DOI: 10.48011/asba.v2i1.975

Classificação de Distúrbios na Qualidade de Energia Usando Modelagem Logística-NARX Multinomial

Pedro H. O. Silva * **Augusto S. Cerqueira** *
Erivelton G. Nepomuceno ** **Arthur F. Oliveira** ***

* *Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Juiz de Fora, MG, Brasil (e-mail: silva.pedro@engenharia.ufjf.br, augusto.santiago@ufjf.edu.br).*
** *Departamento de Engenharia Elétrica, Universidade Federal de São João del-Rei, MG, Brasil (e-mail: nepomuceno@ufsj.edu.br).*
*** *everis NTT Data, SP, Brasil (e-mail:arthur.filgueiras@engenharia.ufjf.br).*

Abstract: In this work, we propose a novel approach for multiclass classification technique, applied in Power Quality disturbance recognition. The Multinomial Logistic-NARX combines logistic regression with the NARX methodology, resulting into a simple and interpretable model, dealing with the multicollinearity problem and predicting categorical variables. The analysis of the Power Quality problem was based on higher-order statistics and Fisher's discriminant ratio for extraction and selection of parameters in Power Quality events. The results showed high performance rates in comparison to other popular classification techniques.

Resumo: O presente trabalho apresenta uma nova técnica de classificação multiclasses, aplicada para o reconhecimento de distúrbios da Qualidade de Energia. A Modelagem Logística-NARX Multinomial combina a metodologia NARX e a regressão logística, resultando em modelos simples e interpretáveis capazes de lidar com o problema multicolinearidade e de predizer variáveis categóricas. No problema de classificação de distúrbios de qualidade de energia são usadas técnicas baseadas nas estatísticas de ordem superior e discriminante de Fisher para a extração e seleção de parâmetros dos eventos da Qualidade de Energia. Os resultados apresentaram altos índices de desempenho em comparação com outras técnicas populares de classificação.

Keywords: power quality; nonlinear system identification; multiclass classfication; disturbance classification; NARX model.

Palavras-chaves: qualidade de energia; identificação de sistemas não-lineares; classificação multiclasse; classificação de distúrbios; modelo NARX.

1. INTRODUÇÃO

A Qualidade de Energia tem sido foco de muitos trabalhos nos últimos anos, recebendo atenção da academia e das indústrias. Isso ocorre devido ao aumento da utilização de cargas não-lineares, ao avanço da utilização da eletrônica de potência e de sistemas microprocessados no sistema elétrico de potência, a proliferação de geração renovável e distribuída, entre outros fatores (Bollen et al., 2017). As cargas de natureza não-linear podem provocar desvios nas formas de onda da tensão e corrente, em contrapartida, os equipamentos microprocessados são sensíveis às perturbações do sistema elétrico (Mishra, 2019). Qualquer desvio nos valores normais dos sinais de energia são considerados um distúrbio da qualidade de energia. Os distúrbios podem ser responsáveis pelo mau funcionamento de equipamentos e paradas de processos industriais, refletindo em significativas perdas financeiras para as indústrias. Nesse contexto, o desenvolvimento e o aperfeiçoamento de sistemas de monitoramento da Qualidade de Energia, com destaque a classificação e detecção de distúrbios, é propriamente justificada e consiste no objetivo do presente trabalho (Nagata et al., 2018).

Geralmente, o processo de classificação e detecção de distúrbios da Qualidade de Energia tem como pré-processamento as etapas de extração e seleção de características. Na extração de parâmetros, diversos trabalhos abordam a técnica da Estatística de Ordem Superior (EOS) como alternativa ao problema de classificação da Qualidade de Energia (Mendel, 1991; Ferreira, 2005). A EOS é amplamente utilizada devido à imunidade aos ruídos gaussianos, elevado desempenho na extração de parâmetros dos sinais de tensão e custo computacional reduzido. Para reduzir o espaço de parâmetros extraído é utilizado o discriminante linear de Fisher, que seleciona os parâmetros mais representativos que apresentam melhor separabilidade entre as classes (Duda et al., 2012). Por fim, diversos trabalhos apresentam técnicas de classificação para o problema da Qualidade de Energia, métodos base-

ados em *support vector machines* (Nagata et al., 2020), redes neurais (Naik and Kundu, 2014), *random forest* (Zhang et al., 2003) e *k-nearest neighbors* (Pan et al., 2017). Os algoritmos apresentados são amplamente utilizados, mas os seus modelos possuem baixa interpretabilidade, ou seja, o entendimento da relação entre as entradas e saídas do preditor é uma tarefa árdua. Recentemente, diversos trabalhos abordam o conceito de interpretabilidade, que consiste na capacidade do modelo em expressar o comportamento do sistema em uma forma compreensível e transparente (Barredo Arrieta et al., 2020). Essa propriedade subjetiva depende principalmente de fatores como a estrutura do modelo, número de termos e relações de entradas e saídas bem definidas (Roscher et al., 2020). Em Ayala Solares et al. (2019) é proposto o método *logistic-NARX* para classificação binária. O algoritmo *logistic-NARX* apresenta desempenho e interpretabilidade altas, aplicando modelos NARX em processos de classificação binários. No entanto, o método é limitado a duas classes e muitos problemas de classificação exigem uma abordagem multiclasse, portanto o presente trabalho apresenta um novo método de classificação multiclasse baseado em modelos NARX (Akinola et al., 2019; Martins et al., 2013). A Modelagem Logística-NARX Multinomial dispõe da eficiência computacional e a interpretação sem esforço fornecida pela metodologia NARX, bem como a capacidade de predição de saídas categóricas presentes nos modelos de regressão logística.

