

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
FACULDADE DE ENGENHARIA
CAMPUS DE ILHA SOLTEIRA**

MARLEIDE FERREIRA ALVES

**PREVISÃO DE CARGAS NÃO RESIDENCIAIS MISTAS POR REDES NEURAIIS
*ARTMAP FUZZY***

Ilha Solteira
2019

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

MARLEIDE FERREIRA ALVES

**PREVISÃO DE CARGAS NÃO RESIDENCIAIS MISTAS POR REDES NEURAIS
ARTMAP *FUZZY***

Tese apresentada à Faculdade de Engenharia de
Ilha Solteira – UNESP como parte dos requisitos
para obtenção do título de doutora em Engenharia
Elétrica: Área de conhecimento: Automação

Anna Diva Plasencia Lotufo
Orientadora

FICHA CATALOGRÁFICA

Desenvolvido pelo Serviço Técnico de Biblioteca e Documentação

A474p Alves, Marleide Ferreira.
Previsão de cargas não residenciais mistas por redes neurais ARTMAP Fuzzy
/ Marleide Ferreira Alves. -- Ilha Solteira: [s.n.], 2019
90 f. : il.

Tese (doutorado) - Universidade Estadual Paulista. Faculdade de Engenharia
de Ilha Solteira. Área de conhecimento: Automação, 2019

Orientador: Anna Diva Plasencia Lotufo
Inclui bibliografia

1. Previsão de cargas elétricas em nível desagregado. 2. Rede neural
ARTMAP Fuzzy. 3. Metodologia híbrida.


Raiane da Silva Santos

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: Previsão de Cargas não Residenciais Mistas por Redes Neurais ARTMAP FUZZY

AUTORA: MARLEIDE FERREIRA ALVES

ORIENTADORA: ANNA DIVA PLASENCIA LOTUFO

Aprovada como parte das exigências para obtenção do Título de Doutora em ENGENHARIA ELÉTRICA, área: Automação pela Comissão Examinadora:


Profa. Dra. ANNA DIVA PLASENCIA LOTUFO
Departamento de Engenharia Elétrica / Faculdade de Engenharia de Ilha Solteira


Pós-doutorando MAURO DE SOUZA TONELLI NETO
Departamento de Engenharia Elétrica / Faculdade de Engenharia de Ilha Solteira


Profa. Dra. MARA LUCIA MARTINS LOPES
Departamento de Matemática / Faculdade de Engenharia de Ilha Solteira


Prof. Dr. KENJI NOZE FILHO
Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas / Universidade Federal do ABC


Prof. Dr. JOSÉ GUILHERME MAGALINI SANTOS DECANINI
Câmpus de Presidente Epitácio / Instituto Federal de São Paulo - IFSP

Ilha Solteira, 11 de março de 2019

DEDICATÓRIA

Dedico este trabalho à minha mãe e ao meu pai (*in memoriam*). Aos meus irmãos Jean e Douglas e aos meus sobrinhos Felipe e Lívia.

AGRADECIMENTOS

Agradeço primeiramente a Deus que está ao meu lado em todos os momentos.

À minha orientadora Anna Diva por todo apoio nesta etapa da minha vida, à toda a dedicação, carinho, amizade e paciência.

Aos meus colegas de laboratório, em especial ao Marquinhos, Thays e Mara que colaboraram com este trabalho e pela amizade maravilhosa.

À minha grande amiga de longa data Efigênia que sempre me apoia em tudo que faço.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

Os sistemas de energia elétrica estão passando por transformações. Aos poucos, técnicas de sistemas de informação estão sendo incorporadas aos sistemas atuais de energia. Basicamente este é o conceito de *smart grid*. Esta incorporação visa aumentar a eficiência dos sistemas de energia elétrica, pois os diversos agentes envolvidos em todo o sistema terão à disposição informações mais completas, precisas e de forma praticamente instantânea. Como consequência, haverá um aumento significativo de dados disponíveis para serem empregados de variadas formas. Um exemplo do uso de dados é a previsão de demanda de energia elétrica. De uma forma geral, previsões servem como suporte para suprir demandas, estimar custos ou justificar investimentos futuros. No campo de previsão de demanda de cargas elétricas existem diversos modelos na literatura, a grande maioria se concentra em níveis mais agregados, que atendem a grandes consumidores em que o fornecimento de energia é feito, por exemplo, por uma subestação. Uma *smart grid* também coloca à disposição as informações de consumo de energia em níveis cada vez menos agregados, como uma residência ou um prédio comercial. Realizar previsões neste nível é um desafio, pois essas demandas são muito influenciadas pelo comportamento humano. Diferentemente dos níveis mais agregados, modelos de previsão para níveis menos agregados, ou desagregados, ainda são poucos. O objetivo deste trabalho é fazer a previsão de cargas elétricas não residenciais mistas em curto prazo utilizando dados de demanda de cargas de uma universidade brasileira. A metodologia escolhida inicialmente foi a rede neural ARTMAP *Fuzzy* e depois foi desenvolvida uma metodologia híbrida entre a rede neural ARTMAP *Fuzzy* e o Método dos Mínimos Quadrados. Os bancos de dados disponíveis para o estudo continham muitos dados ruidosos e *outliers* além disso, havia pouca variabilidade de entradas, neste caso, somente o tempo (em intervalos de 15 minutos), o dia da semana e a carga elétrica. Foram desenvolvidas algumas estratégias para realizar a previsão. Os primeiros resultados das previsões com rede neural ARTMAP *Fuzzy* não foram muito satisfatórios, então verificou-se que havia pouca variedade nos dados. Para contornar esse problema, foram introduzidas os de valores de cargas máxima, mínima e média referentes à cada intervalo de tempo e dia da semana com o objetivo de suprir a pouca variabilidade de entradas. Depois de observado o comportamento da rede neural ARTMAP *Fuzzy* neste cenário foi desenvolvida uma metodologia híbrida entre a rede neural ARTMAP *Fuzzy* e o método dos mínimos quadrados, denominada rede neural ARTMAP *Fuzzy* – *MMQ*. Os testes foram realizados com e sem *outliers* nos conjuntos de dados. Para efeito de comparação, os testes foram feitos com o método dos mínimos quadrados, a rede neural ARTMAP *Fuzzy* e a metodologia híbrida. Além disso, foram feitas análises com dois valores distintos parâmetro de vigilância da rede neural e os valores da escolha das categorias na metodologia híbrida. Dentre todos os testes a metodologia proposta apresentou os melhores resultados.

Palavras-chave: Previsão de cargas elétricas em nível desagregado. Rede neural ARTMAP *Fuzzy*. Metodologia híbrida.

ABSTRACT

Electrical power systems are in transformation nowadays. Gradually, information system technology are being introduced to the energy systems. Basically, this is the concept of smart grid. This new concept aims to improve the efficiency of the energy systems, once the evolved agents will provide complete and precise information instantaneously. This way, a significant increase in data will be available to be employed in several forms. One example in using these data is electric energy demand forecasting. In general, predictions are support to provide electric load demand, estimate costs or justify future investments. Concerning electric load demand, there are several models in the literature, and the majority is concentrated in aggregated levels, attending large consumers, where, for example, the energy supply is provided by a substation. Considering the smart grid, there are consumption information in less aggregated levels as for example residences or commercial buildings. Therefore, realizing predictions in these levels (less aggregated) is a challenge, once the demand is influenced by the human behavior. The models for predicting loads in aggregated levels are common, in the contrary of less aggregated that are few. This work aims to predict short term mixed nonresidential electric loads using data from a Brazilian University. Firstly, Fuzzy ARTMAP Neural Network is chosen to execute the predictions, and afterwards a hybrid methodology containing Fuzzy ARTMAP and Square Minima Method. The dataset used for training contained several noises and outliers, besides few inputs were available, i.e. only time (15min), day of the week and the electric load. To overpass this problem, other information is added, like maximum load, minimum load and medium load referred to each time interval and day of week. The results obtained using the hybrid methodology with Fuzzy ARTMAP and Square Minima were better when compared to the original Fuzzy ARTMAP Neural Network. Analysis were also done varying the neural network parameters such as the vigilance and choice ones.

Keywords: Disaggregated electric load forecasting. Fuzzy ARTMAP Neural Network. Hybrid methodology.

LISTA DE FIGURAS

Figura 1	- Arquiteturas de Redes Neurais da Família ART.....	26
Figura 2	- Esquema Rede Neural ARTMAP Fuzzy	27
Figura 3	- Esquema da fase de diagnóstico da Rede Neural ARTMAP Fuzzy.....	32
Figura 4	- Funcionamento da ARTMAP Fuzzy Modificada – parte 01	34
Figura 5	- Funcionamento da ARTMAP Fuzzy Modificada – parte 02	34
Figura 6	- Funcionamento da ARTMAP Fuzzy Modificada – parte 03	35
Figura 7	- Ilustração treinamento ARTMAP Fuzzy -	40
Figura 8	- Dispersão de todos os elementos de W_b	49
Figura 9	- Dispersão das categorias conforme matriz B.....	49
Figura 10	- Esquema da disposição dos dados para o estudo.	51
Figura 11	- Funcionamento do movimento de janela no banco de dados na etapa de previsão de cargas.....	52
Figura 12	- Esquema da técnica movimento de janela ou janela deslizante.....	53
Figura 13	- Entradas antes da inclusão das cargas máxima, mínima e média.	55
Figura 14	- Entradas após a inclusão das cargas máxima, mínima e média.	56
Figura 15	- Perfil de cargas nível agregado – 1 semana.....	57
Figura 16	- Perfil de cargas nível desagregado – 1 semana.....	58
Figura 17	- Curvas de cargas do bloco 1 – (a) Segundas-feiras e (b) Sextas-feiras	59
Figura 18	- Curvas de cargas do bloco 2 – (a) Sextas-feiras (b) Sábados	59
Figura 19	- Curvas de cargas do bloco 2 – (a) Sextas-feiras (b) Sábados – depois da retirada dos <i>outliers</i> mais evidentes.	60
Figura 20	- Curvas de cargas do bloco 3 – (a) Segundas-feiras (b) Quartas-feiras	60
Figura 21	- Curvas de cargas do bloco 4 – (a) Segundas-feiras (b) Quintas-feiras	61
Figura 22	- Dispersão das categorias selecionadas em W_b	72

LISTA DE TABELAS

Tabela 1	- Valores do parâmetro de escolha de categorias de W_a	44
Tabela 2	- Valores de x_a de todas as categorias de W_a	45
Tabela 3	- Categorias J selecionadas x Categorias K associadas	46
Tabela 4	- Saídas obtidas pelas três metodologias e a saída real	48
Tabela 5	- Constituição dos dados iniciais obtidos pelo medidor	51
Tabela 6	- Dados de exemplo para adoção de outros dados para testes.....	54
Tabela 7	- Valores das cargas máxima, mínima e média- terças e quartas	55
Tabela 8	- MAPEs da primeira etapa de testes dos quatro blocos.	65
Tabela 9	- Resultados obtidos do bloco 01 – com <i>outliers</i>	66
Tabela 10	- Primeiros valores de x_a na escolha de categorias na previsão - $\rho_a = 0,95$	67
Tabela 11	- Quadro da frequência dos valores de $x_a - \rho_a = 0,95$	68
Tabela 12	- Quinze primeiros valores de x_a para a escolha de categorias para a fase de previsão na metodologia híbrida considerando os resultados para $\rho_a = 0,60$.	69
Tabela 13	- Quadro da frequência dos valores de x_a – referente a $\rho_a = 0,60$	69
Tabela 14	- Categorias criadas em ART_a no treinamento da RNA – com <i>outliers</i>	70
Tabela 15	- Frequência dos valores de ρ_a no treinamento do bloco 01 – com <i>outliers</i>	70
Tabela 16	- Resultados obtidos do bloco 01 – sem <i>outliers</i>	71
Tabela 17	- Médias MAPE do Bloco 01 – com <i>outliers</i> x sem <i>outliers</i>	72
Tabela 18	- Frequência dos valores de x_a – referente a $\rho_a = 0,95$ – sem <i>outliers</i>	73
Tabela 19	- Frequência dos valores de x_a – referente a $\rho_a = 0,60$ – sem <i>outliers</i>	73
Tabela 20	- Categorias criadas em ART_a no treinamento da RNA – sem <i>outliers</i>	74
Tabela 21	- Frequência dos valores de ρ_a no treinamento do bloco 01 – sem <i>outliers</i>	74
Tabela 22	- Resultados obtidos do bloco 02	75
Tabela 23	- Resultados obtidos do bloco 03	76
Tabela 24	- Resultados obtidos do bloco 04	76
Tabela 25	- Frequência de x_a no diagnóstico do bloco 02 – Híbrida – com <i>outliers</i>	84
Tabela 26	- Frequência de x_a no diagnóstico do bloco 02 – Híbrida – sem <i>outliers</i>	84
Tabela 27	- Frequência de x_a no diagnóstico do bloco 03 – Híbrida – com <i>outliers</i>	84
Tabela 28	- Frequência de x_a no diagnóstico do bloco 03 – Híbrida – sem <i>outliers</i>	84
Tabela 29	- Frequência de x_a no diagnóstico do bloco 04 – Híbrida – com <i>outliers</i>	85
Tabela 30	- Frequência de x_a no diagnóstico do bloco 04 – Híbrida – sem <i>outliers</i>	85
Tabela 31	- Frequência de x_a no diagnóstico do bloco 02 – Híbrida – com <i>outliers</i>	85
Tabela 32	- Frequência de x_a no diagnóstico do bloco 02 – Híbrida – sem <i>outliers</i>	85
Tabela 33	- Frequência de x_a no diagnóstico do bloco 03 – Híbrida – com <i>outliers</i>	86
Tabela 34	- Frequência de x_a no diagnóstico do bloco 03 – Híbrida – sem <i>outliers</i>	86
Tabela 35	- Frequência de x_a no diagnóstico do bloco 04 – Híbrida – com <i>outliers</i>	86
Tabela 36	- Frequência de x_a no diagnóstico do bloco 04 – Híbrida – sem <i>outliers</i>	86
Tabela 37	- Distribuição de frequência do parâmetro ρ_a bloco 02 – com <i>outliers</i>	87
Tabela 38	- Distribuição de frequência do parâmetro ρ_a bloco 02 – sem <i>outliers</i>	87
Tabela 39	- Distribuição de frequência do parâmetro ρ_a bloco 03 – com <i>outliers</i>	87
Tabela 40	- Distribuição de frequência do parâmetro ρ_a bloco 03 – sem <i>outliers</i>	87

Tabela 41	- Distribuição de frequência do parâmetro ρ_a bloco 04 – com <i>outliers</i>	88
Tabela 42	- Distribuição de frequência do parâmetro ρ_a bloco 04 – com <i>outliers</i>	88
Tabela 43	- Resultados obtidos do bloco 02	89
Tabela 44	- Resultados obtidos do bloco 03	89
Tabela 45	- Resultados obtidos do bloco 04	90

SUMÁRIO

1	INTRODUÇÃO.....	12
2	REVISÃO BIBLIOGRÁFICA.....	17
3	MÉTODOS UTILIZADOS.....	25
3.1	REDE NEURAL ARTMAP <i>FUZZY</i>	25
3.2	REDE NEURAL ARTMAP <i>FUZZY</i> MODIFICADA.....	33
3.3	MÉTODO DOS MÍNIMOS QUADRADOS - MMQ.....	36
4	METODOLOGIA.....	38
4.1	REDE NEURAL HÍBRIDA ARTMAP <i>FUZZY</i> - MMQ.....	38
4.2	CONSTITUIÇÃO E PREPARAÇÃO DOS DADOS.....	50
4.3	CARACTERÍSTICAS DOS DADOS.....	57
5	RESULTADOS.....	62
5.1	MÉTRICA PARA A AVALIAÇÃO DOS RESULTADOS.....	63
5.2	ESPECIFICAÇÃO DOS TESTES.....	64
5.3	RESULTADOS OBTIDOS.....	65
5.3.1	Resultados obtidos com a introdução das cargas máximas, mínima e média.....	65
5.3.2	Resultados obtidos com os conjuntos de dados do bloco 01 com <i>outliers</i>.....	66
5.3.3	Resultados obtidos com os conjuntos de dados do bloco 01 sem <i>outliers</i>.....	71
5.3.4	Resultados obtidos dos demais blocos com e sem <i>outliers</i>.....	75
6	CONCLUSÕES E SUGESTÕES DE TRABALHOS FUTUROS.....	78

6.1	CONCLUSÕES	78
6.2	SUGESTÕES DE TRABALHOS FUTUROS	79
	REFERÊNCIAS	80
	APÊNDICE A – Distribuição de frequência de xa – demais blocos	84
	APÊNDICE B - Distribuição de frequência de ρa – demais blocos	87
	APÊNDICE C – Resultados obtidos para $\rho a = 0,60$ – demais blocos	89

1 INTRODUÇÃO

Previsão de cargas elétricas é uma importante ferramenta na gestão dos sistemas de energia elétrica, como o planejamento estratégico, manutenção do sistema e investimentos. Há muitos trabalhos na literatura que abordam a questão de previsão de cargas e sua grande maioria se concentra em previsões de cargas de cidades ou regiões. Várias metodologias foram utilizadas para realizar esta tarefa ao longo dos anos, como modelos estatísticos, modelos inteligentes e até metodologias híbridas.

O avanço da tecnologia dos últimos anos está possibilitando a incorporação de tecnologia da informação nos atuais sistemas de energia com objetivo de torná-los mais eficientes. Essa incorporação está sendo em todo o sistema, desde a geração até os consumidores finais de forma gradativa aproveitando a rede existente. De forma bem simplificada esse é o conceito de *smart grid* (FARHANGI, 2010; GHARAVI; REZA, 2011). À medida que esse conceito é adotado, uma gigantesca quantidade de informações se torna disponível para serem utilizadas nos mais variados estudos para melhorar a eficiência dos serviços das concessionárias de energia elétrica.

A adoção de medidores inteligentes (*smart meters*) é parte integrante da *smart grid*. Esses medidores, uma vez instalados em unidades consumidoras, como residências, prédios públicos ou comerciais, fornecem informação de demanda em nível local, ou nível desagregado. Esse nível caracteriza-se pela demanda ser mais afetada pelo comportamento humano, como a presença de pessoas no local e a quantidade de aparelhos ligados (ABREU; ALVES *et al.*, 2015).

Como dito anteriormente, a grande maioria dos trabalhos se concentram em realizar previsões de perfis de carga envolvendo cidades ou regiões, com cargas da grandeza de gigawatts ou megawatts. Esses perfis se caracterizam pela união de vários consumidores de energia englobando os residenciais, comerciais, públicos. Esses são os perfis mais agregados de cargas elétricas. De acordo com (SEVLIAN; RAJAGOPAL, 2018) o efeito da agregação diminui a variabilidade dos perfis de carga resultando em curvas mais suaves e que são mais fáceis de prever.

Cargas elétricas em nível desagregado são mais variadas e por esse motivo realizar previsões neste nível é um desafio.

Por ser um campo de pesquisa relativamente novo, trabalhos abordando a questão de previsão de cargas elétricas em nível desagregados ainda são poucos, se comparados aos trabalhos publicados de previsão em níveis mais agregados.

As metodologias empregadas nos trabalhos encontrados de previsão em nível desagregados foram os modelos estocásticos como o modelo ARIMA (PENYA; BORGES *et al.*, 2011), o modelo SARIMA (MÜLLER, 2018), o modelo auto regressivo (AR) (FERNÁNDEZ; BORGES; PENYA, 2011; PENYA; BORGES *et al.*, 2011), o auto regressivo com variáveis exógenas (ARX) (TOURETZKY; PATIL, 2015), modelo auto regressivo não linear (NAR) (FERLITO; ATRIGNA *et al.*, 2015; RUIZ; RUEDA *et al.*, 2015), algoritmo de erro fixo (FERNÁNDEZ; BORGES; PENYA, 2011), modelo auto regressivo não linear com variáveis exógenas (NARX) (RUIZ; RUEDA *et al.*, 2015), regressão linear múltipla (AMBER; ASLAM; HUSSAIN, 2015; KE; JIANG; LU, 2016), regressão polinomial (FERNÁNDEZ; BORGES; PENYA, 2011; KE; JIANG; LU, 2016). Existem trabalhos que utilizam sistemas inteligentes como Adaptive Network based Fuzzy Inference Systems (ANFIS) (DEB, EANG, *et al.*, 2015), programação genética (AMBER, ASLAM e HUSSAIN, 2015), método físico estatístico (LÜ; LU *et al.*, 2015), rede neural de Wavelet auto recorrente (SRWNN) (CHITSAZ, SHAKER, *et al.*, 2015), máquina de vetor suporte (SVM) (FERNÁNDEZ; BORGES; PENYA, 2011), rede neural Elman (RUIZ; RUEDA *et al.*, 2015) rede neural Bayesiana (CHAE; HORESH *et al.*, 2016), perceptron multicamadas (PENYA; BORGES *et al.*, 2011; FERNÁNDEZ; BORGES; PENYA, 2011; ESCRIVÁ-ESCRIVÁ; ÁLVARES-BEL *et al.*, 2011; SCHACHTER; MANCARELLA, 2014; ROLDÁN-BAY; ESCRIVÁ-ESCRIVÁ *et al.*, 2013; GULIN; VASAK *et al.*, 2014; BAGNASCO; FRESI *et al.*, 2015; JETCHEVA, MAJIDPOUR; CHEN, 2014; MÜLLER, 2018), rede neural ARTMAP *Fuzzy* (MÜLLER, 2018).

Este trabalho tem como proposta realizar previsão de cargas elétricas não residenciais mistas em curto prazo, com horizonte de 24h à frente em nível desagregado. Os dados utilizados para este estudo são de quatro blocos de uma universidade brasileira. Eles foram obtidos com medidores instalados junto aos transformadores de cada um dos blocos. Os perfis são bem variados, que são compostos por salas de aula, prédios administrativos, laboratórios de pesquisa, restaurante universitário, cafeteria, agências de banco e correios. Além da variabilidade inerente aos dados desagregados os dados utilizados para este estudo possuem muitos *outliers* (cargas atípicas) e erros de leitura.

No início do trabalho a previsão de cargas foi feita utilizando somente a rede neural ARTMAP *Fuzzy*.

De maneira geral, a rede neural ARTMAP Fuzzy utiliza a semelhança entre seus pesos e o padrão de entrada e saída desejada para realizar seu treinamento. Cada padrão de entrada e saída desejada assimilado pertence a uma categoria que lhe que é mais semelhante. Sendo assim, o papel do treinamento desta rede é separar os padrões em categorias semelhantes.

Devido à variabilidade das cargas e aos *outliers* muitas categorias semelhantes nos padrões históricos de entrada se separaram por causa de saídas históricas bem distintas. Na fase de previsão a rede neural associa a categoria mais similar à entrada apresentada. Esta categoria, por sua vez, indica qual é a saída prevista.

Mas, esta escolha pode não ser a melhor. De fato, nas matrizes de pesos treinados relacionados com a saída desejada foram observadas que haviam saídas melhores se comparadas à saída obtida pela rede neural, mas estas não participaram da previsão.

Assim, para tentar solucionar esta questão, foi desenvolvida uma metodologia híbrida com a rede neural ARTMAP *Fuzzy* e o Método dos Mínimos Quadrados que foi denominada rede neural ARTMAP *Fuzzy* – MMQ.

Nesta metodologia, todo o treinamento é realizado pela rede neural ARTMAP Fuzzy. Na fase de diagnóstico a rede seleciona os pesos adaptados que são mais semelhantes com a entrada apresentada e estes são associados com os respectivos pesos adaptados para a saída desejada e a partir disso realizar a previsão pelo método dos mínimos quadrados.

A ideia da metodologia híbrida é tentar reunir as categorias de entrada mais semelhantes com suas respectivas categorias de saídas possíveis com o objetivo de reduzir o efeito da variabilidade dos dados, e a partir daí fornecer uma equação de regressão por mínimos quadrados para obter uma melhor saída de previsão de acordo com os pontos selecionados.

Uma das contribuições deste trabalho foi o emprego da rede neural ARTMAP *Fuzzy* no problema de previsão de cargas elétricas não residenciais em nível desagregado. Esta metodologia foi escolhida por ser de fácil implementação computacional e de aprendizagem rápida. Possui também as características de estabilidade e plasticidade que permite o aprendizado de novos padrões sem perder o conhecimento previamente adquirido.

(CARPENTER; GROSSBERG; ROSEN, 1991; CARPENTER; GROSSBERG, 1992; CARPENTER *et al.*, 1992).

Outra contribuição é uma análise do comportamento da rede neural ARTMAP *Fuzzy* neste cenário em relação à formação de categorias no treinamento e à escolha de categorias na fase de diagnóstico.

E por fim, o desenvolvimento de uma metodologia híbrida capaz de realizar previsões de cargas elétricas em nível desagregado, mesmo na presença de *outliers* e dados ruidosos.

O trabalho está dividido em seis partes, sendo que a primeira parte é esta introdução que é uma breve apresentação do trabalho, seus objetivos, motivação e os passos adotados para resolver o problema.

A segunda parte é a revisão bibliográfica que mostra os trabalhos da literatura com as mais diferentes metodologias aplicadas na previsão de cargas elétricas em curto prazo em nível desagregado de prédios não residenciais.

A metodologia adotada neste trabalho é apresentada na terceira parte. Primeiramente é apresentado o algoritmo da rede neural ARTMAP *Fuzzy* de Carpenter e Grossberg (1992). Em seguida o detalhamento de uma rede neural ARTMAP *Fuzzy* modificada desenvolvida por Moreno (2010). Esta modificação tem como objetivo tornar o treinamento mais rápido e as matrizes de pesos de tamanhos reduzidos. Nesta parte também é apresentada a metodologia híbrida, com exemplos, e como os dados foram preparados para realizar o treinamento.

