E TAMB E (C) ENTRADAS REDUZIDAS E NUMVENTON.

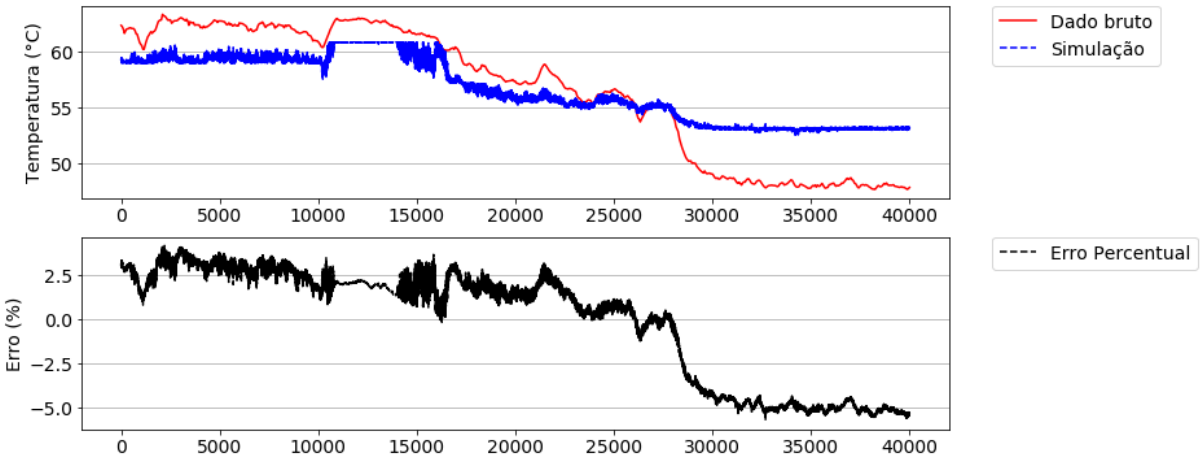


FIGURA 29 - RESULTADO MODELO VERSUS VALOR REAL TSAIDAHT E ERRO PERCENTUAL INSTANTÂNEO DO CENÁRIO 1.

4.2 FUNCIONAMENTO DO MODELO DURANTE A RAMPA DE PARADA E RAMPA DE PARTIDA DA PLANTA

As condições de amostragem e periodicidade foram mantidas para o teste no cenário 2. Neste cenário, a planta estava com programação com parada e partida durante o período de geração exposto. A Figura 30 apresenta os dados de entrada desse período divididos em: (a) dados de potências, (b) dados de temperatura de entrada do radiador e temperatura ambiente, e (c) dados convertidos em componentes principais junto com o número de radiadores.

O modelo do sistema apresentou um bom desempenho, inclusive melhor do que o anterior, apresentando erro médio percentual de 4,41% e erro médio absoluto de 2,37 °C. A partir da Figura 31, é possível ver que em grande parte do tempo o modelo apresentou erro percentual instantâneo abaixo que 5%. Já nos períodos de transições (rampa de parada e rampa de subida), apresentou instantes nos quais o erro ultrapassou os 5% ativando o gatilho para aprender novas regras e assim voltando a baixar o erro. Isso aconteceu pois o modelo não tinha sido treinado em momentos de grande transição, as regras do modelo aumentaram de 591 regras para 754 regras.