O método proposto utiliza o algoritmo *Orthogonal Forward Regression* (OFR) para a determinação da estrutura e estimações dos parâmetros dos modelos NARX. Uma consideração importante ao construir um modelo de regressão logística é a multicolinearidade, ou seja, é importante verificar as correlações entre as variáveis preditoras. No cenário ideal, as variáveis preditoras possuem alta correlação com a variável dependente, contudo não devem ser correlacionadas entre si. Esse problema é resolvido utilizando a abordagem OFR, uma vez que na etapa de seleção da estrutura os termos do modelo selecionados são ortogonais entre si. A abordagem de classificação proposta é aplicada ao problema de Qualidade de energia, em que o desempenho é comparado com outros métodos de classificação multiclasse. Os resultados obtidos mostram que o método proposto é uma alternativa competitiva às outras técnicas de classificação.

O presente trabalho é organizado da seguinte forma. A Seção 2, aborda a simulação de eventos da Qualidade de Energia e técnicas de extração e seleção de parâmetros. A Seção 3, inclui um breve resumo da identificação de sistemas não-lineares, uma discussão do algoritmo *Orthogonal Forward Regression* e a Modelagem Logística-NARX Binomial. Na Seção 4, a nova metodologia Modelagem Logística-NARX Multinomial é descrita. A Seção 5, apresenta o sistema proposto para classficação de eventos da Qualidade de Energia. Na Seção 6, o método proposto é aplicado ao problema da Qualidade de Energia, comparando com outros métodos de classificação. As conclusões são apresentadas na Seção 7.

2. CLASSIFICAÇÃO DE EVENTOS NA QUALIDADE DE ENERGIA

2.1 Simulação dos Eventos de Qualidade de Energia

Os eventos de Qualidade de Energia (QE) abordados pelo trabalho são distúrbios isolados refletidos na forma de onda de tensão. Os distúrbios foram gerados sinteticamente seguindo as normas definidas pelo IEEE (2019), pertencentes a cinco diferentes classes: harmônico (C1), (*Sags*)/(*Swell*) (C2), *spikes* (C3), *notchs* (C4) e puro (C5) (veja Tabela 1). Os eventos foram simulados com uma frequência de amostragem $f_s = 15,360$ Hz, ou seja, 256 amostras por ciclo da componente fundamental de 60 Hz. Foram gerados 250 eventos de cada classe, em que o tamanho de cada evento é igual a 4 ciclos da componente fundamental com $N = 1024$ amostras. Além disso, foi considerado uma relação de sinal ruído (SNR) igual a 30 db. A modelagem dos eventos é representada na Tabela 2, em que os parâmetros adotados na simulação foram escolhidos de forma aleatória em intervalos pré-definidos. Os parâmetros apresentam uma distribuição estatística uniforme, atingindo um elevado índice de generalização no processo de classificação.

Tabela 1. Descrição dos eventos da Qualidade de Energia apresentados e suas causas.

Evento	Descrição
Harmônicos	Tensões ou correntes senoidais com frequências que são múltiplos inteiros da frequência fundamental, causada por entradas retificadas e fontes de alimentação usadas em sistemas eletrônicos.
Sag	Diminuição da tensão RMS, associadas às faltas, chaveamento de cargas e partida de motores.
Swell	Aumento da tensão RMS, causada por faltas no sistema de potência.
Spike	Mudança súbita na condição nominal de tensão, causada por descargas elétricas, comutação de cargas e curto-circuitos.
Notch	Distúrbio de tensão periódico causado pela operação de dispositivos de potência quando ocorre o chaveamento de uma fase para outra.

Tabela 2. Modelos matemáticos e parâmetros dos eventos da Qualidade de Energia.

Evento	Equação	Parâmetros
Harmonic	$\text{sen}(\omega t) + \sum_{n=3}^{7} \alpha_n \text{sen}(n\omega t)$	$0,01 \leq \alpha_n \leq 0,2$
Sag/Swell	$(1 - \alpha(u(t-t_1) -$	$0,1 \leq \alpha \leq 0,8$
	$u(t-t_2)))\text{sen}(\omega t)$	$T \leq (t_2 - t_1) \leq 9T$
Spike	$\text{sen}(\omega t) + \alpha \exp(-$	$0,1 \leq \alpha \leq 0,8$
	$(t-t_1)/\tau)u(t-t_1)$	$8\text{ms} \leq \tau \leq 40\text{ms}$
Notch	$\text{sen}(\omega t) - \text{sign}(\text{sen}(\omega t))$	$0,1 \leq K \leq 0,4$
	$\sum_{n=0}^{4} K(u(t-(t_1+0.02n)) -$	$0,1T \leq (t_2 - t_1)$
	$u(t-(t_2+0.02n)))$	$(t_2 - t_1) \leq 0,05T$
Pure	$A\text{sen}(\omega t), \quad A = 1(\text{p.u.})$	$\omega = 2\pi 60$ rad/s