A quarta parte foi reservada para a apresentação dos resultados. Primeiramente são mostradas as métricas para avaliação dos resultados e os ambientes de testes. Os primeiros resultados apresentados referem-se aos obtidos com a introdução das cargas máxima, mínima e média nos bancos de dados, cargas que foram introduzidas devido à pouca variedade de informação. Esses resultados são comparados aos resultados obtidos sem a introdução destas cargas.

Na quinta parte serão mostrados os resultados. Foram realizados treinamentos com dados com e sem *outliers*. Primeiramente são apresentados os resultados de previsão do bloco 01. A análise dos resultados deste bloco é a mais detalhada. Além de mostrar os resultados de previsão também serão feitas considerações em relação aos valores dos parâmetros observados

no treinamento e diagnóstico. E fechando esta parte serão mostrados os resultados dos demais blocos de modo sintético.

E a sexta e última parte as conclusões e sugestões de trabalhos futuros.

2 REVISÃO BIBLIOGRÁFICA

O objetivo deste capítulo é apresentar trabalhos de previsão de cargas elétricas não residenciais publicados na literatura.

Amber *et al.* (2015) utilizaram modelos de regressão múltipla e programação genética com cinco variáveis exógenas independentes para a previsão de consumo de energia para a administração de prédios do setor de educação superior do Reino Unido. O estudo procura identificar a relação entre consumo de eletricidade com os parâmetros meteorológicos e também com o tipo de dia (úteis, fins de semana e feriados). Destacam a importância da previsão para previsão orçamentária da compra de eletricidade e mostram alguns benefícios como a identificação da economia potencial de eletricidade. No estudo, os autores consideram como variáveis independentes, as variáveis climáticas e o indicador de tipo de dia da semana (dias úteis/ feriados). Ao longo do texto, os autores evidenciam através de gráficos as relações entre cada variável e o consumo de energia. Ambas as técnicas utilizadas tiveram bom desempenho com medidas de erro absoluto total (TAE) de 6% para a programação genética e de 7% para a regressão múltipla.

Amber *et al.* (2018) fazem um estudo para comparar a capacidade de previsão de cinco diferentes técnicas de sistemas inteligentes de um edifício administrativo de Londres. As técnicas utilizadas são regressão múltipla, programação genética, rede neural artificial, rede neural profunda e a máquina de vetor de suporte. Os testes foram feitos com dados medidos ao longo de cinco anos, sendo que quatro anos foram usados para treinamento e um ano para a previsão. A rede neural foi a metodologia que apresentou o melhor desempenho, com valor de erro médio percentual absoluto (MAPE) de 6%. A regressão múltipla teve MAPE de 8,5%, a programação genética 8,7%, a máquina de vetor de suporte 9% e a rede neural profunda 11%.

Amjady *et al.* (2010) propõem uma estratégia de previsão de carga de curto prazo para microgrids em dois níveis da universidade de Calgary no Canadá. Os autores ressaltam a volatilidade das cargas em microgrids inclusive utilizando métricas para fazer comparações entre as medidas de volatilidade das cargas em microgrids e a volatilidade de cargas em sistemas de energia convencionais. Na proposta dos autores um dos níveis (inferior) funciona como previsor de cargas, neste caso os autores utilizam uma metodologia híbrida de rede neural e algoritmo evolucionário. No outro nível (superior) é usada um algoritmo de evolução diferencial melhorado para a seleção de características. A ideia geral da proposta é alimentar o previsor com um subconjunto de entradas mais informativas e descartar as entradas que podem

prejudicar o desempenho do previsor. Os resultados obtidos foram comparados às previsões feitas com o modelo ARIMA, a transformada ARIMA wavelet, rede neural e algoritmo evolucionário e a rede neural wavelet adaptativa (AWNN). A metodologia proposta teve os melhores resultados em relação às outras metodologias citadas.

Bagnasco *et al.* (2015) fazem previsão de cargas de uma instalação hospitalar de Turin. Foi utilizada a rede neural Perceptron Multicamadas com o algoritmo do erro resiliente para a adaptação dos pesos. Os autores fazem uma análise detalhada e um procedimento formal para a seleção de todos os parâmetros da rede neural. Os dados utilizados para o treinamento foram cargas elétricas, tipo do dia (dia útil / feriado), hora do dia, dados meteorológicos. As simulações foram feitas para previsão de cargas elétricas para 15 dias à frente sob dois cenários: (1) sem dados de temperatura e (2) com os dados de temperatura incluídos. São feitas várias simulações com diferentes quantidades de neurônios na camada oculta da rede. Essas quantidades variaram entre 6 e 33 neurônios. Os valores de MAPE (Erro Percentual Absoluto Médio) ficaram entre 7,52% e 8,47% para as simulações sem dados de temperatura e entre 7,68% e 8,31% para as simulações com dados de temperatura.

O trabalho de Chae *et al.* (2016) foi previsão de cargas de prédios comerciais. Eles examinaram nove algoritmos diferentes com o auxílio do coeficiente de correlação entre as variáveis e uma medida de erro. Por fim escolheram o algoritmo de melhor desempenho que foi a rede neural bayesiana com algoritmo de treinamento de Levenberg–Marquardt (LM). Além disso, utilizaram o algoritmo de floresta randômica para selecionar as variáveis mais relevantes para compor os conjuntos de dados de treinamento. Os autores também testaram várias arquiteturas da rede para escolher a quantidade de neurônios na camada intermediária ideal para a previsão. Também fizeram testes com três tipos de janelamento de cargas. Os tipos são janelas estáticas, acumulativas e deslizantes. No geral, para o mês de agosto, os APEs (erros percentuais absolutos) dos dias úteis foram de 2,84% para o janelamento estático, 3,37% do acumulativo e de 3,30% para as janelas deslizantes. Para os fins de semana os valores foram de 8,77% para o estático, 7,28% para o acumulativo e 8,10% para as janelas deslizantes.

Chitsaz *et al.* (2015) propuseram uma Rede Neural Wavelet Auto Recorrente (SRWNN) com algoritmo de Levenberg–Maquardt para adaptação dos pesos para realizar a previsão de cargas de uma microrrede na *British Columbia Institute of Technology* (BCIT) em Vancouver. Esta microrrede possui módulos de energia renovável solar e eólica. Os autores fizeram uma análise do perfil de consumo desta microrrede comparando com o perfil de consumo com

sistemas de energia agregados que são de British Columbia e Califórnia. Nesta análise é mostrada a volatilidade do perfil de consumo na microrrede em relação aos sistemas de energia que são perfis mais suaves. Para comparação de resultados também foram feitas previsões com a Perceptron Multicamadas e com Rede Neural Wavelet. Os resultados do modelo proposto foram superiores aos resultados obtidos pelos outros modelos.

Deb *et al.* (2015) propuseram uma metodologia de previsão de carga de um sistema de resfriamento para três prédios do campus de Singapura, utilizando o sistema de interface adaptativa neuro *Fuzzy* (ANFIS) e uma rede neural *feedforward* com algoritmo de treinamento de regularização Bayesiana. Os dados foram divididos em cinco classes e foram feitos treinamentos de forma individual. Além disso, foram retirados dos dados de entrada as informações dos finais de semana, com a justificativa de serem dias com pouco consumo. Eles mediram a correlação entre a carga e as variáveis climáticas, temperatura do ar, umidade e radiação solar. Os resultados mostraram baixa correlação entre essas variáveis e a carga consumida. Para se ter uma ideia, a temperatura do dia foi a variável que apresentou maior correlação que foi 0,242. Os autores apenas fizeram o treinamento e apresentaram os resultados das correlações obtidas que ficou entre 0,96 e 0,98 nos testes usando o modelo ANFIS e 0,97 e 0,98 para a rede neural, que são correlações muito boas.

Escrivá *et al.* (2011) propõem a rede perceptron multicamadas para previsão de cargas de curto prazo do consumo total da Universidade Politécnica de Valência. A partir dos dados de consumo de toda a universidade, os autores realizaram previsões em separado de unidades consumidoras menores (UEs) levando em consideração que algumas delas possuem comportamento monótono e repetitivo. Os autores tratam cada UE como um processo independente com características próprias. O estudo dos perfis de consumo de cada uma das UEs contribui para fornecer informações detalhadas de cada uma delas. De acordo com os autores, a previsão de uma curva de carga base é uma informação importante para a tomada de decisão mais econômica por parte dos grandes consumidores na negociação de preços de compra de energia.

Ferlito *et al.* (2015) utilizam a rede neural auto regressiva não linear (NAR) para realizar previsão de médio e longo prazo de uma escola pública da cidade de Eboli – Itália. Foram feitas previsões de três, seis e doze meses à frente. Os autores fizeram testes com conjuntos de entrada de cinco, seis e sete dados de carga atrasados para prever a carga seguinte. O melhor resultado obtido para previsão de três meses foi 15,70% de RMSPE (raiz quadrada do erro quadrático

médio). Este resultado foi obtido com os conjuntos de entrada de sete cargas atrasadas. Já para seis meses, o melhor resultado foi de 14,53% de RMSPE com cinco cargas atrasadas. E por fim, um RMSPE de 14,59% na previsão de doze meses com seis dados de carga atrasados. Os autores também exibiram todas as medidas de correlação dos dados. O interessante é que uma maior correlação não significou menor erro de previsão, por exemplo, no caso da previsão de seis meses a medida de correlação foi de $R = 0,89$ para o treinamento com 5 dados de carga atrasados com um RMSPE 14,59%. Entretanto, a correlação para os testes de 7 meses atrasados foi de $R = 0,96$, mas o RMSPE foi de 17,97%.

Fernández *et al.* (2011) utilizaram vários modelos e compararam entre si para realizar a previsão de cargas de um prédio não residencial. Os modelos utilizados foram (1) modelo auto regressivo; (2) polinomial; (3) perceptron multicamadas; (4) máquina de vetor suporte. Os testes foram feitos levando em consideração o tipo de dia (dias úteis, finais de semana). Os treinamentos foram realizados com janelas de dados variáveis para avaliar o número de dias anteriores que otimizam a previsão. Também foi introduzido um algoritmo de pós-processamento que avalia o erro histórico e introduz um erro na saída de previsão para corrigir o erro histórico. Os autores fizeram simulações com quatro bancos de dados diferentes, dois foram da Universidade de Deusto e o restante de duas competições de previsão *Eunite* e *Ashrae*. Foram previsões de horizontes de um a seis dias à frente. Os melhores resultados foram com o modelo autorregressivo em todos os testes. A máquina de vetor de suporte também apresentou bons resultados, com pouca diferença aos resultados apresentados pelo modelo autorregressivo.

Fujiwara e Ueda (2018) propuseram a previsão baseado em horas de trabalho. Usaram o método de Ward para separar as cargas de acordo com o valor. A primeira separação avalia se o dia é de trabalho ou não. Caso o dia seja de trabalho as cargas são clusterizadas de acordo com os valores e as horas de trabalho são determinadas. Além disso, o fluxograma da proposta dos autores conta com uma metodologia de correção de erro. Foram feitas previsões da carga máxima pela análise de regressão múltipla e regressão robusta de vários tipos de perfis de carga. Para comparação dos resultados foi usado um modelo de persistência simples. As previsões foram feitas levando em consideração diferentes perfis de cargas e horizontes. A metodologia proposta foi superior em todos os resultados.

Gao (2018) fizeram testes de previsão de um edifício público com a rede neural perceptron multicamadas com algoritmo *backpropagation* para adaptação dos pesos. Foi adotado um algoritmo evolutivo para otimizar o treinamento da rede com o objetivo de tornar

a convergência mais rápida, eficiente evitando mínimos locais. Também foram feitas análises de correlação entre o consumo de energia e a temperatura, umidade, número de pessoas, tempo de trabalho, concentração de CO₂ e concentração de material particulado fino (PM 2.5). Dentre todas essas variáveis a que apresentou correlação mais baixa foi a PM 2.5 que foi de 0,087. O número de pessoas apresentou a correlação mais alta, que foi de 0,911 e correlação das demais variáveis ficou entre 0,446 e 0,569. Os resultados foram comparados aos testes feitos com a rede neural perceptron multicamadas sem a otimização. De acordo com os autores, o algoritmo evolutivo tornou a rede neural mais eficiente e precisa ao ser comparado aos resultados obtidos pela rede neural sem a otimização.

Gulin *et al.* (2014) utilizam uma heurística para seleção dos dados de entrada de uma rede perceptron multicamadas com o algoritmo de adaptação dos pesos de Levenberg-Maquardt, para realizar a previsão de carga com objetivo de otimizar o fluxo de energia em microgrid da Faculdade de Computação e Engenharia Elétrica da Universidade de Zagreb na Croácia. Foram utilizados dados meteorológicos, tempo e de carga. As variáveis de tempo foram representadas em funções periódicas de seno e cosseno com objetivo de evitar discontinuidades, como por exemplo depois das 23:00 vem 00:00. Foram feitos testes com a seleção das variáveis mais relevantes e também com todos os dados disponíveis. A rede neural treinada com todos os dados foi a que apresentou o melhor desempenho.

O trabalho de Jetcheva *et al.* (2014) é realizar a previsão de cargas elétricas de seis edifícios. São fábricas e edifícios de escritórios comerciais. A estratégia dos autores é utilizar um conjunto de redes neurais perceptron multicamadas, no qual subconjuntos de dados são treinados em redes separadas. Estes subconjuntos foram obtidos por clusterização por meio do algoritmo *k-means*. O resultado é um conjunto de redes neurais, cada uma com melhor desempenho em uma hora determinada do dia. Os resultados obtidos são comparados ao modelo SARIMA. De modo geral a metodologia proposta supera o modelo SARIMA. Para alguns edifícios a diferença entre os resultados é bem pequena enquanto para outros a diferença é bastante significativa, chegando a uma melhora de resultado de até 50% para a metodologia proposta pelos autores.

Jozi *et al.* (2018) fizeram previsões de carga de uma microrrede piloto de vários prédios do campus da Escola Superior de Engenharia da Politécnica do Porto. A metodologia adotada foi a máquina de vetor de suporte (SVM). Além dos dados de consumo histórico, os autores usaram variáveis exógenas como temperatura, luminosidade e sazonalidade. Nesta microrrede

foram obtidos registros de consumo dos sistemas de ar-condicionado, iluminação e tomadas em separado. Foram adotadas duas estratégias para a previsão. A primeira foi fazer a previsão de cada sistema em separado e a segunda estratégia é incluir em dois, dos três sistemas uma terceira variável. No sistema de ar-condicionado foi incluída os valores de temperatura ambiente e no sistema de iluminação a informação de radiação solar. Os resultados foram comparados a sistemas híbridos de inferência *fuzzy* (HyFIS), método de aprendizagem de regras *Fuzzy* (WM) e sistema *fuzzy* genético para aprendizado de regras *fuzzy* baseado na metodologia MOGUL (GFS.FR.MOGUL). Em ambas as estratégias o modelo adotado superou as outras metodologias. O valor do erro percentual médio (MAPE) da previsão da SVM foi de 9,44% para a primeira estratégia enquanto as outras metodologias apresentaram valores entre 10,01% e 18,84%. Na segunda estratégia o valor do MAPE de previsão da SVM foi de 9,11% e das outras metodologias ficou entre 10,01% e 13,16%.

Ke *et al.* (2016) apresentam métodos para construção de carga, análise de perfil e previsão usando dados em intervalos de 15 minutos, coletados da subestação que fornece energia ao campus Centenário da Universidade Estadual da Carolina do Norte. Foram feitos estudos para identificar assinaturas de consumo de energia no prédio. Ao todo, analisaram quatro assinaturas. A primeira foi detectar a relação entre o consumo e as cargas máxima, média e mínima diária. A segunda assinatura foi medir a diferença entre o consumo em um determinado instante e o instante anterior. Essa diferença serve para verificar o quanto a carga varia num intervalo de quinze minutos. A terceira assinatura é a correlação entre temperatura e carga e a última a correlação entre umidade e carga. As revisões de carga foram feitas utilizando três algoritmos: ajuste de curva direta por regressão polinomial, aproximação por dia similar e regressão linear múltipla. Em todo os modelos foram introduzidas medidas de erro de previsão para melhorar a precisão dos resultados. O modelo polinomial escolhido foi um modelo de segunda ordem. Os melhores resultados foram os obtidos pelo algoritmo de aproximação por dia similar e regressão linear múltipla.

A proposta de Müller (2018) foi avaliar o desempenho da rede neural ARTMAP *Fuzzy* considerando dados de consumo de energia elétrica em diferentes resoluções e níveis de agregação, comparando com outras duas metodologias, o modelo SARIMA e rede neural perceptron multicamadas. Nos testes realizados, o autor utiliza o filtro SSA (*Singular Spectrum Analysis*) para remoção de ruídos dos dados utilizados. Foram feitas análises com relação aos parâmetros da rede neural ARTMAP *Fuzzy*. Este trabalho evidencia que quanto maior o nível de agregação, mais assertiva é a previsão.

Penya *et al.* (2011) realizaram previsões de curto prazo de cargas de ar-condicionado, aquecimento e ventilação (HVAC) de prédios não residenciais utilizando quatro modelos: (1) modelo auto-regressivo, (2) ARIMA, (3) rede neural perceptron multicamadas (4) rede Bayesiana. Para evitar a não linearidade utilizam o dia de trabalho como classificador do tipo do dia. Como o consumo de equipamentos HVAC é influenciado pelas condições climáticas, os modelos de previsão utilizam também variáveis meteorológicas como entrada. Inclusive os autores apresentam um estudo da influência deste tipo de variável na previsão. Foram feitos treinamentos com dois conjuntos de dados. Um dos conjuntos com cargas mais regular e outro menos regular. As previsões foram feitas para 24 e 48 horas à frente. De acordo com os autores o modelo auto-regressivo apresentou os melhores resultados no geral, mas de acordo com os quadros apresentados a rede neural teve melhor desempenho.

Roldán-Blay *et al.* (2013) fazem uma modificação no trabalho de Escrivá *et al.* (2011). A modificação está em utilizar a rede neural perceptron multicamadas para a previsão de cargas da Universidade de Valência. Os autores fizeram os treinamentos separando os dados de consumo de aparelhos que são fortemente dependentes de temperatura, como ar-condicionado, ventilação e aquecimento, e os dados de consumo de aparelhos que não são, como computadores. No trabalho foi utilizado um modelo que simula tempo em temperatura (TTC) bem preciso para realizar a previsão de temperatura. Ao invés de usar medidas de temperaturas máximas, médias e mínimas diárias, foi utilizado valores horários de temperatura. Toda essa informação serviu como entrada para uma rede neural perceptron com *backpropagation*. Os autores comparam os resultados com outros dois modelos, um que utiliza valores constantes de temperatura e o outro que utiliza somente dados de cargas locais. No geral, a metodologia proposta apresentou os melhores resultados.

Ruiz *et al.* (2015) utilizaram uma rede neural Elman para previsão de cargas elétricas em um edifício público em conjunto com o algoritmo genético *Cross generational elitist selection, heterogeneous recombination and Cataclysmic Mutation* (CHC) para melhorar o treinamento para reduzir o tempo de convergência da rede neural e evitar mínimos locais. Devido às falhas na coleta dos dados os autores utilizaram dois métodos de tratamento de dados, o filtro de média móvel e a técnica de *sliding windows*, para eliminar ruídos e dados faltantes. Os testes foram feitos com os dados com e sem valores de temperatura. Para comparar os resultados, os autores usaram os modelos NAR (modelo auto regressivo não linear) e NARX (modelo auto regressivo não linear com variáveis exógenas). Cada modelo foi testado cinco vezes. Em todos os experimentos a rede neural Elman com dados de temperatura tiveram

resultados superiores se comparados aos obtidos pelas outras arquiteturas e à rede Elman sem os dados de temperatura.

Schachter e Mancarella (2014) utilizam a rede perceptron multicamadas para um modelo de previsão de demanda de curtíssimo prazo de quatro prédios residenciais no Reino Unido e compara os resultados com a regressão linear. O algoritmo de treinamento escolhido foi o Levenberg-Maquardt. Os treinamentos foram feitos com dados de cargas de bomba de calor e do sistema de ar-condicionado, aquecimento e ventilação (HVAC). Além dos dados de carga históricos também serviram de entrada data, instante que a carga foi medida, dados meteorológicos. Foi feito um estudo de correlação linear entre as variáveis de tempo e meteorológicas e o consumo de carga. Em todos os casos a correlação foi muito baixa, mostrando a baixa dependência linear entre essas variáveis. Os autores fizeram testes com diferentes quantidades de camadas escondidas na rede neural perceptron multicamadas. Também foram feitos testes com diferentes valores de carga atrasados e também com e sem os dados meteorológicos. No geral, ambos os modelos apresentaram bom desempenho com valores de erros abaixo de 10%. É importante observar que as diferentes formas que os dados foram usados para treinamento influenciou de forma muito significativa nos resultados.

Touretzky e Patil (2015) propõem um modelo para previsão de demanda de um prédio comercial. A proposta é o modelo auto regressivo com entradas exógenas (ARX) utilizando variáveis locais como entrada, além disso trabalha em conjunto com a modelagem física existente identificando e incorporando medições como entrada e avaliando a previsão do modelo para supervisionar e controlar a resposta da demanda. A principal proposta dos autores é um modelo que possa ser usado em qualquer tipo de prédio comercial, ou seja uma solução mais genérica para previsão em prédios comerciais. Segundo eles, as soluções encontradas na literatura são soluções para tipos específicos de edifícios residenciais.

3 MÉTODOS UTILIZADOS

Nesta parte serão apresentadas as técnicas da literatura necessárias para a realização deste trabalho, que são a rede neural ARTMAP *Fuzzy*, rede neural ARTMAP *Fuzzy* modificada e o Método dos Mínimos Quadrados (MMQ).

3.1 REDE NEURAL ARTMAP FUZZY

Redes neurais artificiais (RNA) ou simplesmente redes neurais são estruturas computacionais que têm como objetivo simular o comportamento do cérebro humano (HAYKIN, 2001).

Possuem a capacidade de aprender com a experiência armazenando conhecimento através dos seus pesos sinápticos. Têm a vantagem de lidar com problemas complexos, não lineares e em grande escala. Além disso, a capacidade de generalização do conhecimento que permite respostas adequadas a entradas desconhecidas da rede neural em sua fase de diagnóstico. (HAYKIN, 2001).

A história das redes neurais artificiais se inicia em 1943 com o primeiro modelo de neurônio artificial que foi desenvolvido por Warren McCulloch e Walter Pitts. Este modelo é conhecido como o neurônio de McCulloch-Pitts (HAYKIN, 2001). Anos depois, em 1958, Frank Rosenblatt propõe o primeiro modelo de rede neural com aprendizado supervisionado, a rede *Perceptron* (HAYKIN, 2001). Daí em diante diversas arquiteturas de redes neurais foram desenvolvidas.

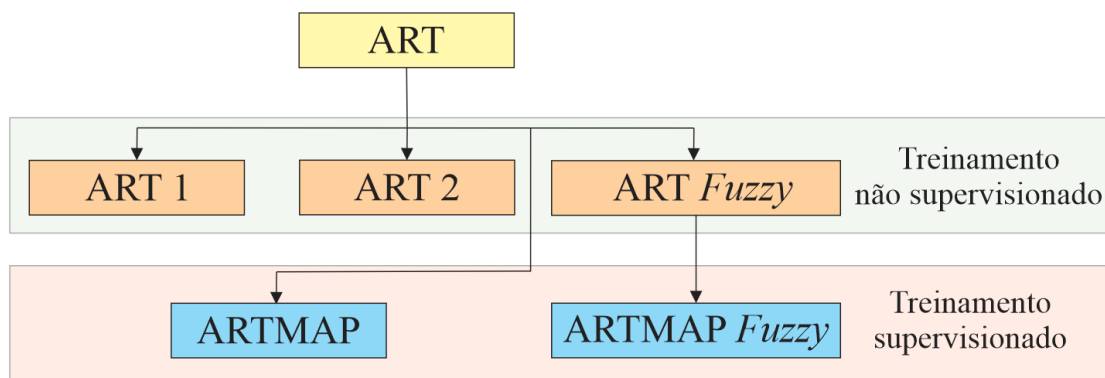
Uma arquitetura de rede neural tem a capacidade de adquirir conhecimento. O processo de adquirir esse conhecimento é chamado de aprendizagem ou treinamento e o resultado do aprendizado é armazenado nos pesos sinápticos da rede neural. De uma maneira geral, o processo de aprendizagem tem por objetivo a adaptação destes pesos. Cada tipo de arquitetura possui um algoritmo próprio de adaptação (HAYKIN, 2001).

Independente das peculiaridades de cada arquitetura de rede neural, os algoritmos de treinamento podem ser divididos em dois grupos principais: algoritmos com treinamento supervisionado e não supervisionado. O primeiro é feito com o auxílio de um professor, ou seja, o conjunto de padrões de entrada para treinamento possui um conjunto de saída desejada, que é o professor ou supervisor. Portanto, os pesos sinápticos de redes neurais que utilizam esse

tipo de treinamento são adaptados de acordo com a saída desejada (HAYKIN, 2001). Já o treinamento não supervisionado não conta com a ajuda de um professor para ajustar os pesos. RNAs com treinamento não supervisionado têm uma regra para ajustar seus pesos, como a regra de aprendizagem competitiva, em que os neurônios disputam entre si a oportunidade de ter seus pesos adaptados (HAYKIN, 2001).

Dentre as arquiteturas destacam-se as redes neurais da família ART (*Adaptive Resonance Theory*) que foram introduzidas inicialmente por Grossberg (1976). O desenvolvimento destas redes foi baseado na resolução do dilema plasticidade e estabilidade. A plasticidade está relacionada com o fato dos sistemas biológicos adquirirem novos conhecimentos sem perder o que foi aprendido anteriormente. Já a estabilidade assegura que todos os padrões serão agrupados em classes (LOPES, 2005). Os principais modelos da família ART são: ART1, ART2, ART *Fuzzy*, estas com treinamento não supervisionado, ARTMAP e ARTMAP *Fuzzy*, com treinamento supervisionado (LOPES, 2005). Um pequeno esquema dessas arquiteturas é mostrado na Figura 1.