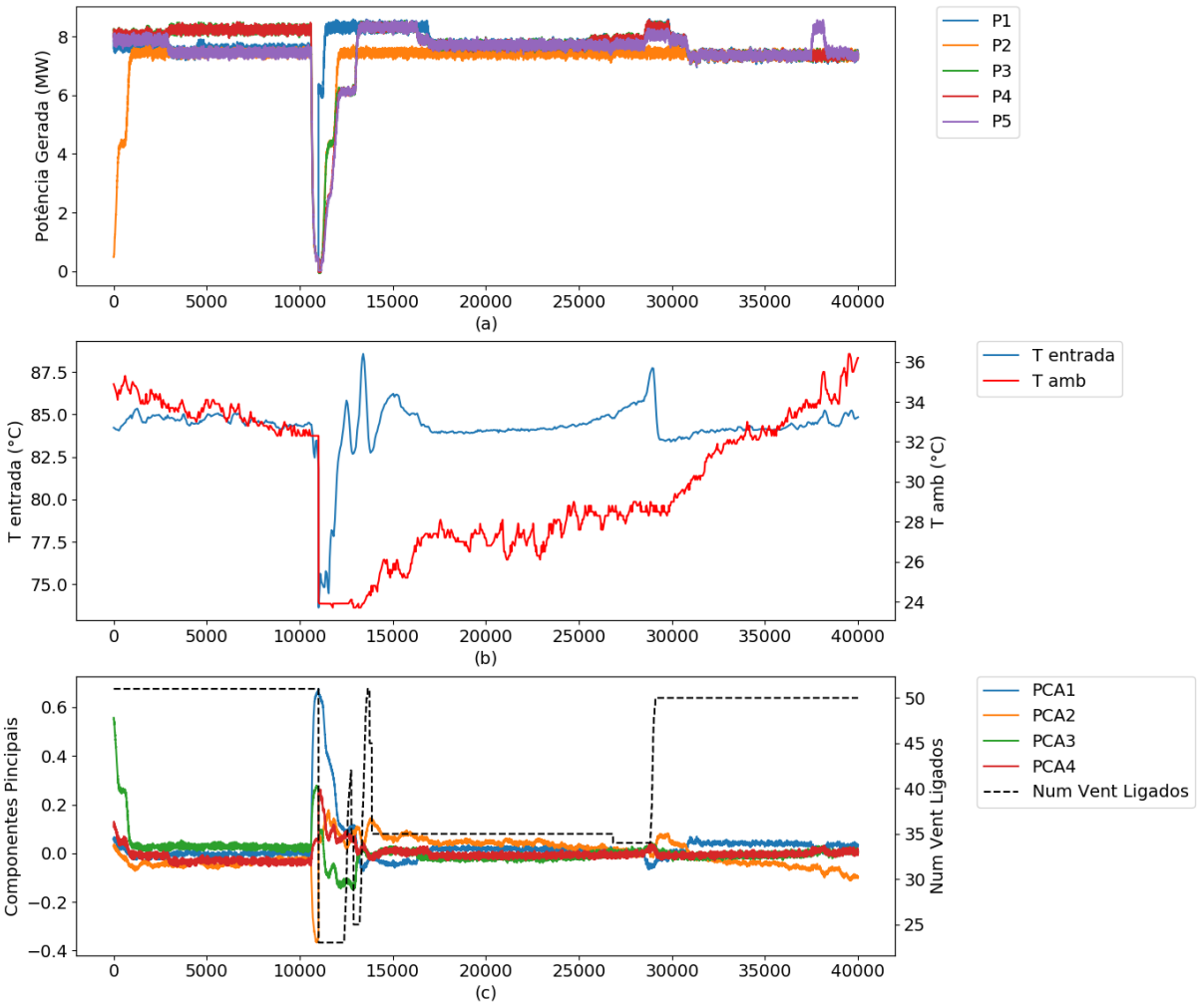


FIGURA 30 - ENTRADAS DO PROCESSO NO CENÁRIO 2, (A) POTÊNCIAS DAS UGD'S, (B) TENTHT E TAMB E (C) ENTRADAS REDUZIDAS E NUMVENTON.

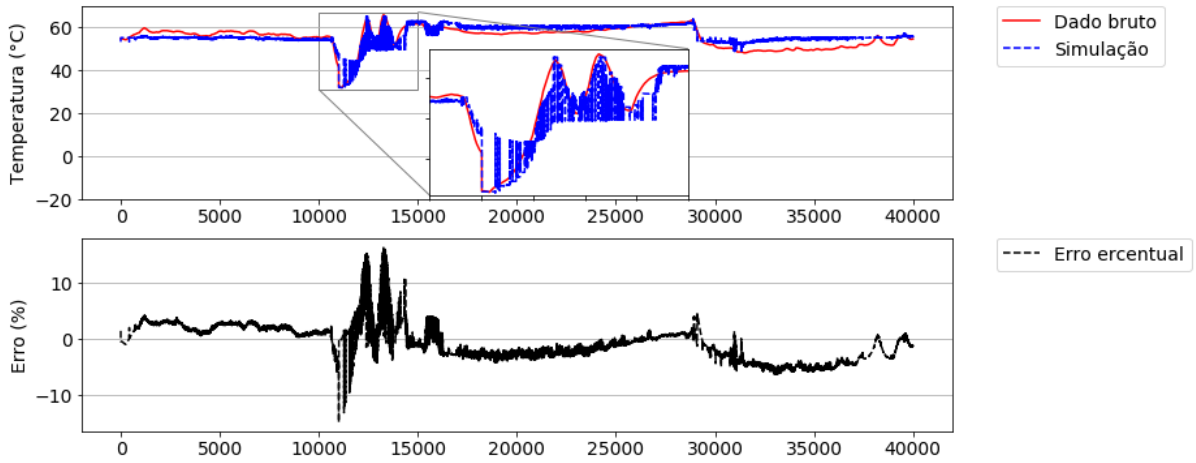


FIGURA 31 - RESULTADO MODELO VERSUS VALOR REAL TSAIDAHT E ERRO PERCENTUAL INSTANTÂNEO DO CENÁRIO 2.

A funcionalidade de aprendizagem durante operação do modelo apresentou grande eficácia e importância, principalmente pensando-se na utilização da abordagem *digital twin* para fornecer sugestões em tempo real ao operador, pois evita que o modelo divirja do real ao longo prazo, sua eficácia no ajuste do erro a curto prazo é verificada ao calcular o valor médio do erro percentual no período simulado.

Foi realizada uma nova simulação utilizando os mesmos dados, a fim de validar a eficiência do método de aprendizagem dinâmica. Os resultados são apresentados na Figura 32. O modelo apresentou resultados melhores durante o período de transiente do que o apresentado na Figura 31, isso devido ao fato do modelo ter aprendido regras durante a operação passada que auxiliaram na operação atual. Na validação, o modelo apresentou uma pequena melhora no desempenho médio do período, com erro médio absoluto de 2,18°C e erro médio percentual de 4,07%, porém, é possível verificar uma melhora no desempenho considerável através da figura do erro instantâneo, principalmente se observado o erro máximo percentual atingido onde no cenário anterior apresentou erro percentual máximo acima de 10% e neste cenário abaixo de 10%.