2.2 Extração de Parâmetros

A extração de parâmetros é um passo essencial para garantir uma análise de processamento de sinal confiável, incluindo a elaboração de um classificador de distúrbios. A pesquisa em Estatísticas de Ordem Superior (EOS) pode ser entendida como uma extensão do conceito de correlação (Mendel, 1991). Uma das vantagens da utilização da EOS é representar um conjunto de dados em um novo espaço de parâmetros, no qual a probabilidade de distinguir as propriedades do sinal é maior que no espaço original (Nagata et al., 2020; Ferreira, 2005). As principais características que suportam a escolha da técnica Estatísticas de Ordem Superior é a aplicabilidade em processos não-gaussianos, sistemas não lineares e a função de extrair características dos sinais, como amplitude e fase. As Estatísticas de Ordem Superior podem ser definidas em termos de cumulantes, assumindo um vetor periódico e finito de comprimento N e um sinal $x[n]$ que possui média nula $E\{x[n]\} = 0$. O cálculos dos cumulantes de segunda e quarta ordem podem ser expressos por:

$$\hat{C}_{2,x}[i]) = \frac{1}{N}\sum_{n=0}^{N-1} x[n]x[\lfloor(n+i)\rfloor_N], \tag{1}$$

$$\begin{aligned}\hat{C}_{4,x}[i]) = \frac{1}{N}\sum_{n=0}^{N-1} x[n]x^3[\lfloor(n+i)\rfloor_N] - \\ \frac{1}{N^2}\sum_{n=0}^{N-1} x[n]x[\lfloor(n+i)\rfloor_N]\sum_{n=0}^{N-1} x^2[n],\end{aligned} \tag{2}$$

em que $i = 0, 1, \cdots, N-1$ e $[\lfloor(n+i)\rfloor_N]$ representa o resto da divisão de $n+1$ por N. As equações podem ser usadas para extrair parâmetros de sinais de tensão para fins de análise da qualidade de energia.

2.3 Análise de Discriminante Linear de Fisher

O discriminante linear de Fisher (FDR - *Fisher Discriminat Ratio*) é uma técnica para discriminação de dados multi-dimensionais. Conforme discutido em Duda et al. (2012) e Ferreira (2005) o FDR é utilizado para selecionar conjuntos representativos de parâmetros, que apresentam melhor separabilidade entre as classes. A função custo do critério FDR como ferramenta de seleção de parâmetros é representada como:

$$J_c = (m_1 - m_2)^2 \odot \frac{1}{\sigma_1^2 - \sigma_2^2}, \tag{3}$$

onde $J_c = [J_1 \cdots J_L]$, L é o número total de parâmetros, m_1 e m_2 são os vetores de média das classes, σ_1 e σ_2 são os vetores de variância das classes e o símbolo $\odot$ refere-se ao produto de Hadamard.

3. IDENTIFICAÇÃO DE SISTEMAS NÃO-LINEARES

A identificação de sistemas, como uma abordagem de modelagem baseada em dados, objetiva encontrar um modelo com base nos dados disponíveis que possa representar o mais próximo possível a relação de entrada e saída do sistema (Billings, 2013; Aguirre, 2007). As técnicas de identificação de sistemas não-lineares avançaram muito desde os anos 80. Em particular, as metodologias NARX

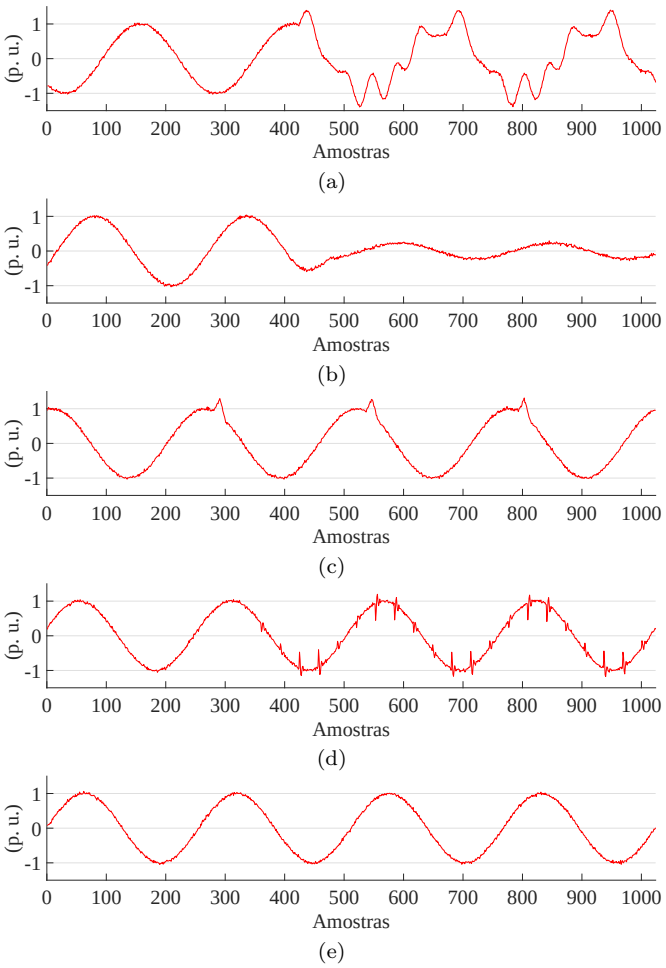

Figura 1. Eventos simulados da QE: (a) Harmônicos, (b) *Sag/Swell*, (c) *Spike*, (d) *Notch* e (e) Puro.

(*nonlinear autoregressive exogenous model*) e NARMAX que provaram ser ferramentas eficientes para identificação de sistemas não-lineares (Wei et al., 2004; Nepomuceno and Martins, 2016). A identificação de sistemas consiste em etapas, incluindo a coleta e processamento de dados, escolha da representação matemática, determinação da estrutura do modelo, estimação de parâmetros e validação do modelo (Söderström and Stoica, 1989). Com relação a escolha da representação matemática, o presente trabalho enfoca os modelos NARX. Para sistemas não-lineares existem inúmeras técnicas para determinação da estrutura do modelo como: algoritmos de agrupamento (*clustering*) (Aguirre and Jácome, 1998), operador de redução e seleção do menor absoluto (Kukreja et al., 2006), *elastic nets* (Zou and Hastie, 2005), programação genética (Sette and Boullart, 2001), metodologia *bagging* (Ayala Solares and Wei, 2015) e o método *Orthogonal Forward Regression* (OFR) usando a abordagem *Error Reduction Ratio* (ERR) (Wei et al., 2004). Uma vez determinada a estrutura do modelo, deve-se estimar seus parâmetros, que pode ser realizada usando o método tradicional dos Mínimos Quadrados, Gradiente Descendente e *Metropolis-Hastings algorithm* (Baldacchino et al., 2012). Por fim, a validação do modelo pode ser realizada mediante testes de correlação estatística, que verificam a validade dos modelos de entrada e saída identificados (Ferreira et al., 2017). Em resumo, a identificação de sistemas é um processo que

cria um modelo parcimonioso que satisfaz um conjunto de testes de acurácia e validade.