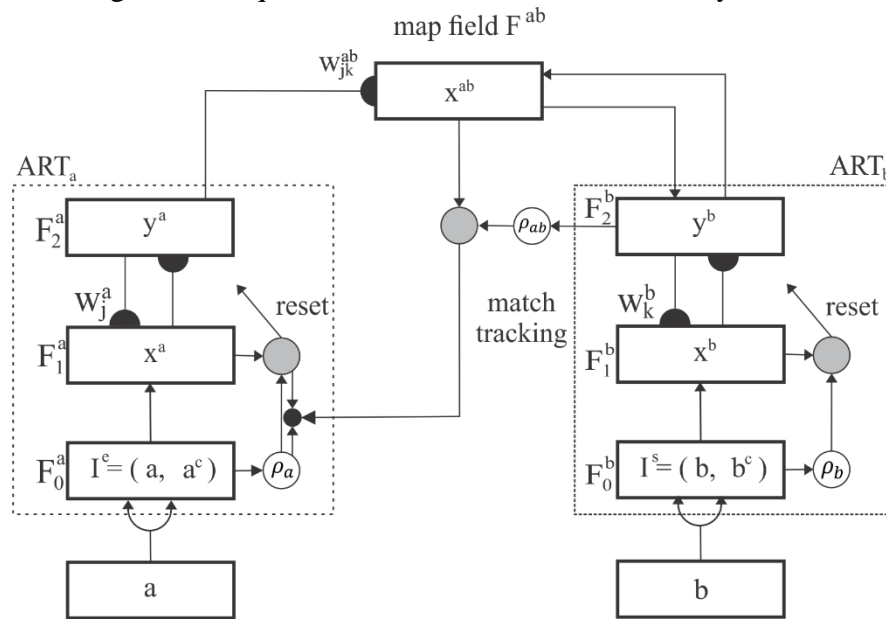
Figura 1 – Arquiteturas de Redes Neurais da Família ART



Fonte – adaptação (SANTOS JÚNIOR, 2017)

A ARTMAP *Fuzzy* foi desenvolvida por Carpenter *et al.* (1992). É uma rede neural de treinamento supervisionado que tem dois módulos ART e um módulo interART. O módulo ART_a é responsável pelo processamento dos padrões de entrada e o módulo ART_b pelo processamento dos padrões da saída desejada. Já o módulo interART realiza o casamento entre esses dois módulos (CARPENTER *et al.*, 1992).

Figura 2 - Esquema Rede Neural ARTMAP Fuzzy



Fonte: Adaptação (CARPENTER *et al.*, 1992)

A Figura 2 é uma ilustração da arquitetura desta rede. Os blocos à esquerda representam o módulo ART_a e à direita os módulos ART_b . Ambos os lados possuem três camadas:

- Camada F_0 – é onde ocorre o pré-processamento dos dados. Nesta camada os dados são normalizados e é feita a codificação complementar que tem por objetivo evitar a proliferação de categorias (CARPENTER *et al.*, 1992).
- Camada F_1 – nesta camada ocorre a escolha das categorias. Uma vez que uma categoria é escolhida ela é avaliada pelo parâmetro de vigilância ρ_a no módulo ART_a e ρ_b no módulo ART_b . Caso a categoria escolhida não seja aceita, ocorre o *reset* e uma nova categoria é escolhida, caso não haja nenhuma categoria que satisfaça ρ , uma nova é criada. O parâmetro de vigilância ρ mede o grau de confiança mínimo que uma categoria deve ser aceita. (CARPENTER *et al.*, 1992)
- Camada F_2 – Se em cada módulo, uma categoria escolhida for confirmada pelos parâmetros de vigilância ρ_a e ρ_b os pesos sinápticos associados à categoria escolhida são adaptados em ART_b . Já a categoria escolhida em ART_a será avaliada no módulo interART pelo processo de *match tracking* através do parâmetro ρ_{ab} . Caso haja confirmação de ρ_{ab} os pesos sinápticos associados à categoria aceita têm seus pesos adaptados, em ambos os módulos ART_a e inter-ART. Caso contrário, ocorre o reset, o parâmetro ρ_a é incrementado,

aumentando sua exigência na avaliação de categoria e uma nova é escolhida em ART_a , ou criada se não for encontrada nenhuma categoria que satisfaça ρ_{ab} . (CARPENTER *et al.*, 1992)

Os parâmetros de vigilância ρ_a e ρ_b são determinantes na formação das categorias em seus respectivos módulos, ou seja, são esses parâmetros que dizem se uma determinada entrada avaliada pertence ou não à uma categoria escolhida. Já o parâmetro ρ_{ab} , do módulo inter-ART, determina se a categoria escolhida em ART_a é adequada à categoria escolhida em ART_b (CARPENTER *et al.*, 1992).

A escolha dos valores dos parâmetros de vigilância ρ_a e ρ_b é uma questão importante. De acordo com Carpenter *et al.* (1992) valores baixos desses parâmetros leva a uma maior generalização, maior compreensão do código e criação de menos categorias. Por outro lado, valores cada vez maiores destes parâmetros conduzem à maior diferenciação na formação de categorias.

É importante observar que a rejeição de uma categoria pelo processo de *match tracking*, no módulo interART também colabora com formação de categorias em ART_a . De fato, se o processo de *match tracking* rejeitar uma categoria avaliada e aceita em ART_a , uma nova categoria será criada para tentar satisfazer o *match tracking*.

A descrição feita nos parágrafos anteriores da rede ARTMAP *Fuzzy* foi realizada de forma bem simplificada, objetivando o entendimento geral de seu mecanismo de funcionamento. Os passos para a implementação computacional, ou seja, o algoritmo, serão feitos nos parágrafos seguintes.

Antes de iniciar o treinamento é necessário um pré-processamento dos dados, a definição dos parâmetros iniciais e a inicialização dos pesos sinápticos da rede.

Na parte de pré-processamento de dados é realizada a normalização de acordo com as equações (1) e a codificação complementar, equações (2). A codificação de complemento evita a proliferação de categorias (CARPENTER *et al.*, 1992).

$$\bar{a} = \frac{a}{|a|} \qquad \bar{b} = \frac{b}{|b|} \qquad (1)$$

$$\bar{a}_i^c = 1 - a_i \qquad \bar{b}_i^c = 1 - b_i \qquad (2)$$

em que a e b são, respectivamente, os padrões de entrada e de saída desejada da rede, e

$$|a| = \sum_i |a_i| \quad (3)$$

O conjunto de padrões de entrada da rede para treinamento é dado por $I^e = [a \bar{a}]$ de dimensão $n \times p$ e o conjunto de padrões de saída desejada por $I^s = [b \bar{b}]$ de dimensão $n \times q$. Em que n é a quantidade de entradas e p e q são, respectivamente, a quantidade de elementos dos padrões de entrada e saída desejada.

Os parâmetros iniciais da rede são:

- $\alpha > 0$: Parâmetro de escolha da categoria.
- ρ_a : Critério de vigilância do módulo ART_a , $\rho_a \in [0,1]$.
- ρ_b : Critério de vigilância do módulo ART_b , $\rho_b \in [0,1]$.
- ρ_{ab} : Critério de vigilância do módulo inter-ART.
- β : Taxa de aprendizagem, $\beta \in [0,1]$.
- ε : Incremento do parâmetro ρ_a

Inicialização das matrizes de pesos de todos os módulos:

$$W^a = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}_{n \times p} \quad W^b = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}_{n \times q} \quad W^{ab} = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}_{n \times n} \quad (4)$$

Indicando que não há categorias ativas.

Fase de treinamento

Módulo ART_b : um padrão de saída desejada é apresentado a este módulo e comparado à toda a matriz de pesos W^b através da equação (5) (CARPENTER; GROSSBERG, 1992).

$$T_k^b(I^s) = \frac{|I^s \wedge w_k^b|}{\alpha + w_k^b} \quad (5)$$

$$K = \arg\{\max(T_k^b : k = 1, \dots, N_b)\} \quad (6)$$

Em que, $I^s \wedge w_k^b = \min(I^s, w_k^b)$ e N_b é o número de categorias em W^b .

A equação (6) seleciona uma categoria K para ser avaliada pelo parâmetro de vigilância ρ_b através da equação (7).

$$x^b = \frac{|I^s \wedge w_k^b|}{|I^s|} \geq \rho_b \quad (7)$$

Uma vez que uma categoria é aceita um índice K em W^b é escolhido a categoria associada a este índice tem seus pesos adaptados de acordo com (8) e é calculado o vetor de atividade y_K^b em F_2 como em (9).

$$w_K^{novo} = \beta(I^s \wedge w_K^{velho}) + (1 - \beta)w_K^{velho} \quad (8)$$

Em que $\beta \in [0,1]$ é a taxa de aprendizagem.

$$y_K^b = [y_1^b, y_2^b, \dots, y_N^b]$$

sendo:

$$y_k^b = \begin{cases} 1, & \text{se } k = K \\ 0, & \text{se } k \neq K \end{cases} \quad (9)$$

Se o parâmetro de vigilância não for satisfeito, uma nova categoria é escolhida para ser avaliada. Este processo se repete até que o parâmetro ρ_b seja satisfeito.

Módulo ART_a : um padrão de entrada é apresentado ao módulo ART_a , e comparado a toda a matriz de pesos W^a através da equação (10) (CARPENTER; GROSSBERG; IIZUKA, 1992).

$$T_j^a = \frac{|I^e \wedge w_j^a|}{\alpha + w_j^a} \quad (10)$$

A equação (11) seleciona uma categoria J para ser avaliada pelo parâmetro de vigilância ρ_a :

$$J = \arg\{\max(T_j^a: j = 1, \dots, N_a)\} \quad (11)$$

N_a é o número de categorias em W^a .

A categoria escolhida é avaliada pelo parâmetro ρ_a de acordo com (12).

$$x^a = \frac{|I^e \wedge w_j^a|}{|I^e|} \geq \rho_a \quad (12)$$

Se a categoria é aceita, um índice J em W^a é escolhido e inicia-se o módulo inter-ART.

Se a categoria avaliada não for aceita, uma nova categoria é escolhida para ser avaliada pelo parâmetro de escolha e este processo se repete até que o parâmetro ρ_a seja satisfeito.

Módulo inter – ART: este módulo é responsável pelo casamento entre o padrão de entrada e sua saída desejada. (CARPENTER; GROSSBERG, 1992)

O parâmetro ρ_{ab} serve para indicar se existe casamento entre a categoria J escolhida em ART_a e a categoria K escolhida em ART_b .

$$x^{ab} = \frac{|y^b \wedge w_j^{ab}|}{|y^b|} \geq \rho_{ab} \quad (13)$$

Caso o parâmetro ρ_{ab} seja atendido a categoria J é aceita, os pesos sinápticos associados à esta categoria em W^a e os pesos sinápticos associados à JK em W^{ab} são adaptados, conforme (14) e (15). O vetor de atividade $y_j^a = [y_1^a, y_2^a, \dots, y_M^a]$ em F_2 é calculado de acordo com (16).

$$w_j^{novo} = \beta(I^e \wedge w_j^{velho}) + (1 - \beta)w_j^{velho} \quad (14)$$

$$W_{jk}^{ab} = \begin{cases} 0, & \text{se } j \neq J; k \neq K \\ 1, & \text{se } j = J; k = K \end{cases} \quad (15)$$

$$y_j^a = [y_1^a, y_2^a, \dots, y_N^a]$$

sendo:

$$y_j^a = \begin{cases} 0, & \text{se } j \neq J \\ 1, & \text{se } j = J \end{cases} \quad (16)$$

Se o parâmetro ρ_{ab} não for atendido o parâmetro ρ_a é incrementado de acordo com (17) e uma nova categoria é selecionada em ART_a .

$$\rho_a = \frac{|I^e \wedge w_j^a|}{|I^e|} + \varepsilon \quad (17)$$

No qual ε é o incremento do parâmetro ρ_a .

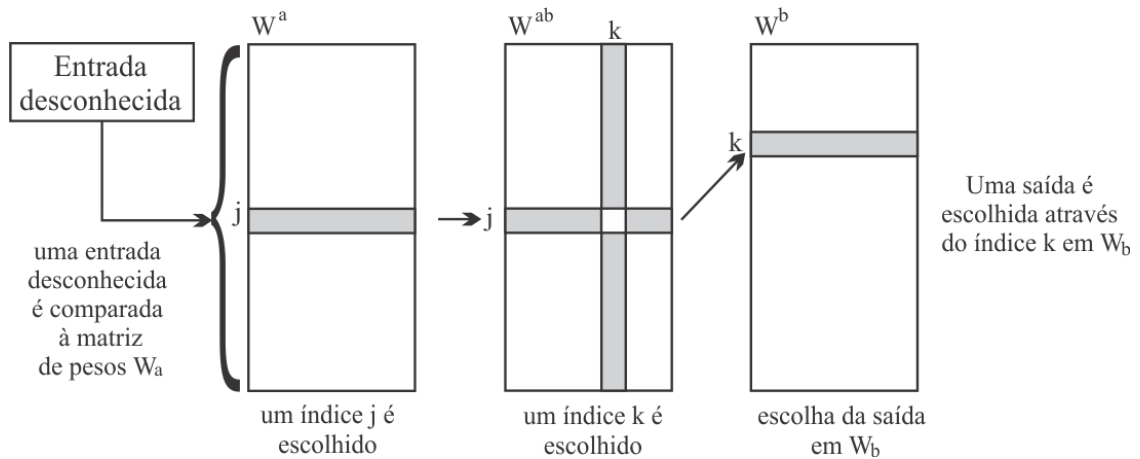
Este processo se repete até que o parâmetro de vigilância ρ_{ab} seja satisfeito.

Fase de diagnóstico

Esta fase ocorre em sua maior parte no módulo ART_a . Uma entrada desconhecida é apresentada à rede. Esta é comparada à toda a matriz W^a e uma categoria J é escolhida para ser avaliada pelo parâmetro de vigilância ρ_a . A confirmação de uma categoria escolhida, conduzirá a uma saída em ART_b de acordo com a ligação que foi feita no módulo interART. Em outras

palavras a categoria J escolhida em W^a indica, na matriz W^{ab} , uma categoria K que mostra a saída em W^b . Todo esse processo está ilustrado na Figura 3.

Figura 3 - Esquema da fase de diagnóstico da Rede Neural ARTMAP *Fuzzy*



Fonte: Elaborada pela própria autora

De uma forma mais prática o algoritmo desta fase é seguinte forma:

Uma entrada desconhecida é apresentada à rede:

- Calcular T_j^a e escolher uma categoria J , de acordo com as equações:

$$T_j^a = \frac{|I^e \wedge w_j^a|}{\alpha + w_j^a} \quad (18)$$

$$J = \arg\{max(T_j^a : j = 1, \dots, N_a)\}$$

- Pelo parâmetro de vigilância ρ_a , avaliar a categoria J escolhida:

$$x^a = \frac{|I^e \wedge w_j^a|}{|I^e|} \geq \rho_a \quad (19)$$

- Se a categoria J escolhida satisfizer o parâmetro de vigilância, então, a linha J de W^{ab} indicará a coluna K que é correspondente à uma categoria em W^b que é a saída y fornecida pela rede.

$$K = \arg\{max(w_j^{ab})\} \quad (20)$$

$$y = W_K^b$$

3.2 REDE NEURAL ARTMAP FUZZY MODIFICADA

MORENO (2010) desenvolveu uma modificação da rede neural ARTMAP *Fuzzy*, que foi denominada rede neural ARTMAP *Fuzzy* modificada. A autora observou que na fase de treinamento todos os neurônios eram testados para a escolha das categorias, ou seja, os ativos e os inativos. Por essa razão, desenvolveu uma modificação que permitisse ativar neurônios de forma gradativa, permitindo que as matrizes de pesos tivessem exatamente a dimensão das categorias ativas.

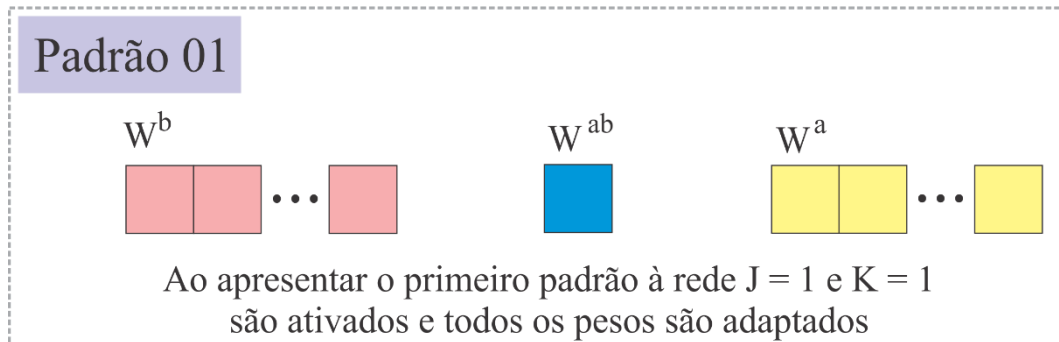
Na modificação proposta a inicialização dos pesos sinápticos da rede é modificada de acordo com (21).

$$W^a = [1 \quad \dots \quad 1]_{1 \times p} \quad W^b = [1 \quad \dots \quad 1]_{1 \times q} \quad W^{ab} = 1 \quad (21)$$

Logo, cada matriz só é inicializada com uma única categoria inativa e à medida que ocorre o treinamento, as categorias são adicionadas de acordo com a necessidade.

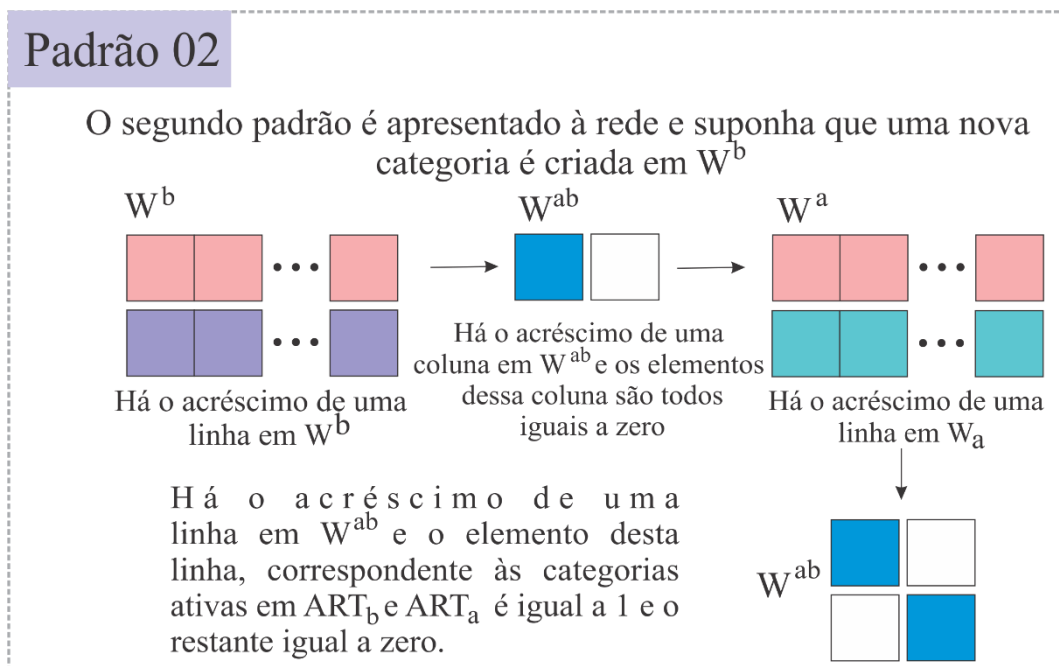
Há duas variáveis que controlam a quantidade de categorias ativas, uma para a matriz W^a e outra para a matriz W^b e essas variáveis são incrementadas à medida que são criadas novas categorias. A variável que controla o número de categorias em W^a também controla o número de categorias em W^{ab} adicionando novas linhas nesta matriz. A variável que controla o número de categorias em W^b também controla o número de categorias em W^{ab} , mas adicionando novas colunas nesta matriz. Em suma, toda vez que uma categoria é criada em W^a automaticamente é adicionada uma linha em W^{ab} e quando uma categoria é criada em W^b uma coluna é adicionada em W^{ab} . Uma ilustração do funcionamento da rede neural ARTMAP *Fuzzy* modificada é mostrada através de um pequeno roteiro, com o auxílio da Figura 6.

Quando o primeiro padrão para treinamento é apresentado à rede a categoria é aceita e tem todos os seus pesos adaptados. (Figura 6).

Figura 4 – Funcionamento da ARTMAP *Fuzzy* Modificada – parte 01

Fonte: Elaborada pela própria autora

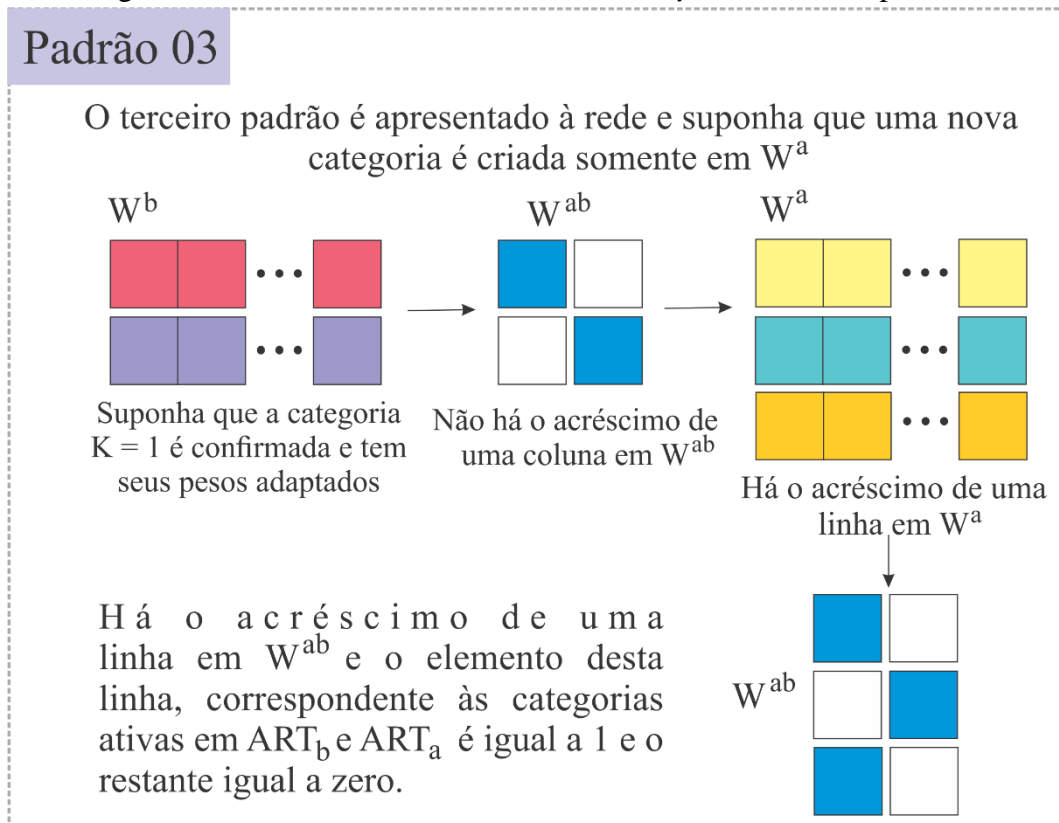
Um segundo padrão é apresentado. Suponha que uma nova categoria é criada em W^b e consequentemente uma nova coluna é adicionada em W^{ab} com todos os valores iguais a zero. Isso é feito desta maneira porque W^{ab} é atualizada apenas no módulo inter-ART no processo de *match-tracking*. Terminada a atualização no módulo ART_b os passos seguintes são no módulo ART_a e o módulo inter-ART. Uma vez que uma coluna nula foi adicionada em W^{ab} nenhuma das categorias ativas em W^a será confirmada no *match-tracking*, assim uma nova categoria será criada em W^a , e W^{ab} terá valor igual a 1 correspondente à linha J ativa em W_a e à coluna K ativa em W^b (Figura 5). Assim, toda vez que uma categoria é criada em ART_b também é criada uma em ART_a

Figura 5 - Funcionamento da ARTMAP *Fuzzy* Modificada – parte 02

Fonte: Elaborada pela própria autora

Um terceiro padrão é apresentado. Suponha agora que não haja criação de nova categoria em W^b e uma nova categoria é criada somente em W^a , automaticamente uma linha é adicionada em W^{ab} e tem seu valor igual a 1 correspondente à categoria J ativa de W^a e à categoria K de W^b (Figura 6 -3).

Figura 6 – Funcionamento da ARTMAP *Fuzzy* Modificada – parte 03



Fonte: Elaborada pela própria autora

Assim, ao final do treinamento da rede, o tamanho de todas as matrizes de pesos será exatamente o mesmo das categorias ativas nos módulos ART_a e ART_b . Além disso, essa estratégia permite que o treinamento continue à medida que o banco seja alimentado com dados mais atualizados.

De agora em diante toda referência que contiver a rede neural ARTMAP *Fuzzy* subentende-se ser a rede neural ARTMAP *Fuzzy* modificada.