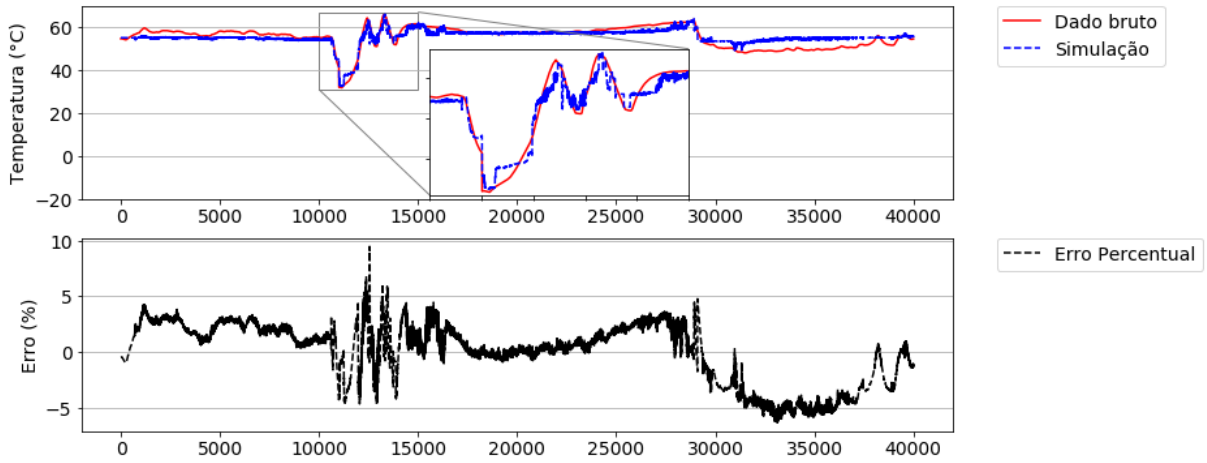


FIGURA 32 - RESULTADO DA SIMULAÇÃO DA AUTO VALIDAÇÃO DO MODELO VERSUS VALOR REAL DE SAÍDA DE TEMPERATURA DO SISTEMA E GRÁFICO DO ERRO PERCENTUAL INSTANTÂNEO, NO CENÁRIO 2.

4.3 FUNCIONAMENTO DO MODELO DURANTE PERÍODO COM PERTURBAÇÃO GERADO POR UM PROBLEMA NA MALHA DE CONTROLE DO SISTEMA DE REFRIGERAÇÃO

No cenário 3, foram extraídos os dados de um período em que o sistema apresentou oscilações anormais devido o desligamento/acionamento constante de um grande grupo de ventiladores devido a uma instabilidade do controle. As condições de amostragem e período do cenário mantiveram-se iguais aos dois cenários apresentados anteriormente, 40000 amostras de um segundo. A Figura 33 apresenta os dados de entrada desse período: (a) dados de potências, (b) dados de temperatura de entrada do radiador e temperatura ambiente, e (c) dados convertidos em componentes principais junto com o número de radiadores.

O modelo obteve uma ótima resposta mesmo em condições anormais de operação do sistema, encontrado entre as amostras 8000 e 12000 aproximadamente, como apresentado na Figuras 34. O erro percentual instantâneo permaneceu entre a faixa de $\pm$ 5% durante a maior parte do tempo, ultrapassando esse valor em poucas situações. Isso ocorreu pelo fato do modelo já ter passado por uma experiência similar a essa situação, no caso uma rampa de parada/partida da planta, ou seja, já existiam regras aprendidas no cenário 2 que ajudaram na decisão nesse período. Mesmo assim, após finalizar a simulação de todo o cenário, o modelo adquiriu um montante de novas regras considerável, totalizando 974 regras no banco.

O modelo apresentou um erro percentual médio neste cenário de 2,77% e 1,65°C, o que mostra uma alta precisão na simulação do sistema mesmo em situações anormais do sistema.