3.1 Modelos NARX

As representações NARX são modelos discretos no tempo que explicam o valor de saída $y(k)$ em função de valores prévios dos sinais de saída e de entrada:

$$\begin{aligned} y(k) = f^l(y(k-1), \cdots, y(k-n_y), \\ u(k-1), \cdots, u(k-n_u)) + e(k), \end{aligned} \tag{4}$$

em que f^l representa uma função não-linear do modelo com grau de não-linearidade $l \in \mathbb{N}$, $y(k)$, $u(k)$ e $e(k)$ são respectivamente a saída, entrada e erro de predição, bem como n_y e n_u os atrasos da saída e da entrada. A maioria das abordagens assume que a função f^l pode ser aproximada por uma combinação linear de um conjunto predefinido de funções $\phi_m(\varphi(k))$, portanto a equação (4) pode ser expressa na seguinte forma linear paramétrica:

$$y(k) = \sum_{m=1}^{M} \theta_m \phi_m(\varphi(k)) + e(k), \tag{5}$$

onde θ_m são os coeficientes a serem estimados, $\phi_m(\varphi(k))$ são as funções predefinidas que dependem do vetor regressor $\varphi(k) = [y(k-1), \cdots, y(k-n_y), u(k-1), \cdots, u(k-n_u)]^T$ de saídas e entradas anteriores e M é o número de funções no conjunto. Um dos modelos NARX mais usados é a representação polinomial, em que (5) pode ser denotado na seguinte forma:

$$\begin{aligned} y(k) = \theta_0 + \sum_{i_1=1}^{n} \theta_{i_1} x_{i_1}(k) + \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \theta_{i_1 i_2} x_{i_1}(k) x_{i_2}(k) + \\ \sum_{i_1=1}^{n} \cdots \sum_{i_l=i_l-1}^{n} \theta_{i_1 i_2 \cdots i_l} x_{i_1}(k) x_{i_2}(k) \cdots x_{i_l}(k) + e(k), \end{aligned} \tag{6}$$

considerando $n = n_y + n_u$,

$$x_m(k) = \begin{cases} y(k-m), & 1 \leq m \leq n_y \\ u(k-m+n_y), & n_y + 1 \leq m \leq n \end{cases} \tag{7}$$

sendo l o grau de não-linearidade. O modelo NARX de ordem l significa que a ordem de cada termo no modelo não é superior a l. O número total de termos potenciais em um modelo NARX polinomial é dado por:

$$M = \frac{(n+l)!}{n! \cdot l!}. \tag{8}$$

Algumas das vantagens dos modelos NARX polinomiais é a simplicidade com que a informação analítica sobre a dinâmica do modelo pode ser obtida e a parcimônia, sendo necessário apenas algumas centenas de amostras de dados para estimar um modelo, o que pode ser importante em aplicações que não é realista executar longos experimentos.

3.2 Orthogonal Forward Regression

Em geral, a determinação da estrutura do modelo e a estimação dos parâmetros são realizadas em conjunto. Um dos algoritmos mais populares para realização das duas etapas para a modelagem NARX é o algoritmo *Orthogonal Forward Regression* (OFR) (Billings, 2013). O algoritmo transforma um conjunto de termos candidatos em vetores ortogonais e os classifica com base em sua contribuição para os dados de saída, identificando e ajustando a um modelo NARX determinístico e parcimonioso que pode ser expresso em uma forma de regressão linear generalizada. O algoritmo original *Orthogonal Forward Regression* utiliza a taxa de redução de erro como métrica de dependência. Esse critério associa a cada termo candidato um índice correspondente à contribuição na explicação da variância dos dados de saída do sistema. A taxa de redução de erro é definida como o coeficiente de correlação de Pearson $C(x, y)$ entre dois vetores associados x e y:

$$C(x, y) = \frac{(x^T y)^2}{(x^T x)(y^t y)}. \tag{9}$$

3.3 Modelagem Logística-NARX Binomial

Problemas de classificação ocorrem nas mais diversas áreas do conhecimento, como finanças, saúde e engenharia, onde o objetivo é identificar um modelo que seja capaz de classificar observações ou medições em diferentes categorias ou classes. Muitos métodos e algoritmos foram desenvolvidos para resolver os problema de classificação, incluindo *logistic regression*, *random forest*, *support vector machines*, *k-nearest neighbors* e *logistic-NARX model for binary classification* (Ayala Solares et al., 2019). Os três primeiros algoritmos são amplamente usados, mas a interpretação de tais modelos é uma tarefa difícil. O método *logistic-NARX*, no entanto, dispõe da eficiência computacional e a interpretação simplificada fornecida pela metodologia NARX e a capacidade de predição de variáveis categóricas presente nos modelos de regressão logística. No método regressão logística os valores preditos são probabilidades, portanto estão restritos a valores entre 0 e 1, e utiliza a função logística definida como:

$$f(x) = \frac{1}{1 + \exp(-x)}, \tag{10}$$

em que $x \in \mathbb{R}$ e $f(x)$ é restrito ao intervalo entre 0 e 1. Um problema encontrado na regressão logística é a multicolinearidade, no qual as variáveis independentes possuem relações lineares exatas ou aproximadamente exatas. Caso as variáveis sejam muito correlacionadas as inferências baseadas no modelo de regressão podem ser errôneas ou pouco confiáveis.