3.3 MÉTODO DOS MÍNIMOS QUADRADOS - MMQ

Dados um conjunto de observações X , também são denominadas de variáveis independentes e um conjunto Y de variáveis de respostas ou independentes uma equação de regressão é da forma: (LARSON e FARBER, 2010)

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (22)$$

Ou alternativamente, na forma matricial:

$$\hat{y} = [1 \quad x_1 \quad x_2 \quad \dots \quad x_n] \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \quad (23)$$

Em que $a_0, a_1, a_2, \dots, a_n$ são os coeficientes da equação de regressão que são obtidos através dos MMQ. Este método obtém os coeficientes de tal forma que a soma do quadrado dos resíduos seja mínima. Resíduos são a diferença entre a saída real e a saída prevista (RUGGIERO e LOPES, 1996).

De acordo com Lipschutz e Lipson (2011) uma solução dos mínimos quadrados de um sistema $Ax = B$ é o vetor de menor norma euclidiana que minimiza $\|Ax - B\|_2$. Este vetor é dado por:

$$X = A^+B \quad (24)$$

No qual A^+ é a pseudoinversa de A ou inversa de Moore-Penrose (LIPSCHUTZ; LIPSON, 2011).

Os próximos dois teoremas enunciados no livro de Lipschutz e Lipson (2011) são a respeito das condições da existência da pseudoinversa e de como é calculada.

Teorema 1: Toda matriz sobre \mathbb{C} tem uma única matriz A^+ de Moore-Penrose quando A tiver posto linha total, ou posto coluna total, tem-se os casos especiais a seguir: (LIPSCHUTZ; LIPSON, 2011)

Teorema 2: (LIPSCHUTZ; LIPSON, 2011)

a) Se A tem posto coluna total (as colunas de A são linearmente independentes), então:

$$A^+ = (A^H A)^{-1} A^H \quad (25)$$

.Se A tem posto linha total (as linhas de A são linearmente independentes), então:

$$A^+ = A^H (A A^H)^{-1} \quad (26)$$

A^H é a matriz transposta conjugada.

Na resolução dos MMQ foi utilizado um software que possui pacotes para lidar com matrizes, incluindo obtenção de inversas, transpostas e resolução de sistemas lineares. Caso se deseje adotar uma linguagem de programação que não possui esses recursos disponíveis, métodos numéricos para solução dos MMQ e para solução de sistemas lineares podem ser encontrados no livro de Ruggiero e Lopes (1996) ou qualquer outro livro de métodos numéricos computacionais.

4 METODOLOGIA

4.1 REDE NEURAL HÍBRIDA ARTMAP FUZZY - MMQ

Devido à sua natureza desagregada, os dados disponibilizados para este estudo possuem perfis diversos. Além disso, estes dados possuem *outliers* dos quais muitos deles são erros de leitura dos medidores. Por essas razões, padrões de entrada muito parecidos possuem saídas bastantes distintas.

Na fase de treinamento da rede neural ARTMAP *Fuzzy* padrões de entrada e de saída desejada são apresentados à rede neural. O módulo ART_b , por meio da escolha da categoria T_k e do parâmetro de vigilância ρ_b , irá identificar uma categoria em W^b mais próxima ao padrão de saída desejado. Se tal categoria existe, a categoria escolhida tem seus pesos adaptados. Se uma categoria não for encontrada, uma nova é criada.

No módulo ART_a acontece praticamente o mesmo, ou seja, a escolha da categoria por T_j e o parâmetro de vigilância ρ_a . Uma vez que uma categoria é escolhida, inicia-se o processo de *match-tracking* no módulo inter-ART. Esse processo é responsável pelo casamento entre os módulos ART_a e ART_b , em outras palavras este módulo indica se a categoria escolhida em ART_a , combina com a categoria escolhida em ART_b .

O *match-tracking* ocorre depois que o parâmetro de vigilância ρ_{ab} é verificado. Se uma categoria escolhida em ART_a é confirmada pelo *match-tracking*, as matrizes W^a e W^{ab} têm seus pesos adaptados, caso contrário o parâmetro ρ_a é incrementado e uma nova categoria é selecionada no módulo ART_a e o processo se repete até que uma categoria seja aceita. Se nenhuma categoria for aceita, uma nova é criada em W^a e em W^{ab} . E assim, gradativamente as categorias são separadas na fase de treinamento da rede neural ARTMAP *Fuzzy*.

Note que, no módulo ART_a se uma categoria for escolhida em W^a significa que ela atingiu o grau de similaridade mínimo exigido para que a categoria seja aceita. Ou seja, esta categoria passou no teste de ρ_a mas, a categoria escolhida pode ser rejeitada pelo módulo inter-ART no processo de *match-tracking*. Obviamente isso ocorre porque a categoria escolhida em W^a não está ligada à categoria escolhida em W^b por W^{ab}

Assim, o processo de *match-tracking* pode separar categorias muito próximas na matriz de pesos W^a que poderão estar ligadas a categorias bem diferentes em W^b .

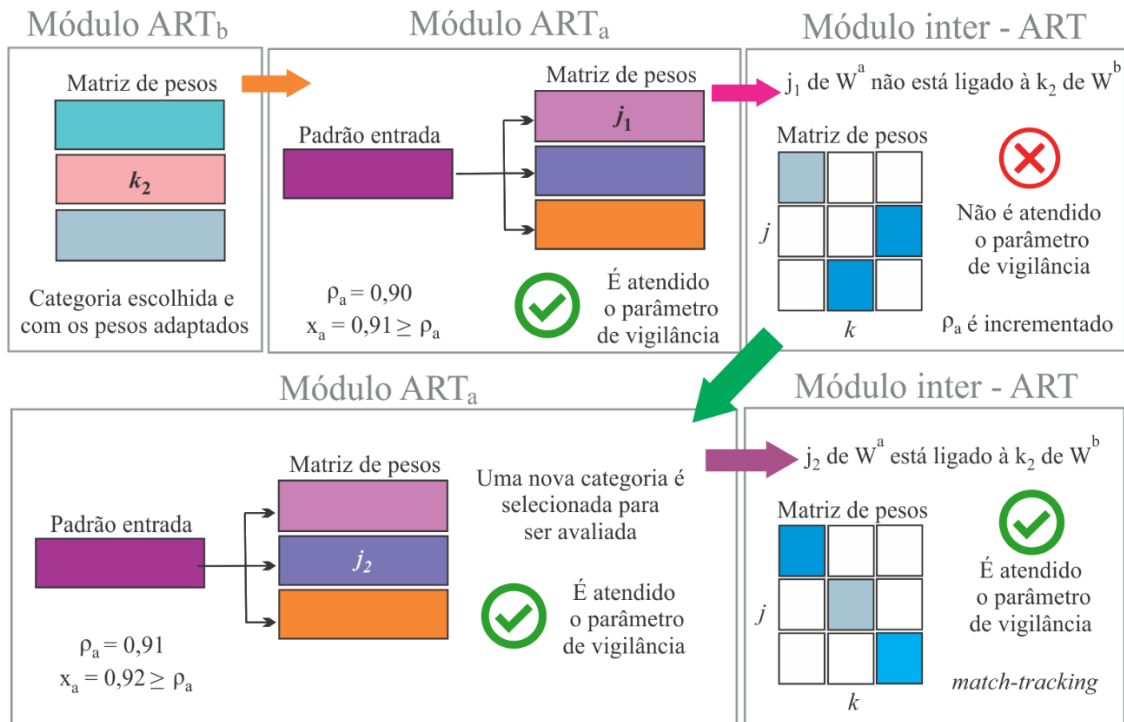
Para exemplificar essas observações considere que um padrão para treinamento seja apresentado à rede neural. Suponha que o módulo ART_b já tenha uma categoria k confirmada e com pesos adaptados. Neste exemplo, chamaremos esta categoria de k_2 , de acordo com a Figura 7, que uma é ilustração deste exemplo.

Para o processo de treinamento no módulo ART_a considere um valor de $\rho_a = 0,90$ e um incremento $e = 0,01$. Considere também que o padrão de entrada apresentado tenha uma similaridade $x_a = 0,91$ com a categoria j_1 selecionada por T_j . Assim, o critério de vigilância ρ_a será atendido confirmando a categoria j_1 . Consequentemente, esta escolha será avaliada pelo módulo inter-ART. Agora, suponha que a categoria j_1 de W^a esteja ligada à categoria k_1 de W^b , indicada pelo primeiro quadrado cinza na Figura 7. Portanto a categoria selecionada em ART_a não satisfaz o critério de vigilância no módulo inter-ART.

Assim, de acordo com o algoritmo da rede ρ_a seria incrementado para $\rho_a = 0,92$ e uma nova categoria seria escolhida em ART_a .

Continuando o treinamento, suponha que uma categoria j_2 selecionada por T_j em W^a atenda o critério de vigilância ρ_a , com um $x_a = 0,92$, e esteja ligada a categoria k_2 em W^b que está representada pelo segundo quadrado cinza na Figura 7. Com isso, o critério de vigilância do módulo inter-ART é satisfeito e a categoria j_2 é confirmada.

Veja que para j_1 tem-se $x_a = 0,91$ e para j_2 tem-se $x_a = 0,92$, ou seja, as categorias j_1 e j_2 são similares entre si e elas podem estar ligadas a categorias W^b distintas, ou não. Para efeito deste exemplo, considere que as categorias k_1 e k_2 são bem distintas que estão representados na Figura 7 pelos retângulos verde e rosa respectivamente.

Figura 7 – Ilustração treinamento ARTMAP *Fuzzy* -

Fonte: Elaborado pela autora

Finalizado o treinamento, a rede neural está apta para entrar em fase de operação. Considere que uma determinada entrada é apresentada para o diagnóstico, com similaridade $x_a = 0,92$ para a categoria j_1 e $x_a = 0,91$ para a categoria j_2 . Considerando um valor de $\rho_a = 0,90$ ambas as categorias passariam pelo parâmetro de vigilância. Caso T_j tenha escolhido a categoria $j = 1$, ela seria confirmada por ρ_a e apontaria como saída k_1 . Mas, se a saída dada por k_2 fosse mais adequada do que a saída dada por k_1 ? Sendo assim, mesmo que a saída dada por k_2 seja a melhor, k_2 não teve a oportunidade de participar da previsão por causa da escolha feita por T_j , mesmo tendo condições de satisfazer o critério de vigilância ρ_a .

Com todas essas observações citadas foi desenvolvida uma metodologia híbrida com o objetivo de permitir que mais categorias tenham a oportunidade de participar do processo de previsão da rede neural ARTMAP *Fuzzy*.

A metodologia híbrida denominada rede híbrida ARTMAP *Fuzzy* – MMQ, utiliza a rede neural ARTMAP *Fuzzy* juntamente com o Método dos Mínimos Quadrados (MMQ).

O método híbrido conserva o treinamento realizado pela ARTMAP *Fuzzy*, mas na fase de diagnóstico a rede neural desempenha o papel de selecionar as categorias para que o MMQ realize a previsão.

A ideia do método, na fase de diagnóstico, é reunir categorias em W^a que atendem o critério de vigilância ρ_a e a partir daí associá-las a suas respectivas saídas em W^b . Assim, obtém-se dois subconjuntos, um com as categorias mais semelhantes com a entrada da previsão e outro subconjunto com as saídas possíveis.

Esses dois subconjuntos serão utilizados como base para a regressão linear. O MMQ é utilizado para encontrar os coeficientes da equação de regressão.

A seguir será mostrado em quatro etapas o funcionamento da arquitetura híbrida e sempre que necessário um trecho do pseudocódigo do método.

1ª etapa: Primeiramente os padrões de entrada e saída desejada são apresentados à rede neural ARTMAP *Fuzzy* para treinamento conforme algoritmo mostrado na seção 3.1

2ª etapa: Nesta fase é iniciado o processo de diagnóstico da rede, que ainda é executada dentro da arquitetura da rede neural ARTMAP *Fuzzy*. Apresenta-se uma entrada I da qual se deseja realizar a previsão. Esta entrada é então comparada à toda a matriz W^a para obter os valores de T_j , ou seja, os valores para a escolha de categorias.

```
Para j = 1 até n
    T(j) = soma(mínimo(I, wa(j)) / (alfa + soma (wa(j)))
Fim
```

A partir deste ponto começa a diferenciação entre a rede neural ARTMAP *Fuzzy* tradicional e a metodologia híbrida. De acordo com o algoritmo da rede neural ARTMAP *Fuzzy* o índice ligado ao maior valor de T_j ganharia a oportunidade de ser avaliado pelo parâmetro de vigilância ρ_a . Na arquitetura híbrida **todos** valores de T_j , serão avaliados pelo parâmetro de vigilância ρ_a , e os índices que satisfazem este parâmetro, serão armazenados em um vetor V .

```
i = 1
Para j = 1 até n
    J = arg(max(T(j))
    xa = soma(mínimo(I, wa(J)) / soma (I)
    Se xa >= ro_a
        V(i) = J
        T(j) = 0
        i = i + 1
    Senão
        T(j) = 0
        i = i + 1
Fim
```

Fim

3ª etapa: Nesta parte inicia-se a preparação dos dados para a regressão linear por mínimos quadrados. O vetor V encontrado servirá de base para a construção de duas matrizes. Uma matriz A , das variáveis independentes, que será composta pelos elementos da matriz de pesos W^a que estão associados a cada um dos índices armazenados em V . E matriz B , das variáveis dependentes, que será composta pelas categorias de W^b que estão ligadas às categorias selecionadas em W^a . Assim, a matriz A terá os elementos mais similares com a entrada para a previsão apresentada, enquanto B pelos elementos com as saídas possíveis.

```
m = i - 1 # este índice i veio do item anterior
Para j = 1 até m
  A = wa(v(j))
  K = arg(max(wab(v(j)))
  B = wb(K)
Fim
```

4ª etapa: Esta última parte é encarregada na preparação dos dados, a regressão linear e a saída calculada.

As matrizes A e B que foram obtidas na etapa anterior estão configuradas de acordo com as matrizes de pesos da rede neural. Para utilizá-las no método dos mínimos quadrados será necessário efetuar duas modificações.

Primeiro, como a equação de regressão é da forma $\hat{y} = a_0 + a_1x_1 + \dots + a_nx_n$, o componente a_0 que é seu coeficiente de deslocamento precisa ser representado na matriz A . Para isso, basta adicionar uma coluna unitária na matriz. Ao final da etapa anterior, a matriz A é representada por (27), em que que os w' representam as categorias selecionadas de W^a . Depois de adicionada uma coluna unitária, a matriz A que passa a ser representada como em (28).

$$A = \begin{bmatrix} w'_{11} & \cdots & w'_{1n} \\ \vdots & \ddots & \vdots \\ w'_{m1} & \cdots & w'_{mn} \end{bmatrix} \quad (27)$$

$$A = \begin{bmatrix} 1 & w'_{11} & \cdots & w'_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & w'_{m1} & \cdots & w'_{mn} \end{bmatrix} \quad (28)$$

O mesmo procedimento deve ser adotado para o vetor de entrada para previsão.

A matriz B tem duas colunas, sendo que a segunda coluna corresponde à codificação complementar que foi herdada da RNA, mas completamente desnecessária para o MMQ, por essa razão esta coluna deve ser retirada de B .

Portanto as matrizes A e B estão preparadas para o MMQ, para obter os coeficientes de melhor ajuste. A solução dos mínimos quadrado é dada pelo Teorema 2(b) apresentada na equação (26)

```
coeficientes_regressao = pseudoinversa(A) * B
saída_calculada = I_e * coeficientes_regressão
```

Como exemplo do algoritmo da metodologia híbrida, suponha as matrizes de pesos adaptados.

$$W^a = \begin{bmatrix} 0,09 & 0,56 & 0,91 & 0,44 \\ 0,12 & 0,58 & 0,88 & 0,42 \\ 0,65 & 0,96 & 0,35 & 0,04 \\ 0,67 & 0,97 & 0,33 & 0,03 \\ 0,65 & 0,95 & 0,35 & 0,05 \\ 0,28 & 0,98 & 0,72 & 0,02 \\ 0,31 & 0,95 & 0,69 & 0,05 \\ 0,27 & 0,91 & 0,73 & 0,09 \\ 0,12 & 0,61 & 0,88 & 0,39 \\ 0,08 & 0,60 & 0,92 & 0,40 \end{bmatrix}$$

$$W^b = \begin{bmatrix} 0,25 & 0,75 \\ 0,38 & 0,62 \\ 0,55 & 0,45 \\ 0,75 & 0,25 \\ 0,83 & 0,17 \\ 0,86 & 0,14 \\ 0,89 & 0,11 \\ 0,95 & 0,05 \end{bmatrix}$$

$$W^{ab} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

E o vetor de entrada para previsão $I = [0,42 \quad 0,90 \quad 0,58 \quad 0,10]$ já normalizado, com codificação complementar e com saída real de $y = 0,87$, os parâmetros possuem os seguintes valores - $\rho_a = 0,85$ e $\alpha = 0,5$.

Assim, para cada elemento de W^a será calculado um valor de T_j para a escolha da categoria, neste exemplo será mostrado em detalhes os cálculos do primeiro elemento de W^a , pois os procedimentos para os outros elementos da matriz são calculados da mesma forma. Logo,

$$T_j = \frac{|I \wedge W_j^a|}{\alpha + |W_j^a|} \quad (29)$$

$$T_{j=1} = \frac{|[0,42 \quad 0,90 \quad 0,58 \quad 0,10] \wedge [0,09 \quad 0,56 \quad 0,91 \quad 0,44]|}{0,5 + |[0,09 \quad 0,56 \quad 0,91 \quad 0,44]|} = \frac{1,33}{2,5} = 0,53$$

Fazendo os cálculos para as demais categorias de W^a , os valores de T_j obtidos estão listados na Tabela 1 a seguir.

Tabela 1 – Valores do parâmetro de escolha de categorias de W^a

Nº categoria	T_j
1	0,53
2	0,55
3	0,68
4	0,67
5	0,68
6	0,71
7	0,73
8	0,73
9	0,56
10	0,54

Fonte: Elaborado pela autora

De acordo com o algoritmo da rede neural ARTMAP *Fuzzy* a categoria escolhida para realizar o teste de vigilância seria a categoria $J = 7$ ou $J = 8$, pois possuem os maiores valores de T_j , porém iguais. O algoritmo da ARTMAP *Fuzzy* diz que, caso isso ocorra a categoria a ser escolhida será a de menor índice, neste caso o índice $J = 7$. Se categoria escolhida atender ao critério de vigilância, uma saída seria dada por de W^{ab} .

A Tabela 2 mostra que o parâmetro de ρ_a estabelecido foi aceito para $J = 7$ e o Quadro 01 mostra que o índice confirmado em W^a conduz a $K = 5$ em W^b . Portanto, a saída da rede neural é $\hat{y} = 0,83$.

Tabela 2 – Valores de x^a de todas as categorias de W^a

Nº categoria	x^a
1	0,66
2	0,69
3	0,85
4	0,84
5	0,86
6	0,89
7	0,92
8	0,92
9	0,70
10	0,68

Fonte: Elaborado pela autora

Continuando o algoritmo, na metodologia híbrida é feito o teste de vigilância para todas as categorias, assim para cada uma das categorias selecionadas é calculado o valor de x_a pela equação (30) para ser avaliado pelo critério de vigilância ρ_a

$$x_a = \frac{|I \wedge W|}{|I|} \quad (30)$$

De maneira análoga que foram realizados os cálculos de T_j , foram feitos os cálculos de x_a e os resultados são mostrados na Tabela 2 – Valores de x^a de todas as categorias de W^a . Tabela 2

Considere um valor de $\rho_a = 0,85$, assim, todas as categorias cujos valores de x_a estejam acima deste valor serão selecionados para fazer parte da matriz que será utilizada como base para a equação de regressão. No caso deste exemplo as categorias são $J = [3 \ 5 \ 6 \ 7 \ 8]$.

Para ilustrar como as categorias K são selecionadas, os valores da matriz W^{ab} estão representadas na Quadro 01. Os valores destacados representam o casamento da categoria J (entrada) com uma categoria K (saída). Para cada linha de J há um único valor unitário que indica uma coluna em K que apontará para a categoria W^b correspondente.

Quadro 01 – Valores tabelados da matriz W^{ab}

		Índices k							
		1	2	3	4	5	6	7	8
Índices j	1	1	0	0	0	0	0	0	0
	2	1	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	1
	4	0	0	0	0	0	1	0	0
	5	0	0	0	0	0	0	1	0
	6	0	0	0	1	0	0	0	0
	7	0	0	0	0	1	0	0	0
	8	0	0	0	0	1	0	0	0
	9	0	1	0	0	0	0	0	0
	10	0	0	1	0	0	0	0	0

Fonte: Elaborado pela autora

De acordo com o Quadro 01, os índices K escolhidos são $K = [4 \ 5 \ 7 \ 8]$, sendo que o índice $K = 5$ aponta para a saída de ambas categorias de $J = 7$ e $J = 8$.

Assim, as associações entre os módulos ART_a e ART_b considerando esses critérios ficam de acordo com a Tabela 3.

Tabela 3 – Categorias J selecionadas x Categorias K associadas

Índices J selecionados	Índices K selecionados
3	8
5	7
6	4
7	5
8	5

Fonte: Elaborado pela autora

De acordo com a tabela acima, as matrizes de pesos selecionadas são:

$$A = \begin{bmatrix} 0,65 & 0,96 & 0,35 & 0,04 \\ 0,65 & 0,95 & 0,35 & 0,05 \\ 0,28 & 0,98 & 0,72 & 0,02 \\ 0,31 & 0,95 & 0,69 & 0,05 \\ 0,27 & 0,91 & 0,73 & 0,09 \end{bmatrix} \quad B = \begin{bmatrix} 0,95 & 0,05 \\ 0,89 & 0,11 \\ 0,75 & 0,25 \\ 0,83 & 0,17 \\ 0,83 & 0,17 \end{bmatrix}$$

Um ponto a observar é que A e B possuem dimensões mais reduzidas. Além disso, as categorias de B estão limitadas entre 0,75 e 0,95 enquanto as categorias de W^b estão entre 0,25 e 0,95. Logo, a equação de regressão será obtida levando em consideração as categorias mais próximas, em ambos os módulos. A saída real $y = 0,87$ também está entre os valores escolhidos em B .

Antes de iniciar o cálculo para a equação de regressão é necessário incluir um vetor coluna unitário na matriz A e excluir a segunda coluna da matriz B , que corresponde à codificação complementar exigida pela rede neural, mas dispensável para o MMQ. Assim, as matrizes que serão utilizadas para o MMQ são:

$$A = \begin{bmatrix} 1 & 0,65 & 0,96 & 0,35 & 0,04 \\ 1 & 0,65 & 0,95 & 0,35 & 0,05 \\ 1 & 0,28 & 0,98 & 0,72 & 0,02 \\ 1 & 0,31 & 0,95 & 0,69 & 0,05 \\ 1 & 0,27 & 0,91 & 0,73 & 0,09 \end{bmatrix} \quad B = \begin{bmatrix} 0,95 \\ 0,89 \\ 0,75 \\ 0,83 \\ 0,83 \end{bmatrix}$$

A e B são as matrizes que serão usadas no método dos mínimos quadrados. Assim, para encontrar os coeficientes de regressão, basta calcular:

$$X = A^+B \quad \text{em que} \quad A^+ = A^H(AA^H)^{-1} \quad (31)$$

Aplicando ao exemplo:

$$X = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} 0,66 \\ 0,51 \\ -0,17 \\ 0,16 \\ 0,84 \end{bmatrix}$$

Que são os coeficientes da equação de regressão:

$$\hat{y} = 0,66 + 0,51x_1 - 0,17x_2 + 0,16x_3 + 0,84x_4 \quad (32)$$

Incluindo o valor um para representar o coeficiente de deslocamento no vetor de entrada I e substituindo na equação de regressão obtida, ou equivalentemente:

$$\hat{y} = [1 \quad 0,42 \quad 0,90 \quad 0,58 \quad 0,10] \begin{bmatrix} 0,66 \\ 0,51 \\ -0,17 \\ 0,16 \\ 0,84 \end{bmatrix} = 0,89$$

No caso deste exemplo, a rede neural ARTMAP *Fuzzy* obteve como saída $\hat{y} = 0,83$, do primeiro índice escolhido e confirmado pelo parâmetro de vigilância ρ_a . Enquanto a metodologia híbrida retornou $\hat{y} = 0,89$ que é um pouco mais próximo da saída real $\hat{y} = 0,87$

Resumindo, o objetivo principal da rede híbrida ARTMAP *Fuzzy* - MMQ é agrupar as categorias de entrada mais próximas com as saídas mais prováveis para o MMQ. Para cada

entrada para diagnóstico, uma equação de regressão diferente é obtida utilizando somente os dados mais similares com a entrada apresentada.