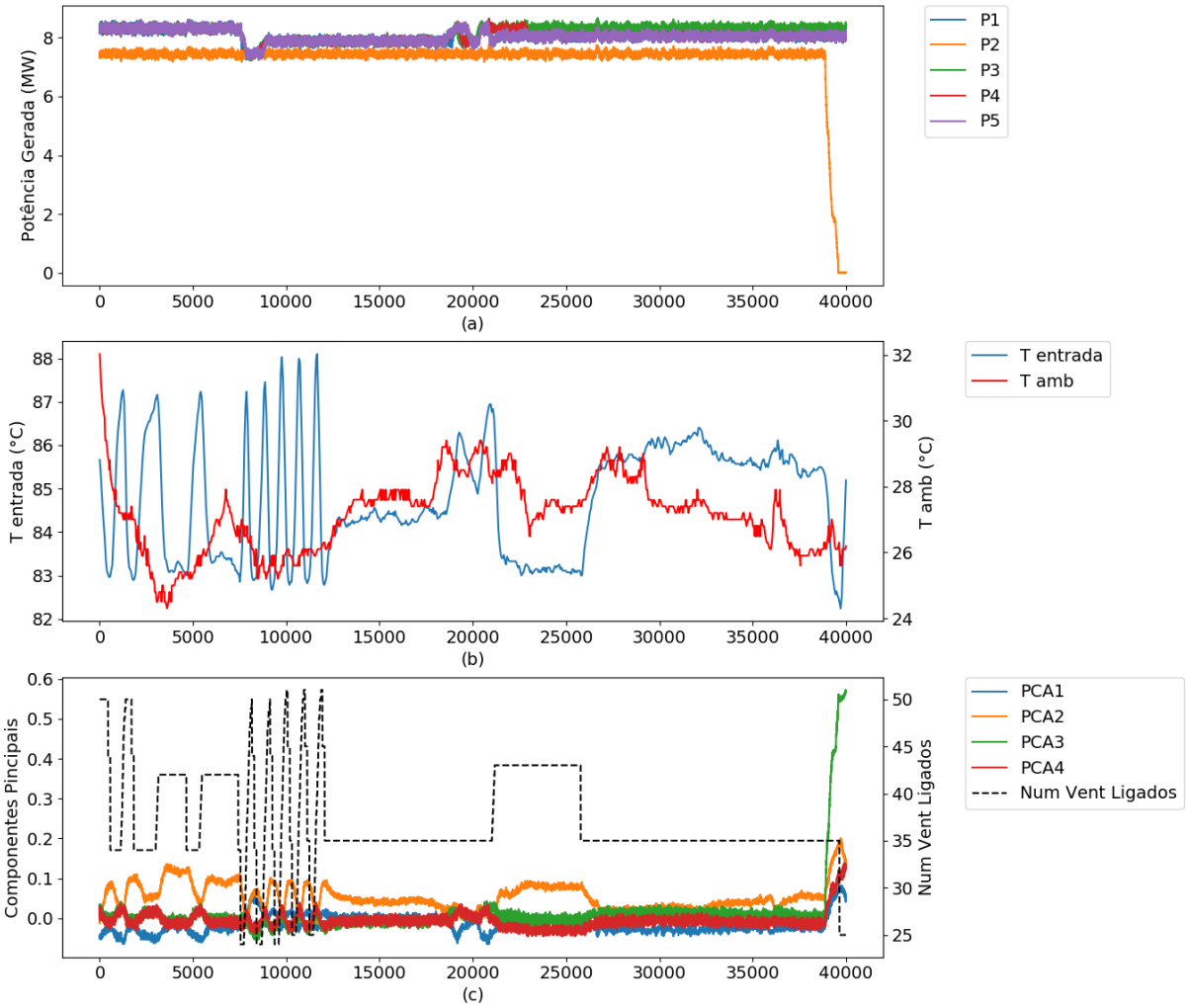


FIGURA 33 - ENTRADAS DO PROCESSO NO CENÁRIO 3, (A) POTÊNCIAS DAS UGD'S, (B) TENTHT E TAMB E (C) ENTRADAS REDUZIDAS E NUMVENTON.

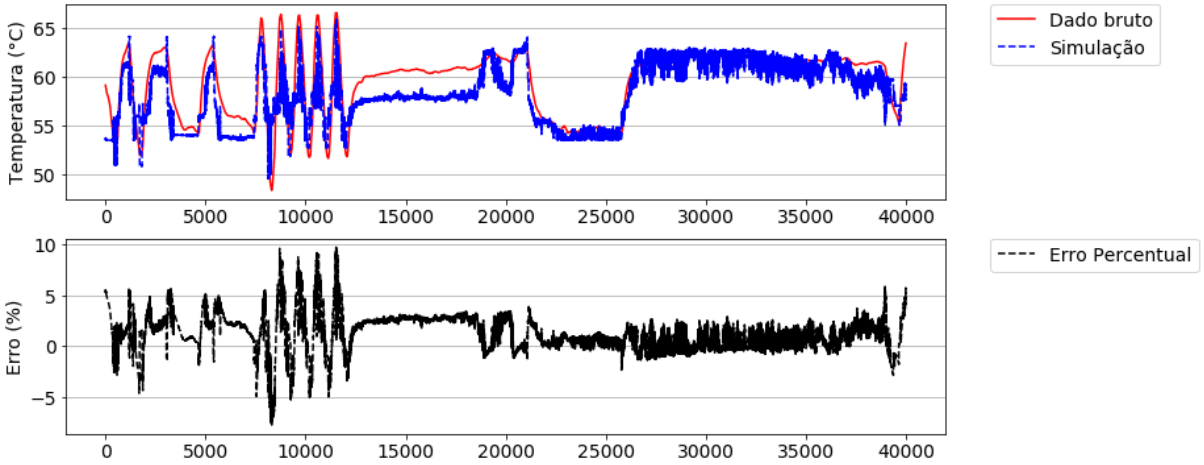


FIGURA 34 - RESULTADO MODELO X VALOR REAL TSAIDAHT E ERRO PERCENTUAL INSTANTÂNEO DO CENÁRIO 3.

4.4 OTIMIZAÇÃO DO PROCESSO A PARTIR DO GÊMEO DIGITAL

A partir do modelo, foi criado um outro sistema de inferência fuzzy Mamdani, sendo este mais simples e básico, a fim de otimizar o controle do acionamento dos ventiladores e obter um menor gasto energético do sistema mantendo a temperatura do processo conforme os parâmetros ditos pelo fabricante.

Dessa forma, foi adicionado um controlador fuzzy, em série com o gêmeo digital criado, o qual possui duas entradas e uma saída sendo como entradas o erro e a variação do erro e a saída um resultado que sugere a variação de ventiladores que deve ser feita para otimizar o processo. A Figura 35 ilustra o diagrama do gêmeo digital com o otimizador fuzzy. A Figura 37 ilustra os conjuntos fuzzy das variáveis do otimizador fuzzy.

Para o otimizador fuzzy, foi criado um banco de regras fixo com 49 regras do tipo “SE...Então” de forma a cobrir todas as possíveis situações. A Tabela 1 mostra a construção das 49 regras do otimizador fuzzy e na Figura 36 se ilustra a superfície gerada pelo sistema de inferência do otimizador, a qual reflete a capacidade de gerar valores contínuos das variações do número de ventiladores.

TABELA 3 - BANCO DE REGRAS FUZZY DO OTIMIZADOR.

δErro / Erro	NG	NM	NP	ZE	PP	PM	PG
NG	NMG	NG	NG	NM	NM	NP	ZE
NM	NG	NG	NM	NM	NP	ZE	PP
NP	NG	NM	NM	NP	ZE	PP	PM
ZE	NM	NM	NP	ZE	PP	PM	PM
PP	NM	NP	ZE	PP	PM	PM	PG
PM	NP	ZE	PP	PM	PM	PG	PG
PG	ZE	PP	PM	PM	PG	PG	PMG

TABELA 4 - LEGENDA DAS REGRAS FUZZY DO OTIMIZADOR.

NMG	Negativo Muito Grande
NG	Negativo Grande
NM	Negativo Médio
NP	Negativo Pequeno
ZE	Zero
PP	Positivo Pequeno
PM	Positivo Médio
PG	Positivo Grande
PMG	Positivo Muito Grande

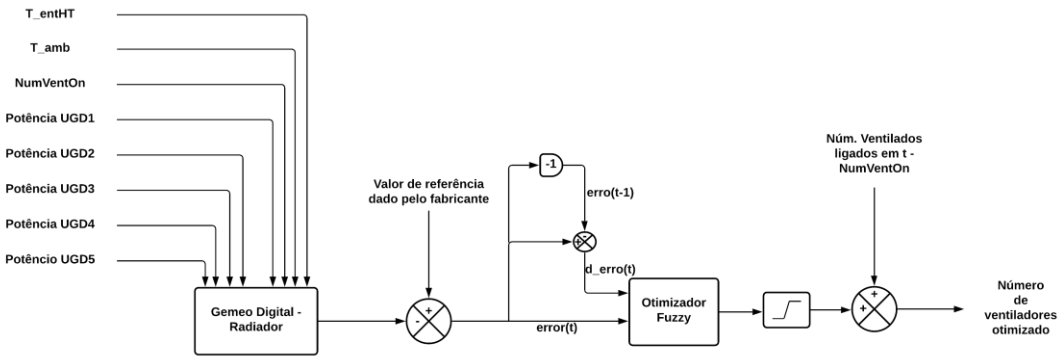


FIGURA 35 - PLANTA DO SISTEMA UTILIZADO PARA OTIMIZAR O NÚMERO DE VENTILADORES DO PROCESSO.