No trabalho Ayala Solares et al. (2019), é apresentado o método *logistic-NARX* que trata do problema de multicolinearidade no procedimento de seleção do modelo, verificando as correlações entre as variáveis preditoras. O método é baseado no algoritmo *Orthogonal Forward Regression* para selecionar os termos e combina a função logística com a representação NARX para obter um modelo de probabilidade:

$$p(x) = \frac{1}{1 + \exp\left[\sum_{m=1}^{M} \theta_m \phi_m\left(\varphi(k)\right)\right]}, \tag{11}$$

em Ayala Solares et al. (2019) foi abordado um problema de classificação binomial e o OFR utiliza o índice da taxa de redução de erro como métrica de dependência, portanto, foi proposto o coeficiente de correlação bisserial para quantificar associação entre uma variável contínua e uma variável dicotômica:

$$r(x, y) = \frac{\overline{X_1} - \overline{X_0}}{\sigma_X} \sqrt{\frac{n_1 n_0}{N^2}}, \tag{12}$$

onde $\overline{X_0}$ e $\overline{X_1}$ são os valores médios da variável contínua X para as observações da classe 0 e 1, σ_X é o desvio padrão da variável X, n_0 e n_1 são os números de observações de cada classe e N é número total de pontos dos dados.

Algoritmo 1 Modelagem Logística NARX para classificação multinomial

1: **Input:** Dicionário dos vetores regressores $D = \{ \phi_1, \phi_1, \cdots, \phi_M \}$, sinal de saída y, máximo número de termos n_{max}
2: **Output:** Modelo logístico NARX com os termos selecionados de D, parâmetros estimados θ e intervalos de separação $Y_v(k)$
3: **for all** ϕ_i in D **do**
4: Defina $w_i = \frac{\phi_i}{\|\phi_i\|_2}$
5: Treine modelo logístico *One-Versus-All* com w_i e y
6: Calcule $r^{(i)}(w_i, y)$: acurácia *k-fold cross-validation*
7: Selecione $j = \max_{1 \le i \le M} \{r^{(i)}(w_i, y)\}$
8: Defina $q_1 = w_j$
9: Defina $p_1 = \phi_j$
10: Remova ϕ_j de D
11: **for** $s = 2$ to n_{max} **do**
12: **for all** ϕ_i in D **do**
13: Ortonormalize ϕ_i em relação a $[\ q_1, \cdots q_{s-1}\ ]$ para obter w_i
14: Treine a regressão logística *One-Versus-All* com w_i e y
15: Calcule $r^{(i)}(w_i, y)$: acurácia *k-fold cross-validation*
16: Selecione $j = \max_{1 \le i \le M-s+1} \{r^{(i)}(w_i, y)\}$
17: Defina $q_s = w_j$
18: Defina $p_s = \phi_j$
19: Remova ϕ_j de D
20: Defina $P = [\ p_1\ p_2\ \cdots\ p_n\ ]$: matriz de termos selecionados
21: Defina $\theta = [\ \theta_1\ \theta_2 \cdots\ \theta_n\ ]^T$: vetor de coeficientes estimados
22: Simule o modelo normalizado obtendo $\hat{y}(k)$
23: Calcule $\underline{\hat{y}_v(k)}$ e $\overline{\hat{y}_v(k)}$
24: Compute a acurácia *k-fold cross-validation* com os dados de validação e selecione o modelo com $n \le n_{max}$ termos com a melhor performance
25: **Return:** matriz de termos selecionados $P = [\ p_1\ p_2\ \cdots\ p_n\ ]$, vetor de coeficientes estimados $\theta = [\ \theta_1\ \theta_2 \cdots\ \theta_n\ ]^T$ e intervalos de separação $Y_v(k)$

4. MODELAGEM LOGÍSTICA-NARX MULTINOMIAL

O presente trabalho aborda um problema de classificação de distúrbios na qualidade de energia, que demanda métodos de classificação multiclasse (multinomial), ou seja, algoritmos que permitem categorizar os dados em mais de duas classes. Desta forma, foi desenvolvido a modelagem logística NARX para classificação multinomial. Problemas de multiclasses são normalmente mais complexos que a classificação binária, devido às suas fronteiras de decisão. A extensão direta do algoritmo binário para uma versão multiclasse nem sempre é possível ou fácil de realizar. Portanto, as formas mais exploradas pela comunidade científica são baseadas na binarização de problemas multiclasses. Um dos métodos de decomposição mais utilizados é o *One-Versus-All* (OVA), que faz o uso de C classificadores binários para resolver um problema de classificação envolvendo C classes. Para o v-ésimo classificador binário, é realizado a distinção entre a classe w_v e as demais classes. Dessa forma, um padrão x é classificado pela seguinte regra de decisão:

$$x \in w_v \Leftrightarrow \arg\max_{1 \le v \le C} f_v(x), \tag{13}$$

define-se f_v como o resultado dado pelo modelo referente à classe v, significando a probabilidade entre 0 e 1 de pertencer à v-ésima classe com relação a uma instância x. Em seguida, verifica-se qual a classe com maior probabilidade dada como resultado na equação (13).