Se os dados presentes nas matrizes de pesos fossem todos utilizados diretamente para obter os coeficientes de regressão, ou seja:

$$W^a = \begin{bmatrix} 0,09 & 0,56 & 0,91 & 0,44 \\ 0,12 & 0,58 & 0,88 & 0,42 \\ 0,65 & 0,96 & 0,35 & 0,04 \\ 0,67 & 0,97 & 0,33 & 0,03 \\ 0,65 & 0,95 & 0,35 & 0,05 \\ 0,28 & 0,98 & 0,72 & 0,02 \\ 0,31 & 0,95 & 0,69 & 0,05 \\ 0,27 & 0,91 & 0,73 & 0,09 \\ 0,12 & 0,61 & 0,88 & 0,39 \\ 0,08 & 0,60 & 0,92 & 0,40 \end{bmatrix} \quad W^b = \begin{bmatrix} 0,25 & 0,75 \\ 0,25 & 0,75 \\ 0,95 & 0,05 \\ 0,86 & 0,14 \\ 0,89 & 0,11 \\ 0,75 & 0,25 \\ 0,83 & 0,17 \\ 0,83 & 0,17 \\ 0,38 & 0,62 \\ 0,55 & 0,45 \end{bmatrix}$$

Neste caso, as devidas associações entre W^a e W^b já estão feitas. Logo, as matrizes para a obtenção dos mínimos quadrados são:

$$A = \begin{bmatrix} 1 & 0,09 & 0,56 & 0,91 & 0,44 \\ 1 & 0,12 & 0,58 & 0,88 & 0,42 \\ 1 & 0,65 & 0,96 & 0,35 & 0,04 \\ 1 & 0,67 & 0,97 & 0,33 & 0,03 \\ 1 & 0,65 & 0,95 & 0,35 & 0,05 \\ 1 & 0,28 & 0,98 & 0,72 & 0,02 \\ 1 & 0,31 & 0,95 & 0,69 & 0,05 \\ 1 & 0,27 & 0,91 & 0,73 & 0,09 \\ 1 & 0,12 & 0,61 & 0,88 & 0,39 \\ 1 & 0,08 & 0,60 & 0,92 & 0,40 \end{bmatrix} \quad B = \begin{bmatrix} 0,25 \\ 0,25 \\ 0,95 \\ 0,86 \\ 0,89 \\ 0,75 \\ 0,83 \\ 0,83 \\ 0,38 \\ 0,55 \end{bmatrix}$$

A equação de regressão obtida pelo MMQ seria:

$$\hat{y} = 0,17 + 0,18x_1 + 0,66x_2 - 0,01x_3 - 0,49x_4 \quad (33)$$

Substituindo os valores da entrada I já com o valor unitário incluído na equação obtida, obtém-se $\hat{y} = 0,78$.

Tabela 4 – Saídas obtidas pelas três metodologias e a saída real

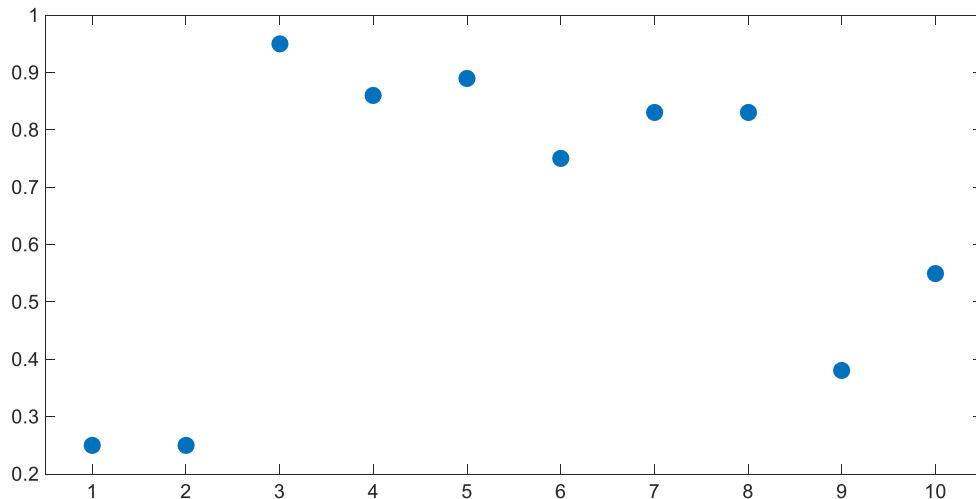
Saída MMQ	Saída RNA	Saída Híbrida	Saída Real
0,78	0,83	0,89	0,87

Fonte: Elaborado pela autora

Se compararmos as saídas obtidas, conforme Tabela 4, a saída da híbrida está mais próxima da saída real. É importante observar aqui a melhora de performance entre a previsão

do MMQ com todas as categorias presentes em comparação ao resultado obtido depois da seleção das categorias mais próximas na metodologia híbrida.

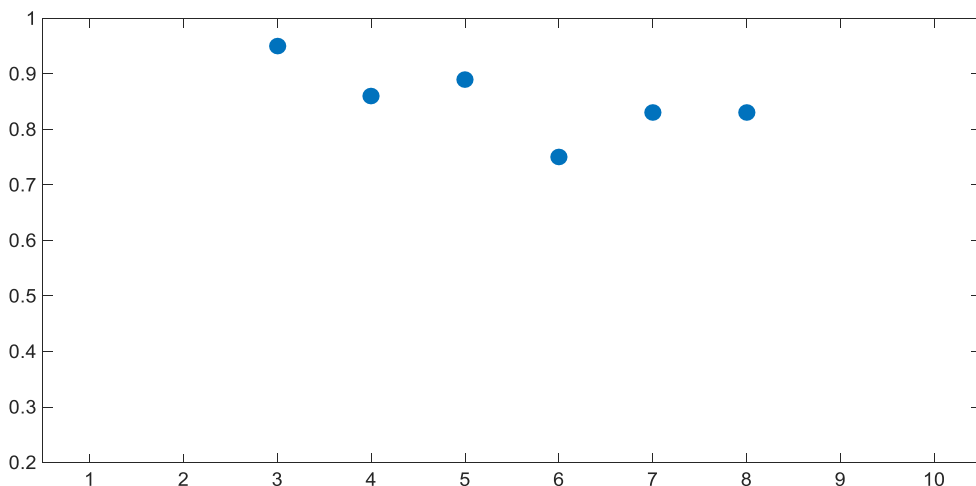
Figura 8 – Dispersão de todos os elementos de W^b



Fonte: Elaborado pela própria autora

A Figura 8 – Dispersão de todos os elementos de W^b é uma representação gráfica de todos os elementos da matriz W^b , ou seja, a representação de todas as categorias que representam a saída desejada. Note que há quatro pontos afastados, dos quais foram removidos pelo modelo híbrido conforme mostrado na Figura 9.

Figura 9 – Dispersão das categorias conforme matriz B



Fonte: Elaborado pela própria autora

Portanto, a grande vantagem da metodologia híbrida é permitir a possibilidade de mais de uma categoria ser utilizada na fase de previsão, ou seja, as categorias com mais afinidade

participam da previsão. Além disso, em relação à regressão linear o MMQ encontrará os coeficientes de melhor ajuste utilizando saídas possíveis próximas.

Neste trabalho, o MMQ foi utilizado por ser um método de fácil implementação e de baixo custo computacional. Porém é necessário cuidado para que as entradas escolhidas tenham posto linha completo, que é a condição necessária para que existência da pseudoinversa (LIPSCHUTZ e LIPSON, 2011). Em todos os testes realizados para esse trabalho, em nenhum caso, houve problemas no cálculo dos coeficientes de regressão, mas, dependendo dos dados utilizados a condição de existência pode não ser cumprida. Contudo, nada impede a adoção de outra metodologia para ser utilizada no lugar dos MMQ, como por exemplo, modelos ARIMA.

4.2 CONSTITUIÇÃO E PREPARAÇÃO DOS DADOS

Nesta seção serão apresentados a constituição dos dados para este estudo e como eles foram arranjados para a realização dos testes.

Os dados cedidos para esta pesquisa são de uma universidade brasileira. Eles foram coletados em quatro blocos de prédios que têm diferentes perfis de carga. Conforme o Quadro 02 um desses blocos possui um restaurante, uma cafeteria, uma agência de banco e uma de correios. Os demais blocos é um misto entre salas de aula, laboratórios de ensino e escritórios administrativos. O período de medição é de três meses, maio a julho de 2013, correspondendo um total de treze semanas.

Quadro 02 - Constituição dos blocos para previsão de cargas

Bloco	Constituição
1	Restaurante, cafeteria, agência de banco e de correios
2	Salas de aula, laboratórios de ensino, escritórios de administração
3	
4	

Fonte: Elaborado pela autora

Esses dados foram obtidos por meio de um medidor multivariável modelo upd-200 instalado junto ao transformador de cada bloco e possui três tipos distintos de entradas. A primeira é o instante t em que a carga foi medida, que neste caso foi a intervalos de 15 minutos,

totalizando 96 entradas para cada dia. A segunda entrada é o dia da semana ds de domingo a sábado. A última entrada é a carga L medida no instante t , conforme mostrado na Tabela 5.

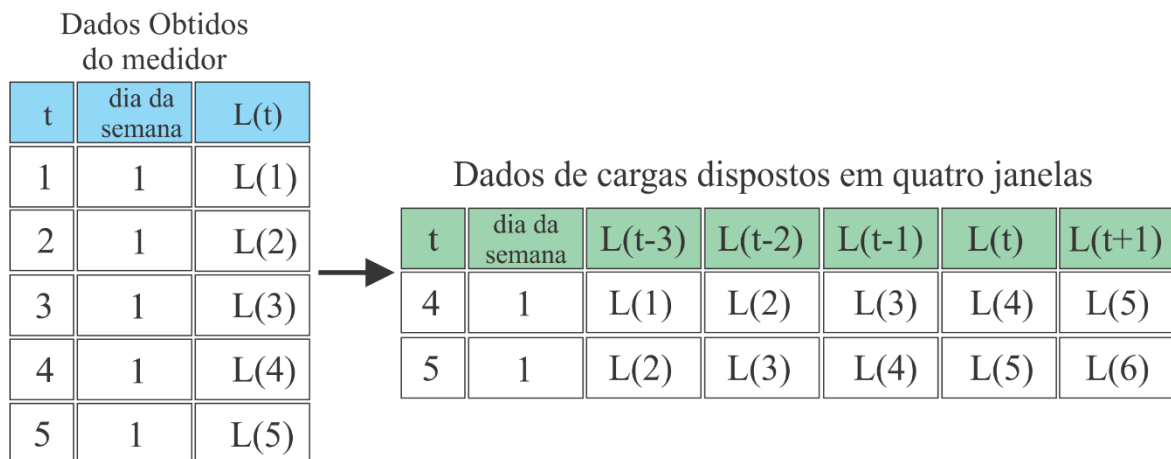
Tabela 5 – Constituição dos dados iniciais obtidos pelo medidor

Tempo (t)	Dia da semana (ds)	$L(t)$
1 a 96	Domingo a sábado (1 a 7)	Quilowatts (kW)

Fonte: Elaborado pela autora

Esses dados serviram de base para montar as séries temporais de cada um dos blocos para os testes de previsão. Na Figura 10, à esquerda, está uma representação dos dados obtidos pelos medidores e à direita um modelo de representação das séries temporais que foram usadas para este estudo. Para a composição das séries as cargas elétricas foram janeladas, para este trabalho foram utilizadas quatro janelas para os dados de entrada e uma para a saída desejada.

Figura 10 – Esquema da disposição dos dados para o estudo.



Fonte: Elaborado pela própria autora

Várias representações de séries temporais foram testadas, desde séries menores a séries bem maiores e, para estes cenários, os resultados não variaram muito entre si. Sendo assim, nos primeiros testes, consideraram-se dados de cargas em quatro janelas, conforme mostrado em (34), em que $X(t)$ é o conjunto dos padrões de entrada e $Y(t)$ o padrão de saída desejada. Mais adiante será mostrado que outros dados foram inseridos para os testes.

$$X(t) = [t \quad ds \quad L(t-3) \quad L(t-2) \quad L(t-1) \quad L(t)] \quad X \in \mathbb{R}^m$$

$$Y(t) = [L(t+1)] \quad Y \in \mathbb{R}^1$$
(34)

Em que

- t : Tempo
 ds : Dia da semana
 $L(t - p)$: Carga no tempo t atrasado em p , em que $p = 0, 1, 2, 3$
 $L(t + 1)$: Carga em $t + 1$

Figura 11 – Funcionamento do movimento de janela no banco de dados na etapa de previsão de cargas

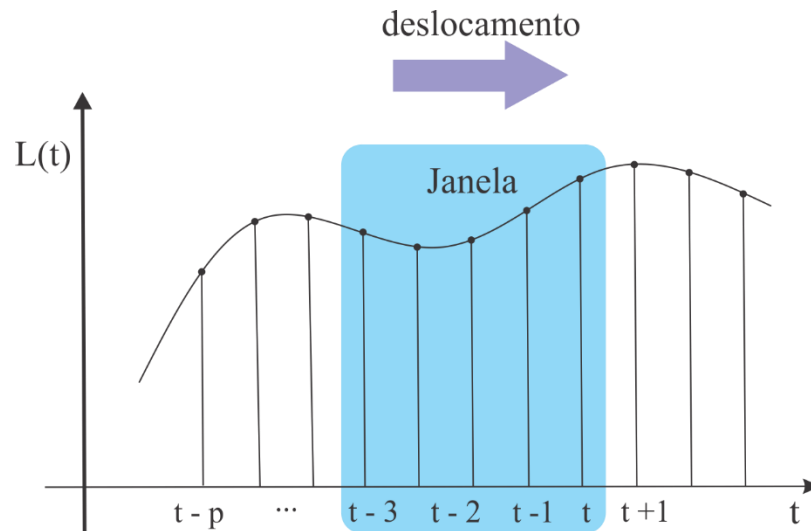
Entradas para a previsão						Saída prevista
t	dia da semana	$L(t-3)$	$L(t-2)$	$L(t-1)$	$L(t)$	$L(t+1)$
4	1	$L(1)$	$L(2)$	$L(3)$	$L(4)$	$\underline{L(5)}$
5	1	$L(2)$	$L(3)$	$L(4)$	$\underline{L(5)}$	$\underline{L(6)}$
6	1	$L(3)$	$L(4)$	$\underline{L(5)}$	$\underline{L(6)}$	$\underline{L(7)}$
7	1	$L(4)$	$\underline{L(5)}$	$\underline{L(6)}$	$\underline{L(7)}$	$\underline{L(8)}$

Fonte: da própria autora

As previsões de cargas feitas para este trabalho foram com um horizonte de 24 horas à frente, ou seja, previsão de curto prazo. A Figura 11 é uma ilustração de como foram construídas cada entrada para previsão. Cada uma das saídas previstas em um determinado instante servirá como entrada para o instante seguinte e também para os instantes subsequentes. Em um determinado momento, as entradas referentes às cargas elétricas serão compostas somente de previsões passadas. Por exemplo, na Figura 12 a saída prevista $L(5)$ para $t = 4$ fará parte das entradas para previsão de $t = 5$, $t = 6$ e $t = 7$, o mesmo procedimento serve para as demais saídas previstas.

Esse procedimento é chamado de técnica de movimento de janela ou janela deslizante. (DJUKANOVIC, RUZIC, *et al.*, 1995; PARK, EL-SHARKAWI e MARKS II, 1991, citados por ANTUNES, 2013). Conforme a previsão avança a observação mais antiga é descartada e em seu lugar é adicionada uma informação mais recente. Cada janela tem um tamanho do conjunto de entrada que participará do movimento de janela, que neste trabalho foram quatro dados de carga (Figura 12).

Figura 12 – Esquema da técnica movimento de janela ou janela deslizante



Fonte: Adaptação (ANTUNES, 2013)

Cada carga estimada possui um erro e este erro estará presente na previsão seguinte. Conforme avança a previsão, o conjunto de entrada tenderá a ter somente cargas previstas e cada uma delas possui seu respectivo erro.

Na fase de previsão a rede neural ARTMAP *Fuzzy* procura uma categoria que seja a mais similar possível com a entrada apresentada, sendo assim, quanto maiores forem os erros das previsões passadas, maiores serão as chances de aumento dos erros nas previsões futuras, uma vez que os conjuntos de entradas tenderão a se afastar mais das observações reais, por causa do movimento de janela.

Qualquer rede neural tem como objetivo extrair características de um conjunto de informações. Todo o processo de extrair essas características é chamado de aprendizado. Para que uma rede neural consiga aprender de modo eficiente é importante que o conjunto, que será submetido a um processo de aprendizagem, seja o mais diverso possível. Por exemplo, se alguém quiser treinar uma rede neural para avaliação de risco de empréstimo a clientes de um banco e, para fazer isso possua dados históricos de empréstimos realizados, informações como idade, sexo e valor do salário mensal podem não ser suficientes para fazer um bom treinamento. Por outro lado, outras informações como estado civil, objetivo do empréstimo, possuir casa própria e grau de escolaridade podem contribuir para que o processo de extração de características da rede neural seja mais eficiente.

Os dados utilizados para este trabalho possuem somente três tipos de entrada o instante t , o dia da semana ds e as cargas janeladas. Outros tipos de entrada, como temperatura, humidade ou ocupação não fazem parte dos conjuntos de dados.

Mesmo que a rede neural ARTMAP *Fuzzy* consiga realizar um bom treinamento com essa variedade de entradas, a previsão pode ser ineficiente por causa do movimento de janela das cargas. De fato, isso ocorreu na maioria dos testes que foram feitos utilizando somente as configurações de entrada de acordo com a Figura 10 à direita.

Há somente duas entradas que não estão sujeitas ao movimento, ou seja, o instante t e o dia da semana ds , pois elas são sempre conhecidas. Alguma outra informação, de preferência que não fizesse parte do movimento de janela ajudaria no treinamento/ diagnóstico da rede neural. E como, no caso deste trabalho, tal informação não existe, foi adotada uma estratégia para contornar essa deficiência.

A estratégia foi a de inserir no conjunto de dados os valores de carga média, mínima e máxima históricos para cada dia da semana ds e cada instante t . Em outras palavras, toma-se todos os dados históricos de um único instante t relacionado a um único dia da semana ds e a partir desses dados calcula-se as cargas máxima, mínima e média. Por exemplo, considere os dados da Tabela 6.

Tabela 6 – Dados de exemplo para adoção de outros dados para testes

Terças-feiras			Quartas-feiras		
t	ds	$L(t)$	t	ds	$L(t)$
5	3	2200	5	4	2100
5	3	3200	5	4	2200
5	3	2400	5	4	3100
5	3	2900	5	4	2500
5	3	3200	5	4	1900
5	3	3100	5	4	2800
5	3	3200	5	4	2300
5	3	3000	5	4	2100
5	3	3100	5	4	2600

Fonte: Elaborado pela autora

Os dados da Tabela 6 referem-se a medidas de cargas de algumas terças-feiras e algumas quartas-feiras, ambas medidas no instante $t = 5$. A partir dessas informações são calculados os valores das cargas máxima, mínima e média de cada um desses dias, os resultados estão na Tabela 7.

Tabela 7 – Valores das cargas máxima, mínima e média- terças e quartas

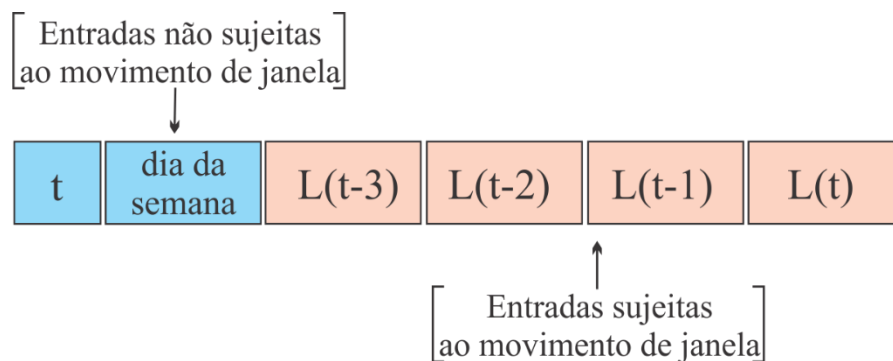
	Terça-feira	Quarta-feira
Máximo	3.200	3100
Mínimo	2.200	1900
Média	2.922	2400

Fonte: Elaborado pela autora

Esses valores são inseridos nos conjuntos de dados e não estão sujeitos ao movimento de janela na fase de previsão. Eles funcionam como uma espécie de estimativa. Por exemplo, ao realizar a previsão de cargas de uma terça-feira em $t = 5$, os dados históricos dizem que para este dia e neste instante, o valor máximo de carga foi de 3.200 KW, o mínimo foi de 2.200 KW e a média foi de 2.922 KW, ou seja, os valores esperados para esta previsão estarão muito provavelmente dentro desses valores.

Assim, a rede neural tem um pouco mais de variedade para treinamento e diagnóstico. Pode não ser uma variedade muito grande por se tratarem também de cargas elétricas, mas estas cargas não fazem parte do movimento de janela e, portanto, não irão propagar erros em previsões subsequentes.

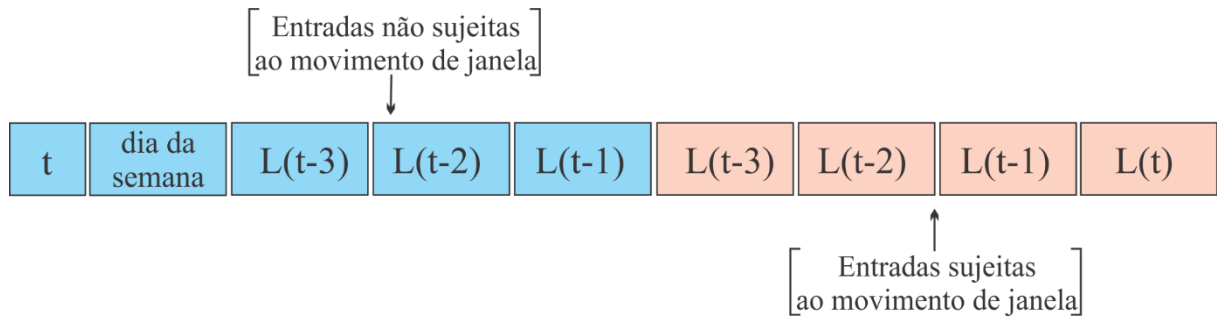
Figura 13 - Entradas antes da inclusão das cargas máxima, mínima e média.



Fonte: Elaborado pela autora

De acordo com a Figura 13 há duas entradas que não estão sujeitas ao movimento de janela, ou seja, t e ds . Já na Figura 14, com a inclusão dos valores de carga média, mínima e máxima, a quantidade de entradas não sujeitas ao movimento de janela passou a ser cinco.

Figura 14 - Entradas após a inclusão das cargas máxima, mínima e média.



Fonte: Elaborado pela autora

Sendo assim, os dados de entrada a partir da inclusão dessas informações, são mostradas em (35).

$$X(t) = [t \ ds \ L(t)_{max} \ L(t)_{min} \ L(t)_{média} \ L(t-3) \ L(t-2) \ L(t-1) \ L(t)] \quad (35)$$

$$Y(t) = [L(t+1)]$$

Dos quais, $X \in \mathbb{R}^m$ e $Y \in \mathbb{R}^1$

Em que:

- t : Tempo
- ds : Dia da semana
- $L(t)_{max}$: Carga máxima em t para cada dia da semana
- $L(t)_{min}$: Carga mínima em t para cada dia da semana
- $L(t)_{média}$: Carga média em t para cada dia da semana
- $L(t-p)$: Carga no tempo t atrasado em p , em que $p = 0, 1, 2, 3$
- $L(t+1)$: Carga em $t+1$

O problema dessa estratégia é que ao longo do tempo novas cargas serão incluídas no banco de dados e conseqüentemente esses valores serão modificados. Sendo assim, sempre que tais modificações se tornem expressivas, recomenda-se novo treinamento da rede com dados mais atualizados.

4.3 CARACTERÍSTICAS DOS DADOS

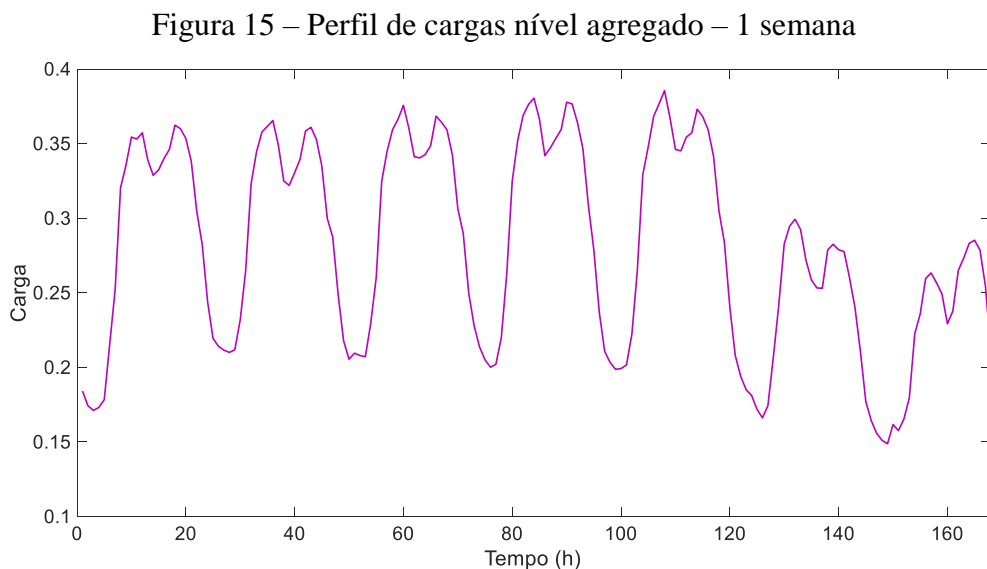
Os perfis de consumo de energia têm alguns níveis de agregação, e em cada nível as curvas de cargas possuem características próprias.

O nível de agregação depende da quantidade de consumidores pertencentes a um perfil de carga. Quanto mais consumidores, mais agregado é o nível. Por exemplo, a curva de carga de uma grande cidade é mais agregada do que a de um bairro, uma vez que há muito mais consumidores na cidade em comparação a um bairro.

Quanto maior o nível de agregação das cargas, mais suaves serão os perfis de cargas e mais fáceis de prever. Ao contrário, os perfis menos agregados são menos suaves e consequentemente são mais difíceis de prever. (SEVLIAN; RAJAGOPAL, 2018; CHITSAZ; SHAKER *et al.*, 2015).

Os dados utilizados para este estudo são de nível desagregado. São compostos de cargas de quatro blocos universitários com variados perfis de cargas. Essas cargas foram medidas utilizando medidores instalados junto ao transformador de cada bloco.