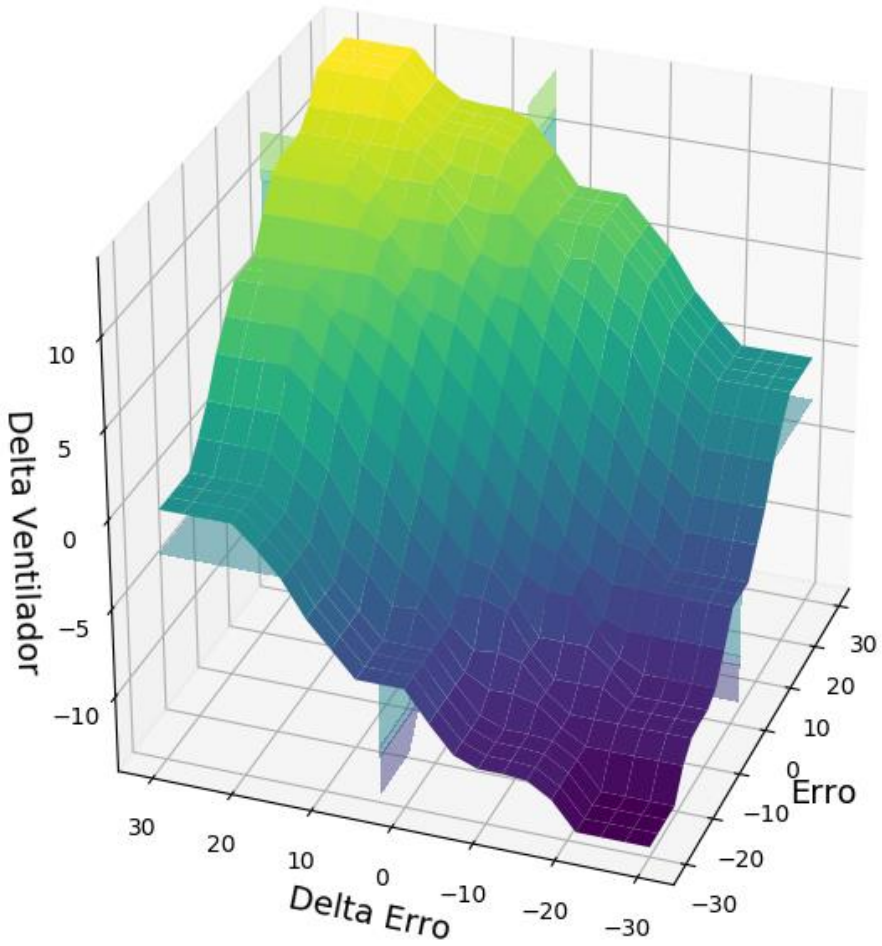


FIGURA 36 - SUPERFÍCIE 3D DO SISTEMA DE INFERÊNCIA DO OTIMIZADOR.

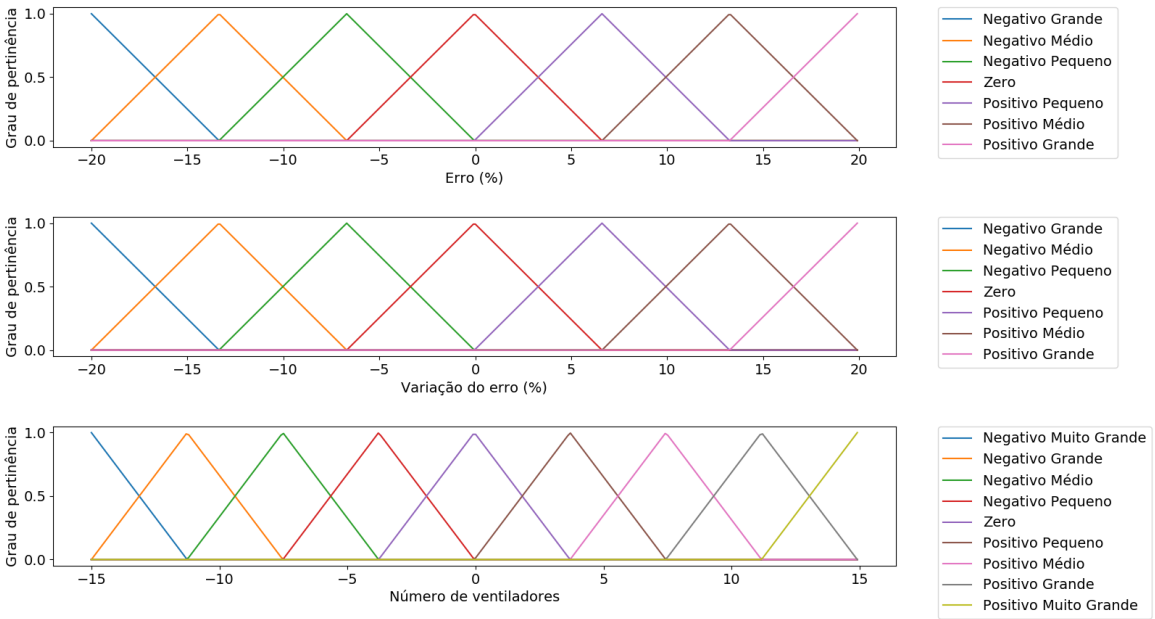


FIGURA 37 - CONJUNTOS FUZZY DO CONTROLADOR OTIMIZADOR DE VENTILADORES.