Com o objetivo de adaptar o método *logistic-NARX* para a classificação multiclasse, alguns aspectos do algoritmo *Orthogonal Forward Regression* devem ser ajustados. O algoritmo OFR apresentado por Ayala Solares et al. (2019) se baseia no coeficiente de correlação bisserial dado pela equação (9) para determinar a contribuição de cada termo candidato. No entanto, essa métrica não é mais útil, visto que a saída é uma variável categórica devido a multiclasse abordada. Para resolver esse problema, é utilizado um modelo logístico simples usando estimativa de máxima verossimilhança. Basicamente, é calculado a acurácia da predição da variável categórica resultante do modelo logístico baseado nas variáveis contínuas. Uma vez que o classificador resultante possui um alto grau de acurácia, pode-se concluir que as duas variáveis são correlacionadas. Para o cálculo da acurácia foi utilizado o *k-fold cross-validation*, a fim de avaliar a capacidade de generalização do modelo.

O pseudocódigo do método proposto é descrito no Algoritmo 1, em que os regressores (termos do modelo) são selecionados mediante o algoritmo de OFR adaptado. No Algoritmo 1, as linhas $3-8$ objetivam selecionar os termos candidatos que mais contribuem na explicação da variância dos dados de saída do sistema, baseando na acurácia *k-fold cross-validation* de predição do modelo logístico usando o *One-Versus-All*. Na linha 14 novos termos candidatos são ortogonalizados usando o método de Gram-Schmidt em relação aos termos do modelo selecionados anteriormente, e em seguida são avaliados usando a regressão logística. O processo é repetido nas linhas $12-20$ até atingir o número máximo especificado n_{max} de termos do modelo. As linhas $21-23$ são usadas para obter a saída do modelo identificado $\hat{y}(k)$ a partir da matriz de termos selecionados e do vetor de coeficientes estimados. A saída $\hat{y}(k)$ é normalizada para evidenciar a separação das C classes pela a saída identificada do modelo, que foram geradas pelos parâmetros e dados rotulados de treinamento. Desta forma, é obtido $\hat{y}_v(k)$ que representa os conjuntos das saídas referentes a cada classe C:

$$\hat{y}(k) = [\ \hat{y}_1(k)\ \hat{y}_2(k)\ \cdots\ \hat{y}_v(k)\ ], \quad v = 1, 2, \cdots, C. \tag{14}$$

Posteriormente, os limites inferiores e superiores que delimitam as classes são calculados e intervalos são estabelecidos para cada classe na linha 24:

$$\underline{\hat{y}_v} = \arg\min_{1 \le v \le C} \hat{y}_v(k), \qquad \overline{\hat{y}_v} = \arg\max_{1 \le v \le C} \hat{y}_v(k), \tag{15}$$

$$Y_v = [\ \underline{\hat{y}_v},\ \overline{\hat{y}_v}\ ]. \tag{16}$$

Por fim, os dados de validação são simulados utilizando o modelo $\hat{y}(k)$ e normalizados conforme o conjunto de regras

estabelecidos pelos intervalos $Y_v(k)$ que representam as fronteiras de separação das classes.

5. SISTEMA PROPOSTO

A figura 2 ilustra as etapas do sistema proposto para a classificação de eventos da Qualidade de Energia. Na simulação de distúrbios na forma de onda de tensão, os sinais $v[n]$ são pré-processados por um filtro *notch*, sintonizado na frequência fundamental de 60 Hz. O sinal $v[n]$ é decomposto em $e[n]$ representando os distúrbios do sinal original. Na simulação são geradas 250 amostras para cada classe, formando um conjunto de dados balanceados. No caso de conjuntos desbalanceados uma alternativa é o uso de abordagens como SMOTE (*Synthetic Minority Oversampling Technique*), que cria observações sintéticas das classes minoritárias (Chawla et al., 2002). Os dados são separados em conjunto de treinamento (1000 amostras), conjunto de validação (250 amostras) e 5 *folds* foi usado na computação da acurácia *cross-validation*. A técnica Estatística de Ordem Superior é usada na etapa de extração de parâmetros, em que os cumulantes de segunda e quarta ordem são extraídos do sinal de tensão. Com o objetivo de reduzir a dimensão dos parâmetros extraídos e a complexidade computacional, o discriminante de Fisher é utilizado.

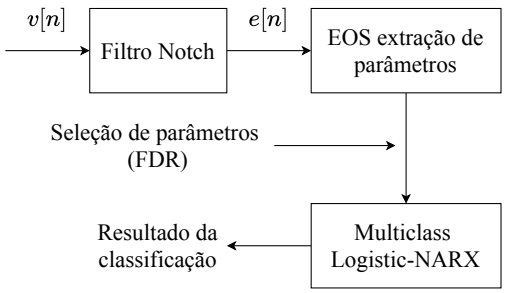

Figura 2. Sistema proposto para classificação da QE.

O discriminante linear de Fisher seleciona os parâmetros que apresentam uma melhor separabilidade entre classes distintas, verificando a distância entre as médias das classes ponderadas pelas suas variâncias. Portanto, é selecionado um conjunto reduzido de dados composto pelos parâmetros mais representativos. Para cada classe foi obtido um vetor J_c contendo $2N$ elementos referentes aos cumulantes de segunda e quarta ordem. A seleção ocorre considerando os maiores valores de J_c relacionados a cada cumulante, obtendo 2 parâmetros para cada classe que resulta em 10 parâmetros por evento. Desta forma, a dimensão original de cada evento foi reduzida de 1024 para 10 amostras. Finalmente, as amostras dos cumulantes selecionados pelo FDR são apresentadas ao método proposto.