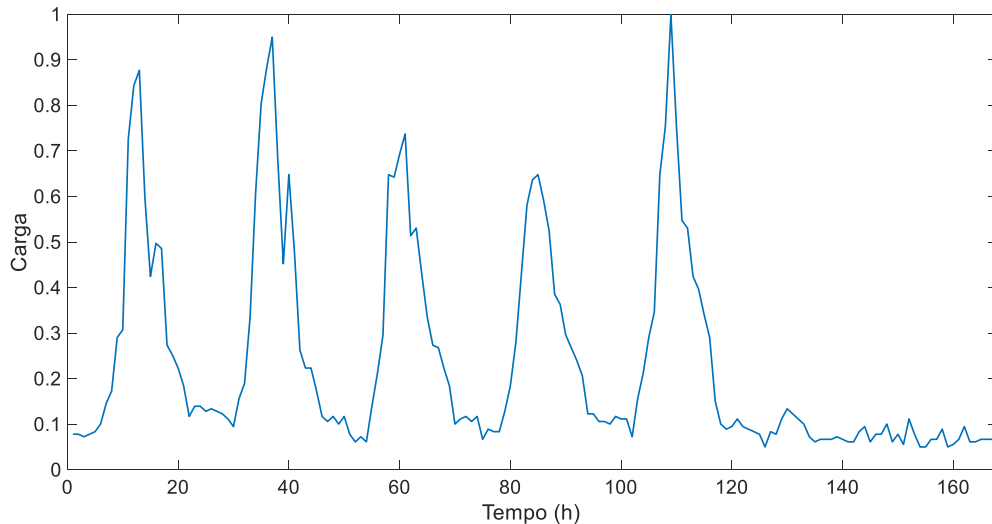
A Figura 15 e a Figura 16 são exemplos de perfis de cargas agregados e desagregados, respectivamente. Por estarem em unidades de medida diferentes os valores usados para essas ilustrações foram normalizados.



Fonte: Elaborado pela autora

As cargas elétricas da Figura 15 foram medidas em megawatts enquanto as cargas elétricas da Figura 16 foram medidas em quilowatts. Esses perfis são correspondentes a uma semana, de segunda a domingo.

Figura 16 - Perfil de cargas nível desagregado – 1 semana

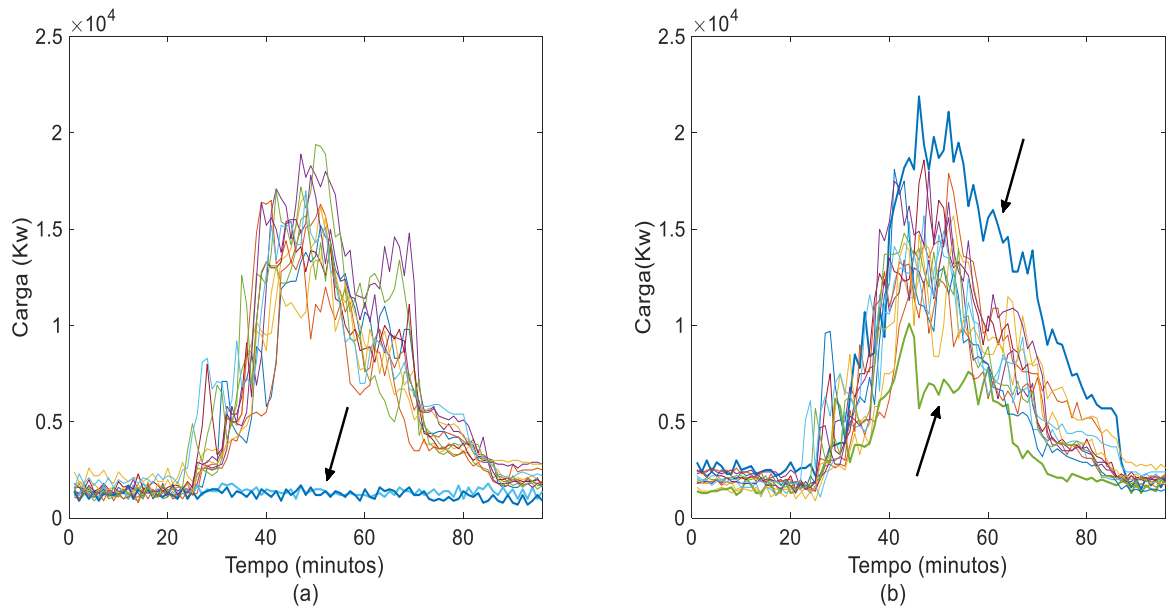


Fonte: Elaborado pela autora

Na Figura 15 nota-se que os perfis de cargas são mais suaves em comparação aos perfis mostrados na Figura 16. Além disso, as curvas de segunda a sexta-feira são similares no nível mais agregado enquanto que no nível menos agregados são ligeiramente distintas. O perfil mostrado na Figura 16 é uma amostra de um dos conjuntos de dados utilizados neste estudo.

Além de serem desagregados, os dados utilizados nesta pesquisa possuem muitos *outliers*, alguns deles por erros de leitura outros por comportamentos atípicos no consumo de cargas. As figuras a seguir são mostradas como exemplos de *outliers* presentes nos dados. Para cada um dos blocos serão mostradas duas figuras lado a lado de dias distintos. Na Figura 17 estão plotadas todas as curvas de todas as segundas e todas as sextas-feiras do bloco 1. No gráfico (a) os *outliers* são as duas curvas que estão próximas de zero, conforme indicado. Essas curvas são provavelmente erros de leitura do medidor ou feriado.

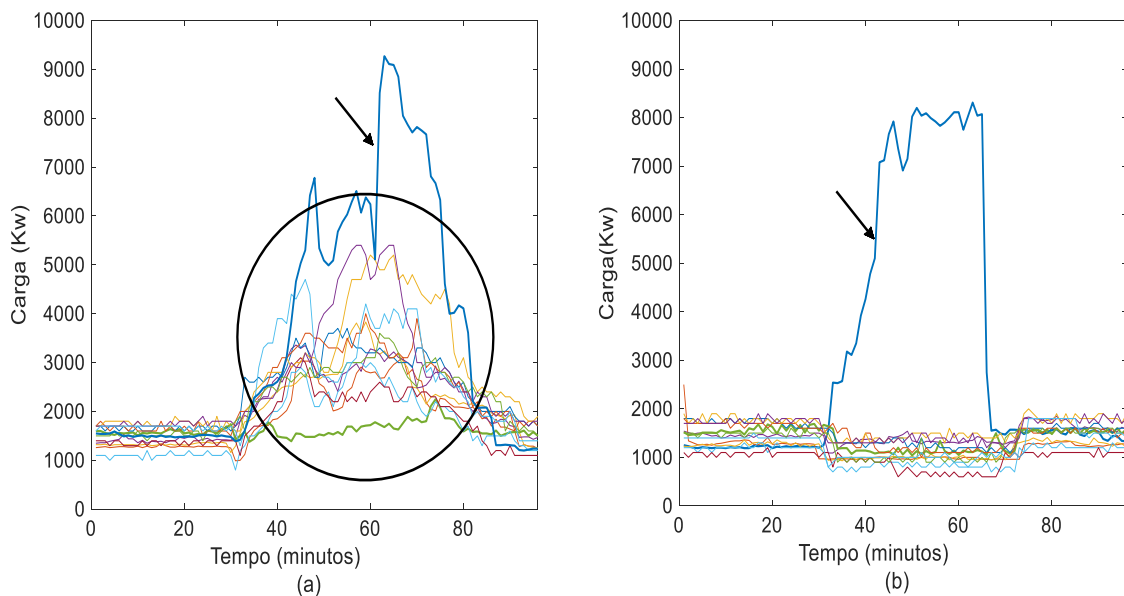
Figura 17 – Curvas de cargas do bloco 1 – (a) Segundas-feiras e (b) Sextas-feiras



Fonte: Elaborado pela autora

Já no gráfico (b) pode-se observar dois *outliers*, um deles abaixo das curvas e um logo mais acima, que pode indicar que foram dias atípicos. Nota-se que mesmo sem a presença de *outliers* as curvas são bem irregulares.

Figura 18 - Curvas de cargas do bloco 2 – (a) Sextas-feiras (b) Sábados

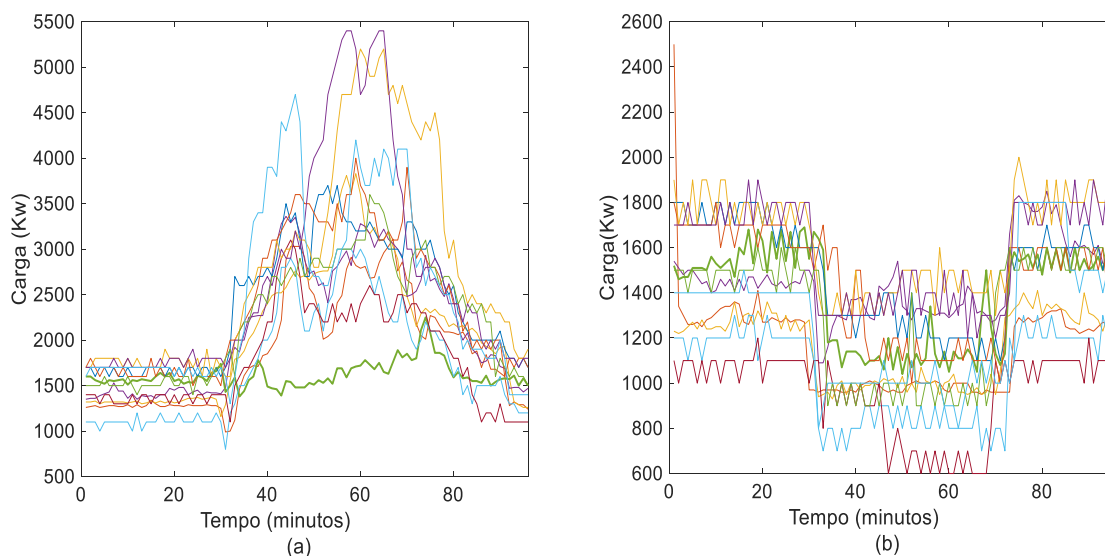


Fonte: Elaborado pela autora

As irregularidades das leituras estão bem evidenciadas no círculo da Figura 18 (a) que representa todas as sextas-feiras do bloco 2. Neste caso está bem difícil destacar *outliers*,

somente um, indicado pela seta, está bem evidente. Já na Figura 18 (b), também há um *outlier* em evidência.

Figura 19 - Curvas de cargas do bloco 2 – (a) Sextas-feiras (b) Sábados – depois da retirada dos *outliers* mais evidentes.

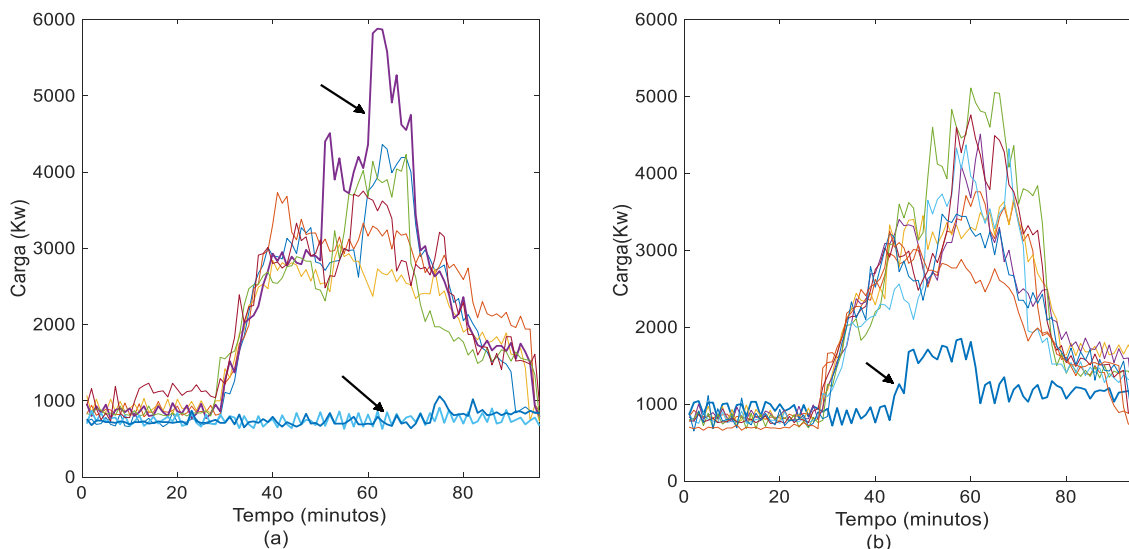


Fonte: Elaborado pela autora

Mesmo com a retiradas dos *outliers* mais evidentes, as leituras de consumo de energia se mostram bastante irregulares (Figura 19).

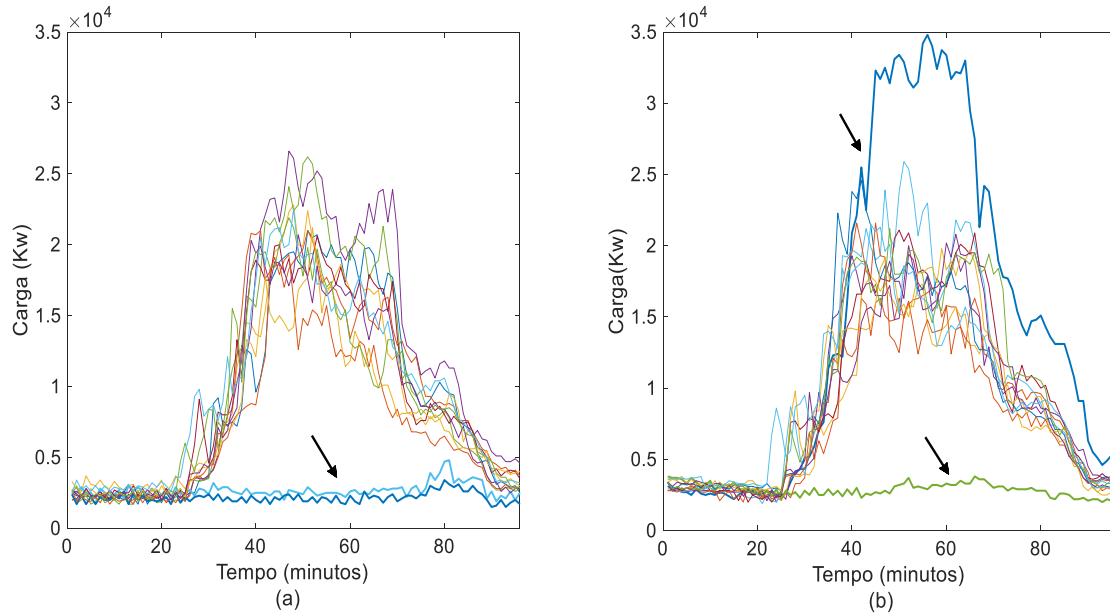
E a mesmas observações são mostradas na Figura 20 e na Figura 21, ou seja, alguns *outliers* por erros de leitura ou comportamentos atípicos. Assim, em todos os blocos, os comportamentos das curvas de carga possuem alguma anomalia de leitura.

Figura 20 - Curvas de cargas do bloco 3 – (a) Segundas-feiras (b) Quartas-feiras



Fonte: Elaborado pela autora

Figura 21 - Curvas de cargas do bloco 4 – (a) Segundas-feiras (b) Quintas-feiras



Fonte: Elaborado pela autora

Diante disso, também foram feitas simulações levando em consideração a remoção dos *outliers*, o que não foi uma tarefa simples, pois a irregularidade dos dados dificultou muito a remoção de alguns deles. Foram removidos dias completos que apresentaram as leituras mais atípicas. Como a quantidade de dados era relativamente pequena, treze semanas, uma remoção mais severa dessa informação poderia acarretar na descaracterização dos dados de estudo.

5 RESULTADOS

Os testes foram realizados utilizando as três metodologias descritas na seção anterior, ou seja, o MMQ, a rede neural ARTMAP *Fuzzy* e a metodologia híbrida. Os resultados foram comparados entre si. Além disso, foram considerados as seguintes situações:

- a) Foram comparados os resultados dos testes com os dados sem a inclusão das cargas máxima, mínima e média *versus* os resultados dos testes com a inclusão dessas cargas;
- b) Foram feitos testes levando em consideração dois valores distintos do parâmetro ρ_a ;
- c) E por fim, foram realizados testes com os dados com *outliers* e sem *outliers*.

As previsões obtidas de acordo com o item a) servem apenas para ilustrar a contribuição das cargas máxima, mínima e média na previsão de cargas. Por este motivo, serão exibidos somente os resultados obtidos pelo MMQ e pela rede neural ARTMAP *Fuzzy* de um único dia de cada bloco.

Os resultados conseguidos observando os itens b) e c) foram feitos nas mesmas simulações, esses itens foram diferenciados com o objetivo de fornecer uma visão mais ampla do estudo.

O parâmetro ρ_a influencia na escolha das categorias em ART_a . Quanto maior este parâmetro, mais categorias serão formadas. Se o valor de ρ_a for muito alto, a quantidade de categorias formadas tenderá a quantidade dos padrões de entrada e de acordo com Carpenter *et al.* (1992) estas categorias serão menos genéricas. Por outro lado, valores menores de ρ_a conduzem a formação de menos categorias, maior generalização e compreensão do código (CARPENTER *et al.*, 1992). De acordo com o algoritmo da rede neural ARTMAP *Fuzzy* o parâmetro ρ_a inicia-se a partir de um valor base mas, é incrementado sempre que uma categoria escolhida em ART_a não passa pelo processo de *match-tracking* no módulo inter-ART.

Assim, o objetivo do item b) é avaliar a quantidade de categorias criadas para $\rho_a = 0,60$ e $\rho_a = 0,95$ e comparar os resultados obtidos no cenário de dados desagregados. Além disso, serão mostrados os valores assumidos por ρ_a a partir de 0,60.

No tocante ao parâmetro ρ_b foi utilizado um valor muito próximo de um, incentivando a formação do maior número de categorias possível. Este valor foi adotado porque valores menores deste parâmetro conduziu à formação de poucas categorias, o que prejudicou a

previsão. Mesmo com a maior exigência na escolha de categorias em ART_b poucas categorias foram criadas, se comparadas à quantidade criada em ART_a .

Em relação ao item c), serão mostrados os resultados que foram obtidos utilizando dados com e sem os *outliers*. Esta parte tem como objetivo mostrar a influência dos *outliers* nos três modelos utilizados neste estudo.

E por fim, fechando esta parte de exibição de resultados serão feitas as comparações dos resultados obtidos entre o MMQ, a rede neural ARTMAP *Fuzzy* e a metodologia híbrida. Com relação à metodologia híbrida serão mostrados também alguns valores de x_a relacionados às categorias escolhidas para o MMQ. A forma de como os testes foram feitos é mostrada no Quadro 03.

Quadro 3 – Quadro da organização dos testes efetuados

Testes	Descrição	Metodologias utilizadas			Valor ρ_a	
1	Previsões para mostrar a contribuição das cargas máxima, mínima e média na previsão	MMQ		RNA	0,95	
2	Previsões utilizando os conjuntos de dados sem a remoção de <i>outliers</i>	MMQ	RNA	Modelo Híbrido	0,60	0,95
3	Previsões utilizando os conjuntos de dados com a remoção de <i>outliers</i>	MMQ	RNA	Modelo Híbrido	0,60	0,95

Fonte: Elaborado pela autora

Com relação aos itens 2 e 3, do Quadro 03, uma análise mais detalhada será realizado considerando os resultados obtidos do bloco 01 com e sem os *outliers*. Em relação aos demais resultados, a análise será mais sucinta pois as conclusões são praticamente as mesmas.

5.1 MÉTRICA PARA A AVALIAÇÃO DOS RESULTADOS

Para avaliar os resultados obtidos foram utilizados o Erro Percentual Absoluto Médio - MAPE pela equação (36).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|L_i(t) - \underline{L}_i(t)|}{L_i(t)} \times 100 \quad (36)$$

em que:

$L(t)$: carga elétrica real no tempo t ;

$\underline{L}(t)$: carga elétrica prevista no tempo t ;

n : número de entradas apresentadas para o diagnóstico.

5.2 ESPECIFICAÇÃO DOS TESTES

Para testar a metodologia híbrida foram realizados vários experimentos com os dados dos quatro blocos universitários do período de 01/05 a 22/07/2013 para os blocos 01, 02 e 04 e o período de 01/05 a 30/06/2013 para o bloco 03. Estes períodos foram escolhidos porque foram os períodos que apresentaram pouquíssimas falhas de leitura. Foram realizados testes ao longo de uma semana, num total de sete dias, conforme apresentado na Quadro 04 para os blocos 01, 02 e 04 e na Quadro 05 para o bloco 03.

Quadro 04 – Especificação dos testes realizados para os blocos 01, 02 e 04.

nº dos testes	Treinamento		Previsão	dia da semana
	Início	Fim		
1	1/5/2013	15/7/2013	16/7/2013	Terça
2	1/5/2013	16/7/2013	17/7/2013	Quarta
3	1/5/2013	17/7/2013	18/7/2013	Quinta
4	1/5/2013	18/7/2013	19/7/2013	Sexta
5	1/5/2013	19/7/2013	20/7/2013	Sábado
6	1/5/2013	20/7/2013	21/7/2013	Domingo
7	1/5/2013	21/7/2013	22/7/2013	Segunda

Fonte: Elaborado pela autora

Havia muitos dados de cargas disponíveis, mas não puderam ser utilizados por possuírem intervalos grandes sem leitura. Em muitos pontos são dias inteiros ou semanas inteiras sem qualquer leitura registrada.

Quadro 05– Especificação dos testes realizados para o bloco 03

nº dos testes	Treinamento		Previsão	dia da semana
	Início	Fim		
1	1/5/2013	23/06/2013	24/06/2013	Segunda
2	1/5/2013	24/06/2013	25/06/2013	Terça
3	1/5/2013	25/06/2013	26/06/2013	Quarta
4	1/5/2013	26/06/2013	27/06/2013	Quinta
5	1/5/2013	27/06/2013	28/06/2013	Sexta
6	1/5/2013	28/06/2013	29/06/2013	Sábado
7	1/5/2013	29/06/2013	30/06/2013	Domingo

Fonte: Elaborado pela autora

5.3 RESULTADOS OBTIDOS

5.3.1 Resultados obtidos com a introdução das cargas máximas, mínima e média

Aqui serão mostrados os resultados obtidos pela inclusão das cargas máxima, mínima e média nos dados usados nas simulações. Os dados utilizados aqui são os com *outliers* e $\rho_a = 0,95$. São resultados ilustrativos com o objetivo apenas para evidenciar a importância na diversificação dos dados. Os resultados mostrados nesta parte são apenas do primeiro teste de cada bloco.

Tabela 8 – MAPEs da primeira etapa de testes dos quatro blocos.

Bloco	Sem máximo, mínimo e média		Com máximo, mínimo e média	
	MMQ	RNA	MMQ	RNA
1	59,26	46,42	38,09	20,70
2	22,65	33,80	11,29	13,69
3	38,98	43,78	34,36	30,47
4	55,38	33,64	53,69	17,06
Média	44,07	39,41	31,12	17,37

Fonte: Elaborado pela autora

Pela Tabela 8 é evidente a melhora significativa nos resultados com a inclusão das cargas máxima, mínima e média. O ganho médio dos valores do MAPE foi de 22,04% para o MMQ e de 48,03% para a RNA.

Esses resultados mostram a importância de se diversificar os conjuntos de entrada de dados, ou seja, quanto mais informação um conjunto possui mais assertivo é um método de previsão. Como foi dito anteriormente, o problema desta técnica é a variação da carga ao longo do tempo. Uma medida que pode ser adotada para minimizar o problema seria adotar um novo treinamento utilizando os dados mais atualizados.

Lembrando que esta estratégia foi adotada por falta de outros tipos de entrada, como variáveis exógenas de tempo ou a taxa de ocupação. Seria interessante fazer testes com dados contendo estes tipos de variáveis para comparar com os resultados obtidos usando as cargas máxima, mínima e média.

5.3.2 Resultados obtidos com os conjuntos de dados do bloco 01 com *outliers*.

Nesta seção, serão mostrados todos os resultados obtidos considerando os dados com *outliers*. Serão feitas comparações entre o MMQ, a rede neural ARTMAP *Fuzzy* e a metodologia híbrida. Além disso, serão feitas considerações em respeito aos resultados dos treinamentos realizados considerando os valores de $\rho_a = 0,60$ e $\rho_a = 0,95$. Para evitar rejeição de categoria na fase de diagnóstico foi considerado um parâmetro de escolha $\rho_a = 0,90$ nas previsões realizadas com os pesos treinados com $\rho_a = 0,95$, indicando que para uma categoria ser selecionada ela deve ser 90% similar à entrada desconhecida, tanto para a rede neural ARTMAP *Fuzzy* quanto para metodologia híbrida. No caso dos diagnósticos feitos com a rede treinada com o parâmetro de escolha menos exigente, o valor permaneceu o mesmo do treinamento, ou seja, $\rho_a = 0,60$. Em todas as tabelas também serão exibidas a média e o desvio padrão dos resultados.

Os resultados do bloco 01 são mostrados na Tabela 9. Esses resultados evidenciam que a metodologia híbrida apresentou os melhores resultados em todas as previsões em ambos valores de ρ_a . Seus desvios padrões revelam que os erros de previsão variaram pouco em relação à média. Diferentemente da rede neural, em que a média dos erros ficou acima de 30% e os desvios-padrões muito superiores ao da metodologia híbrida. Embora o desvio-padrão do MMQ tenha sido relativamente baixo, a média do erro também ficou acima de 30%.