Foi utilizado o cenário 1 para verificar o funcionamento da planta completa. O valor utilizado como referência de temperatura na saída de HT do radiador é de 62°C, tal valor foi retirado do *setpoint* do controlador PID existentes nas unidades geradores para a válvula de três vias na entrada de água do motor. Pensou-se então que para otimizar o uso energético desse sistema, o radiador deverá operar de tal modo que a válvula de três vias de controle de temperatura não retornasse água da saída do motor para esquentar a água de entrada vinda do radiador, conforme pode ser visualizado na Figura 12. Para isso, o sistema com a planta completa fornecerá a sugestão de quantos ventiladores deverá desligar/ligar para que tal *setpoint* seja atingido. O otimizador opera de forma contínua, porém, como não é possível desligar/ligar ventiladores de forma fracionária, a sugestão também é dada na forma discreta. Na Figura 38, pode-se ver o resultado do modelo completo com o otimizador de ventiladores.

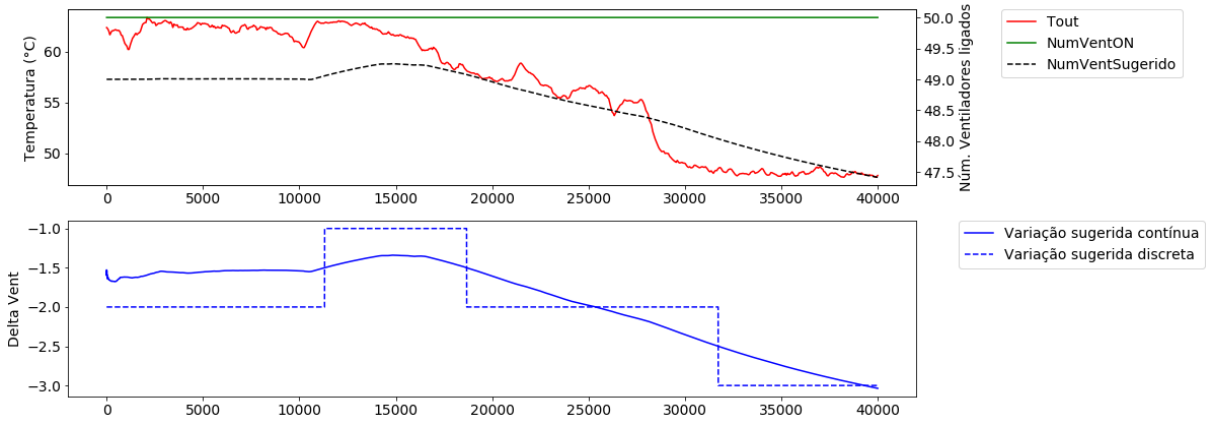


FIGURA 38 - RESULTADO DO OTIMIZADOR DE VENTILADORES PARA O CENÁRIO 1.

Para o cenário estudado, é possível perceber que a temperatura de saída de HT está baixa, mesmo assim o número de ventiladores extraído do histórico permanece inalterado, isso faz com que a válvula de três vias de controle de temperatura do motor faça a mistura da água quente advinda da saída do motor com a água de saída do radiador, desperdiçando energia gasta na troca térmica. A partir dos resultados apresentados é possível verificar que o sistema consegue operar a maior parte do tempo com três ventiladores a menos, cerca de 1,44MWh a menos de

energia por dia por radiador que pode ser gerado para o SIN, em valores monetários, cerca de R$ 1000,00 por dia por radiador de receita.

4.5 CONSIDERAÇÕES E DISCUSSÕES

A partir dos resultados demonstrados na seção anterior, foi feito um comparativo do desempenho do modelo nos cenários simulados, a Tabela 5 mostra essa comparação.

O erro em cenários onde a planta está em regime permanente é bastante baixo em sua maioria, porém, no cenário 1 verificou-se que a partir do instante 27000 o modelo não consegue buscar regras melhores, devido à solução otimização encontrada no AG nos formatos dos conjuntos fuzzy. Isso acontece pois o banco de treinamento utilizado no fitness do AG não possui situações semelhantes - os conjuntos foram otimizados para o regime permamente onde a temperatura ambiente é maior. Para melhorar tal ponto, é possível formar bancos de treinamentos maiores e com cenários mais críticos para a planta, a fim de otimizar os conjuntos para tal cenário, e assim, obter melhores resultados do modelo. Ainda assim, o desempenho do modelo apresentado foi bastante satisfatório na simulação deste cenário

Nos cenários com características mais transitórias ou de anomalias, o modelo operou de forma satisfatória, apresentando erros médios abaixo de 5% para o percentual, abaixo de 3°C para o absoluto e erro quadrático médio de 8. É perceptível que no cenário 3, mesmo com uma situação de oscilação e anomalia bastante brusca, o modelo operou com erros mais baixos do que os outros cenários, isso é devido ao aumento de conhecimento do banco de regras do modelo e ao fato de que o banco de treinamento utilizado no AG apresentava cenário com oscilação.

A abordagem de *digital twins* utilizando lógica fuzzy com extração automática de regras para a modelagem de sistemas se mostrou bastante interessante, pois apresentou algumas vantagens em relação a utilização de outras técnicas de inteligência artificial tal como redes neurais, como a necessidade de uma massa de dados de treinamento menor e a capacidade de atualizar a modelo sem a necessidade de retreinar todo o modelo. É possível utilizar o *digital twin* para diversos benefícios da planta industrial, tais como detecção de falhas preventivamente e otimização de parâmetros do processo, este demonstrado neste trabalho.