6. RESULTADOS

Nesta seção, será realizado a simulação do problema de qualidade de energia, comparando a eficácia da nova metodologia de modelagem logística-NARX em relação aos outros modelos de classificação. A figura 3 demonstra a capacidade de discriminação dos cumulantes, em que as classes são bem caracterizadas utilizando o espaço de parâmetros obtido:

$$u(k) = [\, u_1(k)\ u_2(k)\ \cdots\ u_n(k)\,], \quad n = 1, 2, \cdots, 10. \quad (17)$$

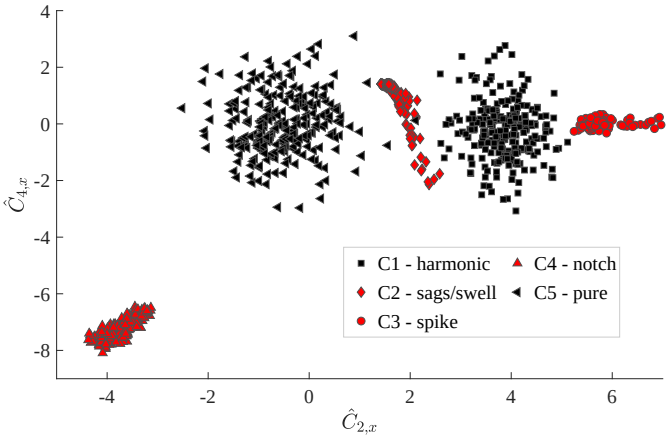

Figura 3. Exemplo dos cumulantes de segunda e quarta ordem selecionados para o espaço de parâmetros dos eventos da Qualidade de Energia.

Após a avaliação do espaço de parâmetros, o algoritmo proposto é aplicado no conjunto de dados da Qualidade de energia. Foi considerado 250 amostras para cada classe de distúrbio $k = 1250$, e grau de não-linearidade $l = 2$ que resulta no espaço de busca com 66 termos potenciais do modelo. O número máximo de termos selecionados é $n_{max} = 10$ e 5 *folds* foi usado na computação da acurácia *cross-validation*. A figura 4 representa a acurácia da validação cruzada na seleção dos termos do modelo na aplicação do Algoritmo 1. O resultado sugere que o modelo com melhor desempenho possui 2 termos no exemplo proposto, mostrando que o algoritmo é capaz de identificar os termos envolvidos na fronteira de decisão do problema de classificação. A tabela 3 apresenta os termos selecionados do modelo identificado $\hat{y}(k)$, o índice de correlação correspondente e seus parâmetros estimados. Desta forma, é possível obter a saída do modelo identificado $\hat{y}(k)$ do conjunto de treinamento e normalizá-la para evidenciar a separação das classes.

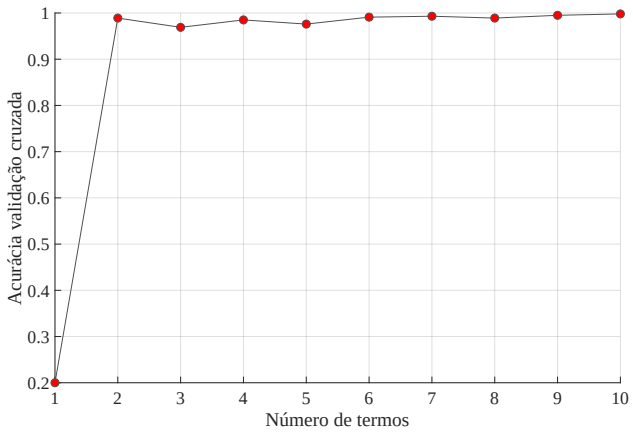

Figura 4. Acurácia da validação cruzada.

A Figura 5 representa os conjuntos de saídas $\hat{y}_v(k)$ referente a cada classe C delimitado na saída do modelo identificado $\hat{y}(k)$. A saída do modelo exibe a separabilidade entre as classes, em que o próximo passo é estabelecer os limites inferiores $\underline{\hat{y}_v(k)}$ e superiores $\overline{\hat{y}_v(k)}$ que as delimitam. Um exemplo é o conjunto $\hat{y}_2(k)$ mostrado na Figura 5, que destaca o intervalo da classe 2, inclusive um erro na

classificação do modelo no treinamento. Os intervalos Y_v são calculados para cada classe e formam um conjunto de regras que representam as fronteiras de separação das classes. Posteriormente, os dados de validação são aplicados no algoritmo proposto e a saída do modelo NARX identificado é processado seguindo as regras estabelecidas pela fase de treinamento. A Figura 6 exibe a simulação do modelo identificado na fase de validação, em que foi observado que as classes 2 e 5 apresentam erros de predição.

Tabela 3. Modelo NARX identificado $\hat{y}(k)$ e os valores dos parâmetros estimados.

Termo do modelo	Correlação	Parâmetro
$u_4(k)$	$0,9910$	$-1,411$
constante	$0,8182$	5

Para verificar o desempenho do algoritmo proposto é realizado um teste com outros métodos de classificação utilizando o mesmo conjunto de dados. A abordagem *random forest* (RF) com máximo número de 500 *splits* e *Gini's diversity index*, *support vector machines* (SVM) com *radial basis kernel* e *k-nearest neighbors* (KNN). Os modelos são comparados usando a acurácia *k-fold cross-validation* com 5 *folds* em um conjunto de dados $k = 1250$.