Tabela 9 – Resultados obtidos do bloco 01 – com *outliers*

Nº teste	MMQ	RNA		Híbrida	
		Valores de ρ_a		Valores de ρ_a	
		0,60	0,95	0,60	0,95
1	38,09	21,11	20,70	19,30	19,69
2	28,57	17,15	15,90	14,77	14,68
3	19,61	54,53	33,72	13,37	14,19
4	33,86	56,25	56,07	13,24	15,57
5	28,16	29,22	31,27	16,90	18,13
6	39,49	17,03	16,31	15,34	15,74
7	30,67	53,60	60,78	19,00	17,62
Média	31,21	35,56	33,54	15,99	16,52
Desvio Padrão	6,76	18,46	18,38	2,49	2,01

Fonte: Elaborado pela autora

Os testes 3, 4 e 7 da rede neural apresentaram MAPEs muito discrepantes em relação aos testes 1 e 2, com erros maiores que 50% na maioria dos casos. Esses valores se devem justamente pela presença dos *outliers* nos dados do estudo, além disso, por causa do movimento

de janela, a inclusão de uma carga prevista t com erro muito alto influencia e muito na previsão das cargas $t + 1$ em diante. Uma vez que a rede neural utiliza a semelhança entre a entrada desconhecida com as categorias armazenadas em W^a , quanto mais distante uma entrada estiver de uma possível realidade, mais errôneas serão as previsões.

No caso da metodologia híbrida a propagação do erro foi menor, pois o MMQ leva a uma equação de regressão com erro quadrático mínimo entre vários pontos selecionados pela rede neural. Mesmo que alguns pontos estejam distantes de uma saída mais adequada, os pontos que estão próximos desta saída conduzirão a uma equação de regressão capaz de retornar uma previsão mais próxima do ideal. E por fim, menores erros de previsão em um instante t , conduzem a melhores seleções das várias categorias para $t + 1$ em diante, de acordo com o movimento de janela.

A Tabela 10 mostra os quinze primeiros valores de x_a das categorias escolhidas por T_j . Estes valores foram usados na confirmação categorias por ρ_a e posteriormente para a previsão pelo MMQ na metodologia híbrida. Essas são exemplos das categorias escolhidas por T_j .

Na realidade foram selecionadas cinquenta categorias para cada instante t nesta etapa da previsão. Esses valores se referem à previsão do primeiro dia do bloco 01. Cada uma das colunas de valores de ρ_a é referente a um instante t a ser previsto. Na prática são noventa e seis colunas, ou seja, uma coluna para cada instante a ser previsto, mas aqui serão mostrados apenas seis instantes para ilustrar o comportamento da escolha de várias categorias para a previsão. O comportamento para todos os resultados obtidos é análogo. Os instantes escolhidos na Tabela 10 não estão necessariamente em sequência até para mostrar os valores de x_a em diferentes instantes t .

Tabela 10 – Primeiros valores de x_a na escolha de categorias na previsão - $\rho_a = 0,95$.

t	1	2	47	48	63	64
1	0,9698	0,9777	0,9700	0,9915	0,9681	0,9705
2	0,9777	0,9762	0,9868	0,9794	0,9863	0,9879
3	0,9777	0,9692	0,9844	0,9857	0,9862	0,9871
4	0,9687	0,9656	0,9723	0,9856	0,9856	0,9845
5	0,9757	0,9698	0,9837	0,9850	0,9856	0,9832
6	0,9662	0,9793	0,9836	0,9849	0,9840	0,9814
7	0,9556	0,9773	0,9832	0,9846	0,9833	0,9808
8	0,9647	0,9556	0,9814	0,9845	0,9823	0,9798
9	0,9766	0,9661	0,9812	0,9844	0,9511	0,9795
10	0,9651	0,9884	0,9776	0,9834	0,9806	0,9793

11	0,9732	0,9821	0,9774	0,9820	0,9801	0,9779
12	0,9757	0,9655	0,9768	0,9816	0,9528	0,9500
13	0,9981	0,9663	0,9755	0,9812	0,9775	0,9764
14	0,9981	0,9985	0,9752	0,9801	0,9600	0,9760
15	0,9831	0,9984	0,9747	0,9794	0,9768	0,9756

Fonte: Elaborado pela autora

A escolha das categorias do índice j é baseado no $\max(T_j)$. Note que na Tabela 10 a primeira escolha de T_j não significou realmente a maior similaridade entre a entrada desconhecida para a previsão e as categorias escolhidas armazenadas em W^a . Por exemplo, no caso das categorias escolhidas para $t = 1$. A primeira escolha de T_j retornou um valor de $x_a = 0,9698$ e a segunda escolha foi de $x_a = 0,9777$, ou seja, a similaridade da segunda escolha é maior que a da primeira escolha. E a segunda escolha ainda não é a maior desta amostra, que é $x_a = 0,9981$. Nesta amostra, qualquer uma das quinze escolhas feitas por T_j passariam pelo parâmetro de vigilância $\rho_a = 0,90$ e de acordo com o algoritmo da rede neural ARTMAP *Fuzzy* a escolha para a saída de previsão seria a primeira escolha de T_j que atendesse ρ_a .

A partir de todas as categorias selecionadas para o MMQ, na metodologia híbrida, foi feito um quadro de frequência dos valores de x_a que é mostrado na Tabela 11.

Tabela 11 – Quadro da frequência dos valores de $x_a - \rho_a = 0,95$

Classe	Frequência	%
0,9231 - 0,9423	97	0,29%
0,9423 - 0,9615	805	2,40%
0,9615 - 0,9807	11898	35,41%
0,9807 - 0,9999	20800	61,90%
Total	33.600	100%

Fonte: Elaborado pela autora

Note que das 33.600 categorias selecionadas para todas as previsões do bloco 01 a maioria dos valores de x_a tem alto grau de similaridade com a entrada desconhecida. Mesmo com o uso de $\rho_a = 0,90$ para a seleção de categorias todos os valores de x_a foram maiores que 0,96.

A Tabela 12 é uma amostra dos quinze primeiros valores de x_a obtidos nos testes utilizando $\rho_a = 0,60$ na fase de treinamento. Os resultados são similares aos obtidos para $\rho_a = 0,95$.

Tabela 12 - Quinze primeiros valores de x_a para a escolha de categorias para a fase de previsão na metodologia híbrida considerando os resultados para $\rho_a = 0,60$

t	1	2	46	47	61	62
1	0,9637	0,9752	0,9689	0,9698	0,9674	0,9703
2	0,9751	0,9718	0,9923	0,9868	0,9875	0,9910
3	0,9789	0,9637	0,9890	0,9849	0,9869	0,9613
4	0,8225	0,9751	0,9878	0,9725	0,9637	0,9870
5	0,9748	0,8225	0,9754	0,9836	0,9862	0,9543
6	0,9709	0,9789	0,9859	0,9833	0,9842	0,9840
7	0,9676	0,9854	0,9849	0,9832	0,9839	0,9835
8	0,9792	0,9800	0,9847	0,9823	0,9836	0,9835
9	0,9840	0,9680	0,9815	0,9807	0,9836	0,9833
10	0,9754	0,9608	0,9811	0,9779	0,9823	0,9829
11	0,9602	0,9684	0,9785	0,9772	0,9822	0,9805
12	0,9972	0,9988	0,9784	0,9768	0,9819	0,9801
13	0,9954	0,9987	0,9762	0,9760	0,9819	0,9794
14	0,9675	0,9975	0,9749	0,9749	0,9783	0,9788
15	0,9837	0,9826	0,9744	0,9748	0,9782	0,9787

Fonte: Elaborado pela autora

Note que para $t = 4$ tem-se $x_a = 0,8225$ que é o valor mais baixo da tabela. Veja que depois deste valor há valores de x_a bem maiores do que o valor para $t = 4$. Assim, mais uma vez, foi observado que o primeiro valor escolhido por T_j não significou similaridade alta entre a entrada desconhecida e a categoria escolhida em W^a .

Tabela 13 - Quadro da frequência dos valores de x_a – referente a $\rho_a = 0,60$

Classe	Frequência	%
0,8109 - 0,86788	380	1,13%
0,8734 - 0,9339	20	0,06%
0,9339- 0,9999	33200	98,81%
Total	33600	100%

Fonte: Elaborado pela autora

A distribuição de frequência de x_a mostrada na Tabela 13 refere-se aos testes do bloco 01 levando em consideração a rede treinada com $\rho_a = 0,60$. Da mesma forma que ocorreu na distribuição de frequência mostrada na Tabela 11, as categorias escolhidas se concentraram nos maiores valores de x_a .

Como dito anteriormente, a ideia de fazer testes com dois valores diferentes de ρ_a , um próximo de um e outro mais distante, foi para verificar o comportamento na escolha da quantidade de categorias num cenário mais desagregado.

Tabela 14 – Categorias criadas em ART^a no treinamento da RNA – com *outliers*

Nº teste	Nº padrões	Categorias criadas em ART^a $\rho_a = 0,60$	Categorias criadas em ART^a $\rho_a = 0,95$	% Diferença
1	7292	5435	5455	0,27%
2	7388	5411	5550	1,88%
3	7484	5504	5640	1,81%
4	7580	5693	5802	1,43%
5	7676	5767	5886	1,16%
6	7772	5717	5900	2,35%
7	7868	5797	5961	2,08%

Fonte: Elaborado pela autora

A quantidade de categorias formadas em ambos os valores de ρ_a podem ser visualizadas na Tabela 14. Nela é fácil perceber a pouca diferença na quantidade de categorias formadas entre os dois diferentes valores de ρ_a . No que se refere aos resultados, o treinamento com $\rho_a = 0,95$ leva uma ligeira vantagem. A média e o desvio padrão são um pouco menores nos resultados da rede neural, enquanto que na metodologia híbrida a média foi um pouco maior, mas o desvio padrão foi menor.

A Tabela 15 mostra qual foi a frequência dos valores que ρ_a assumiu no treinamento de acordo em todos os testes do bloco 01. Nela, observa-se que os valores de ρ_a foram, em sua grande maioria, maiores que 0,90. Assim, pode-se dizer que os valores de ρ_a deste experimento se aproximou dos valores de $\rho_a = 0,95$ mesmo sendo iniciado em 0,60. Provavelmente por causa da rejeição no módulo inter-ART.

Tabela 15 - Frequência dos valores de ρ_a no treinamento do bloco 01 – com *outliers*

Classes	Frequência	%
0,60 - 0,70	5589	10,95%
0,70 - 0,80	42	0,08%
0,80 - 0,90	1284	2,52%
0,90 - 1,00	44129	86,45%
Total	51044	100%

Fonte: Elaborado pela autora

A adoção de um valor inicial de $\rho_a = 0,60$ não implicou numa redução drástica do número de categorias formadas em ART^a e conseqüentemente maior generalização do código. Pelo contrário, mesmo com um valor de ρ_a menos exigente a quantidade de categorias formadas se aproximou do número de categorias formadas nos testes com um ρ_a mais exigente.

5.3.3 Resultados obtidos com os conjuntos de dados do bloco 01 sem *outliers*

Neste subtópico serão mostrados os resultados obtidos nos testes do bloco 01 com os *outliers* removidos dos dados de estudo. Da mesma forma que ocorreu no item anterior serão mostrados os valores do MAPE, sua média e desvio padrão. Esses resultados serão comparados aos obtidos dos dados com *outliers*. Além disso, também serão mostrados os resultados dos testes com os dois valores de ρ_a escolhidos, mostrando a quantidade de categorias formadas para cada valor deste parâmetro e a variação para $\rho_a = 0,60$ no treinamento. E por fim, os valores de x_a obtidos para a seleção da metodologia híbrida.

Diante dos resultados mostrados na Tabela 16 ficou evidente a influência que os *outliers* têm no banco de dados.

Tabela 16 - Resultados obtidos do bloco 01 – sem *outliers*

Nº Teste	MMQ	RNA		Híbrida	
		Valores de ρ_a		Valores de ρ_a	
		0,60	0,95	0,60	0,95
1	17,51	22,38	22,38	16,66	18,88
2	24,53	23,68	23,68	18,96	13,57
3	15,90	20,92	20,92	14,04	13,89
4	14,38	18,55	18,55	14,17	12,63
5	16,25	19,62	19,62	14,69	20,12
6	18,59	15,55	15,55	21,35	16,50
7	18,51	22,06	22,06	16,62	16,04
Média	17,95	20,39	20,39	16,64	15,95
Desvio Padrão	3,02	2,54	2,54	2,51	2,60

Fonte: Elaborado pela autora

A melhora é muito acentuada para o MMQ que reduziu em mais de 40% a média do erro e o desvio padrão. No caso da rede neural a redução da média também ficou por volta de 40%, em relação ao desvio padrão houve uma redução drástica de mais de 85%. Note que com a remoção dos *outliers* os resultados da rede neural se tornaram menos discrepantes.

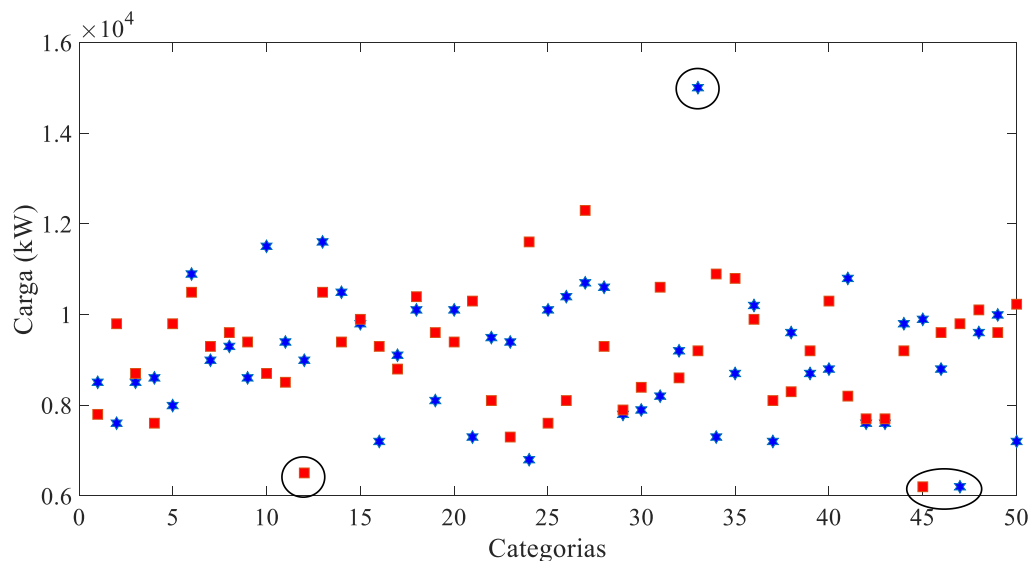
Já em relação à metodologia híbrida a variação na média dos resultados entre os testes com e sem *outliers* variou muito pouco, como pode ser visto na Tabela 17. Nestes testes as médias se aproximaram das obtidas pelo MMQ, mas ainda sim com média de erro e desvio padrão menores.

Tabela 17 – Médias MAPE do Bloco 01 – com *outliers* x sem *outliers*

Com outliers					
	MMQ	RNA		Híbrida	
		Valores de ρ_a		Valores de ρ_a	
		0,60	0,95	0,60	0,95
Média	31,21	35,56	33,54	15,99	16,52
Desvio Padrão	6,76	18,46	18,38	2,49	2,01
Sem outliers					
Média	17,95	20,39	20,39	16,64	15,95
Desvio Padrão	3,02	2,54	2,54	2,51	2,60

Fonte: Elaborado pela autora

Como a metodologia híbrida seleciona várias entradas para a previsão o efeito da presença de algum *outlier* acaba sendo minimizado pelas outras entradas que não são classificadas como tal. A Figura 22 mostra a dispersão das categorias selecionadas em W^b , ou seja, as saídas possíveis para a previsão de uma carga em t . As estrelas se referem às saídas dos dados com *outliers* enquanto os quadrados aos dados sem *outliers*. Nesta amostra nota-se que as dispersões são semelhantes com alguns pontos discrepantes. A média das cargas para a seleção com *outliers* foi de 9.126kW com desvio padrão de 1.529 kW enquanto a média das cargas para a seleção sem *outliers* foi de 9.172,66 kW com desvio padrão de 1.241 kW. Nota-se que a média entre as duas seleções não variou muito. Os desvios apresentaram variação um pouco maior, mas isso a um ponto muito distante, como se pode ver na Figura 22 .

Figura 22 – Dispersão das categorias selecionadas em W_b 

Fonte: Elaborado pela autora

Do ponto de vista da rede neural a saída seria de 10.500 kW para os dados sem *outliers* e de 15.600 kW para os dados com *outliers*. A saída da metodologia híbrida foi de 8.631 kW com *outliers* e 8.773 kW sem *outliers* para uma saída real de 8900 kW.

Na Figura 22 observa-se em destaque pelo menos quatro pontos distantes da grande maioria dos dados, dois representando a saída com *outliers* e dois sem os *outliers*. Um desses pontos, há um ponto muito acima da grande maioria dos dados. Nem a presença deste ponto em particular impediu uma previsão de boa qualidade em relação do treinamento com *outliers*.

Sendo assim, a presença dos *outliers* nos dados não influenciou de forma significativa a previsão da metodologia híbrida, pois a dispersão tende a ser parecida em ambos os casos, ao contrário do que aconteceu com a rede neural e o MMQ dos quais estes modelos não tiveram bom desempenho com os *outliers*.

A seleção de categorias para a metodologia híbrida, sem *outliers*, tem um comportamento semelhante aos testes com *outliers*, ou seja, a maioria das categorias escolhidas por T_j possuem semelhança com as entradas de previsão acima de 93%, como se pode ver na Tabela 18 e na Tabela 19.

Tabela 18 - Frequência dos valores de x_a – referente a $\rho_a = 0,95$ – sem *outliers*

Classes	Frequência	%
0,9011 - 0,9339	95	0,28%
0,9339 - 0,9668	4531	13,49%
0,9668 - 0,9997	28974	86,23%
Total	33600	100,00%

Fonte: Elaborado pela autora

Da mesma forma do que ocorreu nos testes com *outliers* nem sempre a primeira escolha de T_j conduziu realmente a maior semelhança entre a categoria escolhida e a entrada para a previsão.

Tabela 19 - Frequência dos valores de x_a – referente a $\rho_a = 0,60$ – sem *outliers*

corte	Frequência	%
0,8049 - 0,8698	36	0,11%
0,8698 - 0,9347	156	0,46%
0,9347 - 0,9995	33408	99,43%
Total	33600	100,00%

Fonte: Elaborado pela autora

Com a remoção dos *outliers* esperava-se uma maior diferença na formação de categorias entre os treinamentos com $\rho_a = 0,60$ e $\rho_a = 0,95$, uma vez que a remoção de *outliers* diminuiu as grandes discrepâncias dos dados, e consequentemente esperava-se uma maior generalização das categorias. De acordo com Tabela 20, não foi isso que aconteceu. Pelo contrário, a diferença foi menor e mais uniforme. Sendo assim, a remoção dos *outliers* não contribuiu para maior generalização do código para $\rho_a = 0,60$. Logo em nível mais desagregado, a rede neural ARTMAP *Fuzzy* tenderá à formação de muitas categorias, com ou sem *outliers*. Curiosamente, os resultados médios foram exatamente iguais e isso também aconteceu para os outros blocos, exceto para o bloco 04.

Tabela 20 - Categorias criadas em ART^a no treinamento da RNA – sem *outliers*

Nº teste	Nº padrões	Categorias criadas em	Categorias criadas em	Diferença	% Diferença
		ART ^a $\rho_a = 0,60$	ART ^a $\rho_a = 0,95$		
1	5856	4163	4179	16	0,27%
2	5952	4194	4214	20	0,34%
3	6048	4248	4269	21	0,35%
4	6144	4355	4377	22	0,36%
5	6240	4468	4501	33	0,53%
6	6336	4543	4576	33	0,52%
7	6432	4552	4589	37	0,57%

Fonte: Elaborado pela autora

A frequência dos valores atribuídos a ρ_a no treinamento é mostrada na Tabela 21. Da mesma forma que ocorreu com o treinamento dos dados com *outliers* a maior concentração de valores assumidos por ρ_a foi para $\rho_a > 0,90$, sendo assim, mesmo partindo de $\rho_a = 0,60$ o resultado do treinamento se aproximou do resultado para $\rho_a = 0,95$.

Tabela 21- Frequência dos valores de ρ_a no treinamento do bloco 01 – sem *outliers*

classes	Frequência	%
0,60 - 0,70	4950	12,29%
0,70 - 0,80	21	0,05%
0,80 - 0,90	36	0,09%
0,90 - 1,00	35285	87,57%
Total	40292	100%

Fonte: Elaborado pela autora

5.3.4 Resultados obtidos dos demais blocos com e sem *outliers*

Os comentários referentes aos resultados obtidos considerando os dados do bloco 01 foram bem detalhados com o objetivo de fornecer uma visão maior do funcionamento da metodologia híbrida e da rede neural ARTMAP *Fuzzy*. Para evitar redundâncias, as análises com relação aos papéis do parâmetro ρ_a e a escolha das categorias na metodologia híbrida serão omitidas daqui em diante. As conclusões para os outros blocos em relação a essas questões são as mesmas. As distribuições das frequências dos valores x_a para a escolha das categorias na metodologia híbrida e os valores assumidos por ρ_a a partir de 0,60 no treinamento da rede neural estão no apêndice I e apêndice II respectivamente.

Os resultados que serão mostrados daqui em diante serão focados nas previsões feitas com as três metodologias com e sem *outliers* e considerando um valor de $\rho_a = 0,95$. Os resultados apresentados com o parâmetro $\rho_a = 0,60$ estão no apêndice III. As médias e os desvios-padrões serão mostrados no resumo dos resultados.

Tabela 22 - Resultados obtidos do bloco 02

Nº teste	MMQ <i>outliers</i>		ARTMAP <i>Fuzzy</i> <i>outliers</i>		Modelo Híbrido <i>outliers</i>	
	Sim	Não	Sim	Não	Sim	Não
1	11,29	12,21	13,69	13,57	9,90	12,10
2	14,20	11,16	9,52	10,96	10,60	12,08
3	21,52	16,97	15,91	17,88	9,06	8,41
4	15,17	10,79	20,72	24,71	10,40	12,25
5	37,74	20,95	18,76	17,89	14,54	16,96
6	39,35	25,78	11,32	32,32	15,05	15,97
7	15,00	10,81	8,88	12,40	9,70	9,44
Média	22,04	15,52	14,11	18,53	11,32	12,46
Desvio padrão	11,69	5,94	4,56	7,62	2,43	3,12

Fonte: Elaborado pela autora

Nos resultados do bloco 2, mostrados na Tabela 22, a rede híbrida superou todas as metodologias nos testes realizados. Considerando os resultados da rede ARTMAP *Fuzzy*, este bloco apresentou resultados menos discrepantes em comparação aos resultados do bloco 01, neste caso, a rede ARTMAP *Fuzzy* foi capaz de fornecer melhores saídas mesmo na presença dos *outliers*. Note que a média e o desvio padrão dos resultados foram maiores nos testes sem os *outliers*. Esse aumento foi causado principalmente pelo resultado obtido no teste 6.

Tabela 23 - Resultados obtidos do bloco 03

Nº teste	MMQ		ARTMAP Fuzzy		Modelo Híbrido	
	<i>outliers</i>		<i>outliers</i>		<i>outliers</i>	
	Sim	Não	Sim	Não	Sim	Não
1	34,36	11,77	30,47	23,26	15,48	10,35
2	40,05	12,09	10,84	11,68	12,27	11,32
3	35,87	21,42	17,19	18,71	18,90	18,36
4	36,07	19,87	24,90	18,03	18,00	17,43
5	32,20	16,48	13,96	20,47	13,99	14,94
6	33,32	10,44	27,72	15,23	15,83	10,35
7	41,64	14,47	9,42	25,52	12,35	10,69
Média	36,22	15,22	19,21	18,99	15,26	13,35
Desvio padrão	3,47	4,22	8,46	4,68	2,59	3,50

Fonte: Elaborado pela autora

Os resultados do bloco 03 estão exibidos na Tabela 23. A remoção dos *outliers* contribuiu para a melhora significativa nos resultados dos testes com o MMQ, mas não fez muita diferença para a rede neural ARTMAP *Fuzzy*. A metodologia híbrida também conseguiu manter um bom desempenho neste bloco.

Tabela 24 - Resultados obtidos do bloco 04

Nº teste	MMQ		ARTMAP Fuzzy		Modelo Híbrido	
	<i>outliers</i>		<i>outliers</i>		<i>outliers</i>	
	Sim	Não	Sim	Não	Sim	Não
1	53,69	15,35	17,06	32,94	24,84	15,67
2	30,72	16,05	24,83	14,40	11,90	9,04
3	29,80	17,37	54,53	16,16	10,42	11,10
4	34,57	37,47	17,86	17,36	14,67	25,13
5	35,99	15,78	18,72	17,00	17,39	11,80
6	46,28	12,08	23,08	18,81	11,43	9,97
7	28,94	17,27	54,32	10,54	18,13	14,06
Média	37,14	18,77	30,06	18,17	15,54	13,83
Desvio padrão	9,38	8,43	16,88	7,04	5,06	5,48

Fonte: Elaborado pela autora

Da mesma forma que ocorreu nos resultados do bloco 01, também houve discrepâncias em alguns resultados fornecidos pela rede neural do bloco 04, conforme mostra a Tabela 24. E mais uma vez as discrepâncias foram minimizadas pela remoção dos *outliers*. Quase todos os resultados obtidos pelo MMQ também foram melhores sem os *outliers*. Já os resultados da metodologia híbrida se mantiveram dentro da mesma média, tanto com *outliers* quanto sem eles.