TABELA 5 - COMPARAÇÃO DOS ERROS ENTRE OS CENÁRIOS

	Cenário 1	Cenário 2	Cenário 2 – Autovalidação	Cenário 3
Erro médio percentual	5,12%	4,41 %	4,07%	2,18 %
Erro médio absoluto	2,76 °C	2,37 °C	2,18 °C	1,65 °C
Erro médio quadrático	10,06	7,9	7,18	4,45

5 CONCLUSÃO

5 CONCLUSÃO

Para demonstrar a capacidade de construção de um *digital twin*, foram utilizados os dados de uma usina termelétrica do sistema de refrigeração compartilhada com cinco unidades geradoras a combustão interna a diesel, com amostragem de um segundo e período equivalente a 11 dias de geração. Nesse período, a usina não operou de forma contínua e apresentava dados referentes a períodos de usina desligada como também transientes de usina em rampa.

Foi essencial o pré-processamento dos dados para não de extrair regras que possam prejudicar o modelo, assim, foram removidos os dados de usina desligada e aplicado um algoritmo de detecção de *outliers* baseado na média móvel.

Também foi utilizada a técnica de redução de dimensão por análise de componentes principais, a fim de otimizar o uso computacional do modelo e tornar o banco de regras menos denso, reduzindo cerca de 83% da quantidade de regras em comparação ao modelo sem a redução. A utilização de tais técnicas se mostrou importante na obtenção do resultado da ferramenta computacional, reduzindo o tempo necessário de treinamento do modelo durante a aprendizagem dinâmica, podendo também utilizar essa abordagem na criação de ferramentas para auxílio em tempo real da operação do sistema.

Foi feita a extração automática de regras seguindo a metodologia de Wang-Mendel, que foi capaz de minimizar o erro do modelo durante a execução em tempo real a partir de um algoritmo de atualização de regras, esse método se mostrou muito importante na mitigação do erro, tornando o modelo – “gêmeo digital” mais confiável e robusto em situações anormais do processo. O *digital twin* criado neste trabalho se mostrou bastante robusto e com uma qualidade interessante, a necessidade de poucos dados para se ter um modelo com resultados de erros percentuais igual ou abaixo de 5%. Nos testes foram utilizados um banco de treinamento inicial de 86700 amostras, equivalente a um dia de dados.

Por fim, foi utilizado um algoritmo otimizador no número de conjuntos fuzzy das variáveis de entrada e saída, o algoritmo de estratégias evolutivas, com o intuito de obter uma acurácia maior do modelo a partir da otimização da forma desses conjuntos. Tal algoritmo conseguiu reduzir ainda mais o número de regras com baixo impacto no erro do modelo. A determinação do número de conjuntos não foi

considerada no algoritmo genético porque tornaria o espaço de otimização muito abrangente tornando o processo de otimização muito longo e maçante. Assim, optou-se por uma análise da característica do erro quadrático em função dos números de conjuntos para limitar a área de busca da solução.

Mesmo com a utilização de várias estratégias de minimizar o número de conjuntos e de regras, não foi possível manter uma das principais características de um sistema fuzzy, a interpretabilidade. Inclusive a utilização de uma das estratégias para diminuir a dimensionalidade do sistema, a análise de componentes principais, teve efeitos negativos na interpretabilidade, em contrapartida, teve um impacto importante na redução do esforço computacional do modelo, parâmetro esse que será importante para a implementação real da ferramenta.

Como considerações para trabalhos futuros, tem-se:

- Aplicação de técnicas para melhorar a interpretabilidade do sistema fuzzy e utilizar o máximo do potencial da ferramenta [Ishibashi e Nascimento, 2012], [Guillaume,2001] e [Jin e Sendhoff, 2003];
- Aplicação do algoritmo genético de forma mais abrangente (otimização do número e formato dos conjuntos);
- Estudar um *set* de entradas e saída que minimize a dimensionalidade do sistema sem a utilização do PCA;
- Implementar a ferramenta na usina em modo sombra e verificar a eficácia.

REFERÊNCIAS

ANTONELLI, Michela et al. On the influence of feature selection in fuzzy rule-based regression model generation. Information Sciences, v. 329, p. 649-669, 2016.

ARMENDIA, Mikel et al. Twin-Control: A Digital Twin Approach to Improve Machine Tools Lifecycle. Springer Nature, p. 4, 2019.

BACK, Thomas; HAMMEL, Ulrich; SCHWEFEL, H.-P. Evolutionary computation: Comments on the history and current state. IEEE transactions on Evolutionary Computation, v. 1, n. 1, p. 3-17, 1997.

BONDARENKO, Andrey; ALEKSEJEVA, Ludmila. Methodology for Knowledge Extraction from Trained Artificial Neural Networks. Information Technology & Management Science (RTU Publishing House), v. 21, 2018.

CASTRO, Juan Luis et al. Extraction of fuzzy rules from support vector machines. Fuzzy Sets and Systems, v. 158, n. 18, p. 2057-2077, 2007.

CORDÓN, Oscar; HERRERA, Francisco; VILLAR, Pedro. Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base. IEEE Transactions on fuzzy systems, v. 9, n. 4, p. 667-674, 2001.

CRAVEN, Mark W.; SHAVLIK, Jude W. Using sampling and queries to extract rules from trained neural networks. In: Machine learning proceedings 1994. Morgan Kaufmann, 1994. p. 37-45.

DE ALMEIDA, Cássio; MACIEL, Vinicius; CYBIS, Luiz. Global warming potential assessment for operation of thermoelectric power plant in Manaus. Revista Latinoamericana en Análisis de Ciclo de Vida Latin-american Life Cycle Assessment Journal, p. 44, 2017.

DE ENERGIA ELÉTRICA, Câmara de Comercialização. CCEE. Relatórios do Reajuste do Custo Variável Unitário Revisado para o Programa Mensal da Operação, 07/07/2020.