Quadro 06- Resumo dos resultados obtidos para os blocos 02, 03 e 04

Bloco 02						
		MMQ	RNA		Híbrida	
			Valores de ρ_a		Valores de ρ_a	
Outliers			0,60	0,95	0,60	0,95
Sim	Média	22,04	16,74	14,11	11,55	11,32
	Desvio Padrão	11,70	6,30	4,56	3,16	2,43
Não	Média	15,53	18,53	18,53	12,11	12,46
	Desvio Padrão	5,94	7,62	7,62	3,77	3,12
Bloco 03						
		MMQ	RNA		Híbrida	
			Valores de ρ_a		Valores de ρ_a	
Outliers			0,60	0,95	0,60	0,95
Sim	Média	36,22	19,01	19,21	14,37	12,46
	Desvio Padrão	3,47	3,78	8,46	2,68	3,12
Não	Média	15,22	18,99	18,99	13,63	13,35
	Desvio Padrão	4,22	4,68	4,68	3,12	3,50
Bloco 04						
		MMQ	RNA		Híbrida	
			Valores de ρ_a		Valores de ρ_a	
Outliers			0,60	0,95	0,60	0,95
Sim	Média	37,14	39,34	30,06	15,69	15,54
	Desvio Padrão	9,39	16,39	16,88	4,85	5,06
Não	Média	18,77	21,91	18,17	14,81	13,83
	Desvio Padrão	8,43	13,03	7,04	5,01	5,48

Fonte: Elaborado pela autora

Diante de todos esses resultados e do resumo mostrado no Quadro 06 fica evidente que a metodologia híbrida apresentou os melhores resultados em todos os cenários testados. Considerando os valores dos desvios-padrões esta metodologia teve os resultados mais estáveis. Além disso, houve pouca variação nos resultados entre os testes com os conjuntos de treinamento com e sem *outliers*.

Ainda assim, apareceram alguns resultados ligeiramente discrepantes, como no teste 1, com *outliers* e o teste 4 sem *outliers* do bloco 04. Uma sugestão para reduzir essas discrepâncias seria incluir algum filtro nos dados selecionados no módulo ART_b , ou seja, as saídas possíveis, para remover algum *outlier* uma vez que, grande similaridade em ART_a com a entrada para a previsão, não garante o mesmo em ART_b com uma saída mais adequada.

6 CONCLUSÕES E SUGESTÕES DE TRABALHOS FUTUROS

6.1 CONCLUSÕES

Inicialmente, a proposta deste trabalho foi realizar previsão de cargas elétricas mistas em nível desagregado em curto prazo, com horizonte de 24 horas, de quatro blocos de prédios de uma universidade brasileira utilizando a rede neural ARTMAP *Fuzzy*. Esta metodologia havia sido escolhida devido ao rápido treinamento e à sua capacidade de aprendizado em modo *online*.

Nos dados disponíveis para testes havia muitas falhas de leitura e *outliers* o que prejudicou algumas das previsões obtidas pela rede neural, que ao invés de dar uma saída adequada possivelmente estava classificando algum *outlier*.

Depois de algumas observações dentro das matrizes de pesos da rede neural foi notado que saídas mais adequadas estavam presentes, mas como o previsor escolhe somente uma, a que tem maior grau de similaridade com a entrada apresentada, ele não conseguia alcançar uma melhor saída.

Para contornar esse problema foi desenvolvida uma metodologia híbrida utilizando a rede neural ARTMAP *Fuzzy* e o método dos mínimos quadrados, do qual a rede neural desempenha o papel de seletor de variáveis e a previsão é realizada pelo MMQ. Esta metodologia foi denominada rede neural ARTMAP *Fuzzy* – MMQ.

Por falta de diversidade de informação nos bancos de dados de estudo, foram introduzidos valores da carga máxima, mínima e média para cada instante e dia da semana. De acordo com os resultados mostrados a inclusão dessas cargas nos dados contribuiu para a melhora dos resultados que foi de 22,04% para o MMQ e de 48,03% para a rede neural ARTMAP *Fuzzy*.

Quanto aos testes feitos com os dois valores diferentes do parâmetro de vigilância, ou seja, $\rho_a = 0,60$ e $\rho_a = 0,95$, houve pouca diferença na formação de categoria para os testes envolvendo esses dois valores. Todos os quadros de distribuição de frequência mostram que mesmo iniciando em 0,60, a imensa maioria dos valores que ρ_a acabaram assumindo valores maiores de 0,90. Quanto aos resultados, as médias e os desvios não se diferenciaram muito entre si.

Nos estudos em relação à metodologia híbrida foi interessante observar a enorme quantidade de categorias que são altamente similares com as entradas usadas para a previsão. De fato, todos os quadros de distribuição de frequência mostraram que a grande maioria das classes selecionadas por T_j tinham similaridade de mais de 90% com as entradas apresentadas.

Os *outliers* influenciaram muito nas previsões feitas pelo MMQ e a rede neural ARTMAP Fuzzy e mesmo com a remoção deles alguns resultados foram discrepantes. Em praticamente todos os testes a rede neural ARTMAP Fuzzy – MMQ apresentou os melhores resultados com as menores médias de MAPE, além disso, se comparadas às outras metodologias, os desvios-padrões foram baixos mostrando que os resultados variaram pouco em relação à média mesmo na presença dos *outliers*.

6.2 SUGESTÕES DE TRABALHOS FUTUROS

Mesmo diante dos bons resultados da metodologia proposta ainda existem pontos que precisam ser explorados.

Um deles é empregar filtros para remover *outliers* dentro do conjunto selecionado, como as saídas possíveis no módulo ART_b na metodologia híbrida. Assim, os dados realmente similares entre si farão parte do processo de previsão, pois mesmo que as categorias selecionadas em ART_a sejam semelhantes entre si, não garante que o mesmo ocorre em ART_b , até porque os dados, por serem desagregados, são muito variados.

Testar a metodologia híbrida com dados diferentes, considerando também variáveis exógenas se possível.

Empregar outra metodologia ao invés do MMQ na previsão da rede híbrida, como por exemplo, o modelo ARIMA.

REFERÊNCIAS

- ABREU, T. *et al.* **Residential electric load curve profile based on fuzzy systems**. In: IEEE PES INNOVATIVE SMART GRID TECHNOLOGIES LATIN AMERICA (ISGT LATAM), 2015, Montevideo. **Proceedings** [...] Piscataway: IEEE, 2015. p. 591-596.
- ALVES, M. F.; LOTUFO, A. D. P.; LOPES, M. L. M. Seleção de variáveis stepwise aplicadas em redes neurais artificiais para previsão de demanda de cargas elétricas. **Proceeding Series of the Brazilian Society of Computational and Applied Mathematics**. Fortaleza, v. 1, n. 1, p. 16, 2013.
- AMBER, K. P. *et al.* Intelligent techniques for forecasting electricity consumption of buildings. **Energy**, Rondebosch, v. 157, p. 886-893, 2018.
- AMBER, K. P.; ASLAM, M. W.; HUSSAIN, S. K. Electricity consumption forecasting models for administration buildings of the UK higher education sector. **Energy and Buildings**, Amsterdam, v. 90, p. 127-136, 2015.
- AMJADY, N.; KEYNIA, F.; ZAREIPOUR, H. Short-Term Load Forecast of Microgrids by a New Bilevel Prediction Strategy. **IEEE Transactions on Smart Grid**, Piscataway, v. 1, p. 286 - 294, 2010.
- ANTUNES, J. F. **Previsão de Carga Multinodal Usando a Rede Neural usando ART-ARTMAP Fuzzy**. 2013. 101 f. Tese (Doutorado em Engenharia Elétrica) - Faculdade de Engenharia de Ilha Solteira, Universidade Estadual Paulista Júlio de Mesquita Filho, Ilha Solteira, 2013.
- BAGNASCO, A. *et al.* Electrical consumption forecasting in hospital facilities: an application case. **Energy and Buildings**, Amsterdam, v. 103, p. 261-270, 2015.
- CARPENTER, G. A.; GROSSBERG, S.; ROSEN, D. B. Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. **Neural Networks**, Kidlington, v. 4, n. 6, p. 759-771, 1991.
- CARPENTER, G. *et al.* Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps. **Neural Networks**. **IEEE Transactions on**, Piscataway, v. 3, n. 5, p. 698-713, 1992.
- CARPENTER, G.; GROSSBERG, S. A self-organizing neural network for supervised learning, recognition, and prediction. **Communications Magazine**, Piscataway, v. 30, p. 38-49, 1992.
- CARPENTER, G.; GROSSBERG, S.; IIZUKA, K. Comparative Performance Measures of Fuzzy ARTMAP, Learned Vector Quantization, and Back Propagation for Handwritten Character Recognition. **IJCNN International Joint Conference on Neural Networks**, Piscataway, v. 1, n. 7, p. 794-799, 1992.
- CHAE, Y. T. *et al.* Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. **Energy and Buildings**, Amsterdam, v. 111, p. 184-194, 2016.

CHITSAZ, H. et al. Short-term electricity load forecasting of buildings in microgrids. **Energy and Buildings**, Amsterdam, v. 99, p. 50-60, 2015.

DEB, C. et al. Forecasting Energy Consumption of Institutional Buildings in Singapore. **Procedia Engineering**, Amsterdam, v. 121, p. 1734-1740, 2015.

DJUKANOVIC, M. et al. A neural-net based short-term load forecasting using moving window procedure. **Electrical Power & Energy System**, Guildford, v. 17, n. 6, p. 391-397, 1995.

ESCRIVÁ-ESCRIVÁ, G. et al. New artificial neural network prediction method for electrical consumption forecasting based on building end-uses. **Energy and Buildings**, Amsterdam, v. 43, p. 3112-3119, 2011.

FARHANGI, H. The path of the smart grid. **IEEE Power and Energy Magazine**, Piscataway, v. 8, n. 1, p. 18 - 28, 2010.

FERLITO, S. *et al.* Predictive models for building's energy consumption: an Artificial Neural Network (ANN) approach. *In: AISEM ANNUAL CONFERENCE*, 18th., 2015, Trento. **Proceedings [...]** Piscataway: IEEE, 2015.

FERNÁNDEZ, I.; BORGES, C. E.; PENYA, Y. K. Efficient Building load forecasting. *In: EMERGING TECHNOLOGIES & FACTORY AUTOMATION (ETFA)*, IEEE CONFERENCE ON., 16th., 2011, Toulouse. **Proceedings [...]** Piscataway: IEEE, 2011. p. 1-8.

FUJIWARA, T.; UEDA, Y. Load forecasting method for Commercial facilities by determination of working time and considering weather information. *In: INTERNATIONAL CONFERENCE ON RENEWABLE ENERGY RESEARCH AND APPLICATIONS (ICRERA)*, 7th, 2018, Paris. **Proceedings [...]** Piscataway: IEEE, 2018. p. 336-341.

GAO, Y. *et al.* A Prediction Approach on Energy Consumption for Public Buildings Using Mind Evolutionary Algorithm and BP Neural Network. *In: IEEE DATA DRIVEN CONTROL AND LEARNING SYSTEMS CONFERENCE (DDCLS)*, 7th., 2018, [s. l.]. **Proceedings [...]** Piscataway: IEEE, 2018. p. 385-389.

GHARAVI, H.; REZA, G. Smart grid: The electric energy system of the future. **Proceedings of the IEEE**, Piscataway, v. 99, n. 6, p. 917-921, 2011.

GROSSBERG, S. Adaptive pattern classification and universal recoding, II: feedback, expectation, olfaction, and illusions. **Biological Cybernetics**, Heidelberg, v. 23, n. 4, p. 187-202, 1976.

GULIN, M. et al. Load forecast of a university building for application in microgrid power flow optimization. *In: ENERGY CONFERENCE (ENERGYCON)*, IEEE INTERNATIONAL, 2014, Cavtat. **Proceedings [...]** Piscataway: IEEE, 2014. p. 1223-1227.

HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001.

JETCHEVA, J. G.; MAJIDPOUR, M.; CHEN, W. P. Neural network model ensembles for building-level electricity load forecasts. **Energy and Buildings**, Amsterdam, v. 84, p. 214-223, 2014.

JOZI, A. *et al.* Day-ahead forecasting approach for energy consumption of an office building using support vector machines. *In: IEEE SYMPOSIUM SERIES ON COMPUTATIONAL INTELLIGENCE (SSCI)*, 2018, Bangalore. **Proceedings [...]** Piscataway: IEEE, 2018. p. 1620-1625.

KE, X.; JIANG, A.; LU, N. Load profile analysis and short-term building load forecast for a university campus. *In: POWER AND ENERGY SOCIETY GENERAL MEETING (PESGM)*, 2016, Boston. **Proceedings [...]** Piscataway: IEEE, 2016. p. 1-5.

LARSON, R.; FARBER, B. **Estatística aplicada**. 4. ed. São Paulo: Pearson Prentice Hall, 2010.

LI, H. Z. *et al.* A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm. **Knowledge-Based Systems**, Amsterdam, v. 37, p. 378-387, 2013.

LIPSCHUTZ, S.; LIPSON, M. L. **Álgebra linear**. Porto Alegre: Bookman, 2011.

LOPES, M. L. M. **Desenvolvimento de redes neurais para previsão de cargas elétricas de sistemas de energia elétrica**. Ilha Solteira: UNESP, 2005.

LÜ, X. *et al.* Modeling and forecasting energy consumption for heterogeneous buildings using a physical–statistical approach. **Applied Energy**, Kidlington, v. 144, p. 261–275, 2015.

MORENO, A. L. **Análise de estabilidade transitória via rede neural ART-ARTmap fuzzy Euclidiana modificada com treinamento continuado**. 2010. 143 f. Tese (Doutorado em Engenharia) - Faculdade de Engenharia de Ilha Solteira, Universidade Estadual Paulista, Ilha Solteira, 2010.

MÜLLER, M. R. **Análise de desempenho da rede neural artificial artmap fuzzy aplicada para previsão multi-step de cargas elétricas em diferentes níveis de agregação**. 2018. 97 f. Tese (Doutorado em Engenharia) - Faculdade de Engenharia de Ilha Solteira, Universidade Estadual Paulista, Ilha Solteira, 2018.

PARK, D. C.; EL-SHARKAWI, M. A.; MARKS II, R. J. Electric load forecasting using an artificial neural network. **IEEE Transaction on Power System**, New York, v. 6, n. 2, p. 442-449, 1991.

PENYA, Y. K. *et al.* Short-term load forecasting in air-conditioned non-residential Buildings. *In: INDUSTRIAL ELECTRONICS (ISIE), IEEE INTERNATIONAL SYMPOSIUM ON GDANSK*, 2011, Poland. **Proceedings [...]** Piscataway: IEEE, 2011. p. 1359-1364.

ROLDÁN-BAY, C. *et al.* Upgrade of an artificial neural network prediction method for electrical consumption forecasting using an hourly temperature curve model. **Energy and Buildings**, Amsterdam, v. 60, p. 38-46, 2013.

RUGGIERO, M. A. G.; LOPES, V. L. D. R. **Cálculo numérico**: aspectos teóricos e computacionais. 2. ed. São Paulo: Makron, 1996.

RUIZ, L. G. B. et al. Energy consumption forecasting based on Elman neural networks with evolutive optimization. **Applied Energy**, Amsterdam, v. 147, p. 466–477, 2015.

SANTOS JÚNIOR, C. R. D. **Uma nova abordagem de treinamento online para rede neural artificial ARTMAP Fuzzy**. 2017. 57 f. Tese (Doutorado em Engenharia) - Faculdade de Engenharia de Ilha Solteira, Universidade Estadual Paulista, Ilha Solteira, 2017.

SCHACHTER, J.; MANCARELLA, P. A short-term load forecasting model for demand response applications. *In*: EUROPEAN ENERGY MARKET (EEM), INTERNATIONAL CONFERENCE ON THE, 11th, 2014, Krakow. **Proceedings** [...] Poland: [s. n.]. 2014. p. 1-5.

SEVLIAN, R.; RAJAGOPAL, R. A scaling law for short term load forecasting on varying levels of aggregation. **International Journal of Electrical Power & Energy Systems**, London, v. 98, p. 350-361, 2018.

TOURETZKY, C. R.; PATIL, R. Building-level power demand forecasting framework using building specific inputs: Development and applications. **Applied Energy**, Amsterdam, v. 147, p. 466-477, 2015.

APÊNDICE A – DISTRIBUIÇÃO DE FREQUÊNCIA DE x_a – DEMAIS BLOCOS

Neste apêndice estão as tabelas de distribuição de frequência de x_a na seleção de entradas do MMQ na metodologia híbrida da Tabela 25 a Tabela 30 dispõem-se aos treinamentos que usaram o parâmetro de vigilância $\rho_a = 0,60$

Tabela 25 – Frequência de x_a no diagnóstico do bloco 02 – Híbrida – com *outliers*

Classes	Frequência	%
0,9384 - 0,9537	44	0,13%
0,9537 - 0,9690	125	0,37%
0,9690 - 0,9843	4206	12,52%
0,9843 - 0,9997	29225	86,98%
Total	33600	100,00%

Fonte: Elaborado pela autora

Tabela 26 - Frequência de x_a no diagnóstico do bloco 02 – Híbrida – sem *outliers*

Classes	Frequência	%
0,9274 - 0,9454	22	0,07%
0,9454- 0,9635	265	0,79%
0,9635 - 0,9816	4251	12,65%
0,9816 - 0,9997	29062	86,49%
Total	33600	100%

Fonte: Elaborado pela autora

Tabela 27 - Frequência de x_a no diagnóstico do bloco 03 – Híbrida – com *outliers*

Classes	Frequência	%
0,9585 - 0,9689	153	0,46%
0,9689 - 0,9792	2061	6,13%
0,9792 - 0,9895	8278	24,64%
0,9895 - 0,9999	23108	68,77%
Total	33600	100,00%

Fonte: Elaborado pela autora

Tabela 28 - Frequência de x_a no diagnóstico do bloco 03 – Híbrida – sem *outliers*

Classes	Frequência	%
0,9086 - 0,9313	10	0,03%
0,9313 - 0,9540	671	2,00%
0,9540 - 0,9767	4824	14,36%
0,9767 - 0,9994	28095	83,62%
total	33600	100,00%

Fonte: Elaborado pela autora

Tabela 29 - Frequência de x_a no diagnóstico do bloco 04 – Híbrida – com *outliers*

Classes	Frequência	%
0,8175 - 0,8782	348	1,04%
0,8782 - 0,9390	6	0,02%
0,9390 - 0,9998	33246	98,95%
Total	33600	100,00%

Fonte: Elaborado pela autora

Tabela 30 - Frequência de x_a no diagnóstico do bloco 04 – Híbrida – sem *outliers*

Classes	Frequência	%
0,8321 - 0,8740	29	0,09%
0,8740 - 0,9159	52	0,15%
0,9159 - 0,9578	997	2,97%
0,9578 - 0,9997	32522	96,79%
Total	33600	100,00%

Fonte: Elaborado pela autora

Da Tabela 31 a Tabela 36 estão as distribuições frequências de x_a são referentes aos treinamentos que utilizaram o parâmetro de vigilância $\rho_a = 0,95$

Tabela 31 - Frequência de x_a no diagnóstico do bloco 02 – Híbrida – com *outliers*

Classe	Frequência	%
0,9535	47	0,14%
0,9689	174	0,52%
0,9843	4261	12,68%
0,9997	29118	86,66%
Total	33600	100,00%

Fonte: Elaborado pela autora

Tabela 32 - Frequência de x_a no diagnóstico do bloco 02 – Híbrida – sem *outliers*

Classes	Frequência	%
0,9449	30	0,09%
0,9632	257	0,76%
0,9815	4211	12,53%
0,9997	29102	86,61%
Total	33600	100%

Fonte: Elaborado pela autora

Tabela 33 - Frequência de x_a no diagnóstico do bloco 03 – Híbrida – com *outliers*

Classes	Frequência	%
0,9688	71	0,21%
0,9791	2110	6,28%
0,9895	8278	24,64%
0,9998	23141	68,87%
Total	33600	100,00%

Fonte: Elaborado pela autora

Tabela 34 - Frequência de x_a no diagnóstico do bloco 03 – Híbrida – sem *outliers*

Classes	Frequência	%
0,9444	38	0,11%
0,9627	1354	4,03%
0,9810	8470	25,21%
0,9993	23738	70,65%
Total	33600	100,00%

Fonte: Elaborado pela autora

Tabela 35 - Frequência de x_a no diagnóstico do bloco 04 – Híbrida – com *outliers*

Classes	Frequência	%
0,9344 - 0,9507	274	0,82%
0,9507 - 0,9670	955	2,84%
0,9670 - 0,9833	9328	27,76%
0,9833 - 0,9995	23043	68,58%
Total	33600	100,00%

Fonte: Elaborado pela autora

Tabela 36 - Frequência de x_a no diagnóstico do bloco 04 – Híbrida – sem *outliers*

Classes	Frequência	%
0,9079 - 0,9309	18	0,05%
0,9309 - 0,9538	593	1,76%
0,9538 - 0,9767	7023	20,90%
0,9767 - 0,9997	25966	77,28%
Total	33600	100,00%

Fonte: Elaborado pela autora

APÊNDICE B - DISTRIBUIÇÃO DE FREQUÊNCIA DE ρ_a – DEMAIS BLOCOS

Distribuição de frequência dos valores assumidos por ρ_a no treinamento da rede neural ARTMAP *Fuzzy* com ρ_a a partir de 0,60.

Tabela 37 – Distribuição de frequência do parâmetro ρ_a bloco 02 – com *outliers*

Classes	Frequência	%
0,60 - 0,70	9657	18,20%
0,70 - 0,80	14	0,03%
0,80 - 0,90	14	0,03%
0,90 - 1,00	43361	81,74%
Total	53046	100,00%

Fonte: Elaborado pela autora

Tabela 38 - Distribuição de frequência do parâmetro ρ_a bloco 02 – sem *outliers*

Classes	Frequência	%
0,60 - 0,70	8021	19,26%
0,70 - 0,80	84	0,20%
0,80 - 0,90	28	0,07%
0,90 - 1,00	33503	80,47%
Total	41636	100,00%

Fonte: Elaborado pela autora

Tabela 39 - Distribuição de frequência do parâmetro ρ_a bloco 03 – com *outliers*

classes	Frequência	%
0,60 - 0,70	1216	3,19%
0,70 - 0,80	247	0,65%
0,80 - 0,90	1178	3,09%
0,90 - 1,00	35525	93,08%
Total	38166	100,00%

Fonte: Elaborado pela autora

Tabela 40 - Distribuição de frequência do parâmetro ρ_a bloco 03 – sem *outliers*

Classes	Frequência	%
0,60 - 0,70	937	3,40%
0,70 - 0,80	21	0,08%
0,80 - 0,90	28	0,10%
0,90 - 1,00	26545	96,42%
Total	27531	100,00%

Fonte: Elaborado pela autora

Tabela 41 - Distribuição de frequência do parâmetro ρ_a bloco 04 – com *outliers*

Classes	Frequência	%
0,60 - 0,70	3674	8,55%
0,70 - 0,80	42	0,10%
0,80 - 0,90	81	0,19%
0,90 - 1,00	39183	91,17%
Total	42980	100%

Fonte: Elaborado pela autora

Tabela 42 - Distribuição de frequência do parâmetro ρ_a bloco 04 – com *outliers*

Classes	Frequência	%
0,60 - 0,70	3619	8,04%
0,70 - 0,80	30	0,07%
0,80 - 0,90	1145	2,54%
0,90 - 1,00	40218	89,35%
Total	45012	100%

Fonte: Elaborado pela autora

APÊNDICE C – RESULTADOS OBTIDOS PARA $\rho_a = 0,60$ – DEMAIS BLOCOS

Resultados obtidos com a rede neural treinada utilizando o parâmetro $\rho_a = 0,60$ dos blocos 02, 03 e 04.

Tabela 43 – Resultados obtidos do bloco 02

Nº teste	MMQ		RNA		Híbrida	
	não	sim	não	sim	não	sim
1	11,30	12,26	10,26	13,57	9,99	14,45
2	14,20	11,16	14,07	10,96	10,31	11,14
3	21,52	16,97	12,05	17,88	8,94	8,20
4	15,17	10,80	20,78	24,71	11,42	9,51
5	37,75	20,96	18,76	17,89	16,04	16,62
6	39,36	25,79	13,01	32,32	15,83	16,67
7	14,97	10,77	28,25	12,40	8,29	8,16
Média	22,04	15,53	16,74	18,53	11,55	12,11
Desvio padrão	11,70	5,94	6,30	7,62	3,16	3,77

Fonte: Elaborado pela autora

Tabela 44 - Resultados obtidos do bloco 03

Nº teste	MMQ		RNA		Híbrida	
	não	sim	não	sim	não	sim
1	34,36	11,77	23,01	23,26	16,85	11,85
2	40,05	12,09	12,83	11,68	10,03	11,68
3	35,87	21,42	17,91	18,71	17,47	17,99
4	36,07	19,87	20,66	18,03	15,67	17,38
5	32,20	16,48	20,10	20,47	14,50	14,85
6	33,32	10,44	15,66	15,23	14,27	10,00
7	41,64	14,47	22,90	25,52	11,78	11,67
Média	36,22	15,22	19,01	18,99	14,37	13,63
Desvio padrão	3,47	4,22	3,78	4,68	2,68	3,12

Fonte: Elaborado pela autora

Tabela 45 - Resultados obtidos do bloco 04

Nº teste	MMQ		RNA		Híbrida	
	não	sim	não	sim	não	sim
1	53,70	15,35	48,46	16,60	24,81	19,31
2	30,72	16,05	52,83	15,12	11,79	12,95
3	29,80	17,37	54,00	50,96	11,38	11,54
4	34,58	37,47	18,23	17,28	13,59	24,04
5	35,99	15,78	29,18	18,95	17,49	11,42
6	46,29	12,08	19,51	13,62	12,50	10,44
7	28,92	17,27	53,14	20,85	18,26	13,99
Média	37,14	18,77	39,34	21,91	15,69	14,81
Desvio padrão	9,39	8,43	16,39	13,03	4,85	5,01

Fonte: Elaborado pela autora