DEMŠAR, Jure; LEBAR BAJEC, Iztok. Evolution of collective behaviour in an artificial world using linguistic fuzzy rule-based systems. Plos one, v. 12, n. 1, p. e0168876, 2017.

EIBEN, Agoston E.; SMITH, James E. Introduction to evolutionary computing. springer, 2015.

EPE – Expansão da geração. Disponível em: <https://www.epe.gov.br/pt/areas-de-atuacao/energia-eletrica/expansao-da-geracao/fontes>. Acessado em 30/07/2020.
TOLMASQUIM, Mauricio Tiomno et al. Energia termelétrica: gás natural, biomassa, carvão, nuclear. Rio de Janeiro: EPE, 2016.

FARSI, Maryam et al. (Ed.). Digital Twin Technologies and Smart Cities. Springer, 2020.

FU, LiMin. Rule generation from neural networks. IEEE Transactions on Systems, Man, and Cybernetics, v. 24, n. 8, p. 1114-1124, 1994.

FU, LiMin. Rule Learning by Searching on Adapted Nets. In: AAAI. 1991. p. 590-595.
KACIMI, Mohand Akli et al. New mixed-coding PSO algorithm for a self-adaptive and automatic learning of Mamdani fuzzy rules. Engineering Applications of Artificial Intelligence, v. 89, p. 103417, 2020.

GUILLAUME, Serge. Designing fuzzy inference systems from data: An interpretability-oriented review. IEEE Transactions on fuzzy systems, v. 9, n. 3, p. 426-443, 2001.

ISHIBASHI, Rogério; NASCIMENTO, Cairo Lúcio. Knowledge extraction using a genetic fuzzy rule-based system with increased interpretability. In: 2012 IEEE 10th

International Symposium on Applied Machine Intelligence and Informatics (SAMI). IEEE, 2012. p. 247-252.

JIN, Yaochu; SENDHOFF, Bernhard. Extracting interpretable fuzzy rules from RBF networks. Neural Processing Letters, v. 17, n. 2, p. 149-164, 2003.

KRUHOL, Mykola et al. Group Regulation Efficiency Analysis for Thermal Power Plant Auxiliaries. In: 2019 IEEE 6th International Conference on Energy Smart Systems (ESS). IEEE, 2019. p. 289-293.

LEE, C. S. G.; LIN, C. T. Neural fuzzy systems: a neuro-fuzzy synergism to intelligent systems. 1996.

NEGRI, Elisa; FUMAGALLI, Luca; MACCHI, Marco. A review of the roles of digital twin in cps-based production systems. Procedia Manufacturing, v. 11, p. 939-948, 2017.

NEISSE, Anderson Cristiano; HONGYU, Kuang. Aplicação de componentes principais e análise fatorial a dados criminais de 26 estados dos EUA. E&S Engineering and Science, v. 5, n. 2, p. 105-115, 2016.

ONS – Histórico DA OPERAÇÃO. Disponível em:<http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/geracao_energia.aspx>. Acessado em 30/07/2020.

PEREIRA, A. C.; ROMERO, Fernando. A review of the meanings and the implications of the Industry 4.0 concept. Procedia Manufacturing, v. 13, p. 1206-1214, 2017.

POURJAVAD, Ehsan; SHAHIN, Arash. The application of Mamdani fuzzy inference system in evaluating green supply chain management performance. International Journal of Fuzzy Systems, v. 20, n. 3, p. 901-912, 2018.

RHEINGANTZ, Paulo Afonso. Lógica fuzzy e variáveis linguísticas aplicadas na avaliação de desempenho de edifícios de escritório. Ambiente Construído, v. 2, n. 3, p. 41-55, 2002.

R. TANSCHEIT. Sistemas Fuzzy. VI Simpósio Brasileiro de Automação Inteligente (SBAI 2003), Minicurso (35 pgs.), Bauru, SP, setembro/2003.

SAHA, Balaram; PATEL, Vivek; CHATERJEE, Kalyan. Assessment of process parameter to improving power plant performance. In: 2014 Innovative Applications of Computational Intelligence on Power, Energy and Controls with their impact on Humanity (CIPECH). IEEE, 2014. p. 441-446.

TAN, Shing Chiang; LIM, Chee Peng. Application of an adaptive neural network with symbolic rule extraction to fault detection and diagnosis in a power generation plant. IEEE transactions on energy conversion, v. 19, n. 2, p. 369-377, 2004.

THRUN, Sebastian. Extracting provably correct rules from artificial neural networks. Sekretariat für Forschungsberichte, Inst. für Informatik III, 1993.

TICKLE, Alan B.; ORLOWSKI, Marian; DIEDERICH, Joachim. DEDEC: decision detection by rule extraction from neural networks. QUT NRC, 1994.

TOWELL, Geoffrey G.; SHAVLIK, Jude W. Extracting refined rules from knowledge-based neural networks. Machine learning, v. 13, n. 1, p. 71-101, 1993.

VILLANUEVA, Juan Moises Mauricio et al. Determinação de aproximação linear por partes de funções não lineares para sistemas embarcados utilizando algoritmos genéticos. 2005. Dissertação de Mestrado. Universidade Federal da Paraíba.

WANG, L.-X.; MENDEL, Jerry M. Generating fuzzy rules by learning from examples. IEEE Transactions on systems, man, and cybernetics, v. 22, n. 6, p. 1414-1427, 1992.

WANG, L.-X.; MENDEL, Jerry M. Generating fuzzy rules by learning from examples. IEEE Transactions on systems, man, and cybernetics, v. 22, n. 6, p. 1414-1427, 1992.

ZHOU, Keliang; LIU, Taigang; ZHOU, Lifeng. Industry 4.0: Towards future industrial opportunities and challenges. In: 2015 12th International conference on fuzzy systems and knowledge discovery (FSKD). IEEE, 2015. p. 2147-2152.