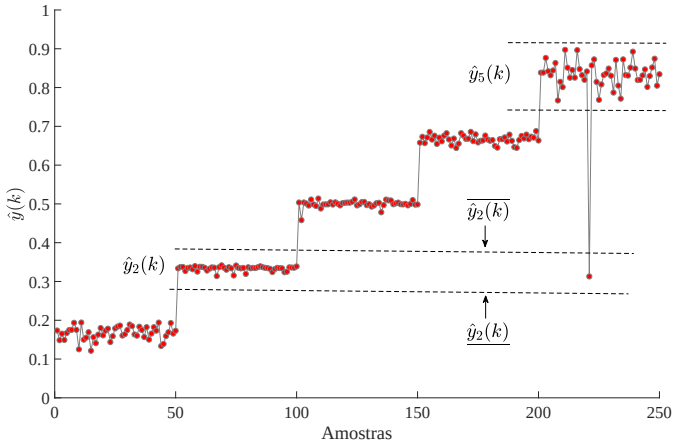

Figura 5. Conjunto de saídas $\hat{y}_v(k)$ referentes a cada classe C da simulação do modelo identificado $\hat{y}(k)$ usando o conjunto de treinamento.

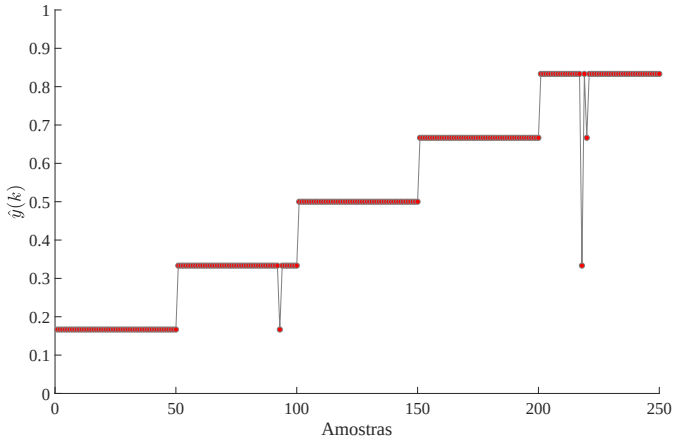

Figura 6. Validação do modelo identificado $\hat{y}(k)$ usando o conjunto de validação.

Os índices de desempenho referentes a cada método de classificação são apresentados na Tabela 4. O método proposto possui desempenho semelhante às demais técnicas, tornando uma alternativa para problemas de classificação multiclasse. O desempenho das técnicas *random forest* e *k-nearest neighbors* foram inferiores para o problema proposto. Embora que o método apresentado obteve desempenho inferior ao SVM, o modelo logístico-NARX é naturalmente interpretável e a contribuição individual dos regressores pode ser conhecida. A Tabela 5 mostra a matriz

Tabela 4. Comparação de desempenho entre diferentes métodos de classificação.

Método	Acurácia	Sensibilidade	Especificidade	Precisão
Proposto	0,9976	0,9976	0,9994	0,9976
RF	0,9936	0,9936	0,9984	0,9936
SVM	0,9984	0,9984	0,9996	0,9984
KNN	0,9944	0,9944	0,9986	0,9944

de confusão na validação do algoritmo, foi possível observar que a classe 5 apresenta erros superiores de predição em comparação às demais. O número de erros na predição estão associados a dispersão das saídas do modelo $\hat{y}_v(k)$ identificado referente a cada classe. Por exemplo, a Figura 5 mostra uma maior dispersão da saída do modelo referente a classe 5 em comparação as outras. A metodologia proposta para classificação de distúrbios da qualidade de energia composta pela extração e seleção de parâmetros, obteve sucesso em conjunto com o método proposto de classificação.

Tabela 5. Matriz de confusão gerada pelo algoritmo proposto na validação.

	C1	C2	C3	C4	C5
C1	250				
C2		250			
C3			250		
C4				250	
C5		2		1	247

7. CONCLUSÃO

Neste trabalho, foi proposto um sistema de classificação de distúrbios da Qualidade de Energia, abordando uma nova metodologia de classificação multiclasse. O algoritmo apresentado lida com multicolinearidade e fornece modelos NARX simplificados para realizar a predição de variáveis categóricas. O modelo obtido é relativamente simples e intuitivo de interpretar, fornecendo *insights* (coeficientes) que explicam claramente o impacto incremental de uma variável preditora na variável de resposta. Em geral, os resultados mostram que o sistema proposto é apropriado para aplicações de Qualidade de Energia, obtendo índices de desempenho em 99%. Finalmente, o algoritmo proposto superou ou pelo menos foi concorrente com técnicas populares como *random forest*, *support vector machines* e *k-nearest neighbors*.

Como proposta futura deseja-se avaliar os modelos identificados correspondente a cada classe usando técnicas mais recentes de validação da identificação de sistemas, podendo interpretar a qualidade dos preditores e adicionar as informações no processo seleção. Outra abordagem desejada é a aplicação do método em dados desbalanceados, visto que é típico em muitas aplicações reais a dominância de classes em relação as demais. Diversas abordagens estão disponíveis para lidar com o problema de dados desequilibrados (Chawla et al., 2002; Kuhn and Johnson, 2013).

AGRADECIMENTOS

Agradecemos à CAPES, CNPq, INERGE, FAPEMIG e à Universidade Federal de Juiz de Fora (UFJF) pelo apoio.

REFERÊNCIAS

Aguirre, L. (2007). *Introdução à identificação de sistemas - Técnicas lineares e não-lineares aplicadas a sistemas reais*. Editora UFMG.

Aguirre, L. and Jácome, C. (1998). Cluster analysis of NARMAX models for signal-dependent systems. *IEE Proceedings - Control Theory and Applications*, 145(4), 409–414.

Akinola, T.E., Oko, E., Gu, Y., Wei, H.L., and Wang, M. (2019). Non-linear system identification of solvent-based post-combustion CO2 capture process. *Fuel*, 239.

Ayala Solares, J.R. and Wei, H.L. (2015). Nonlinear model structure detection and parameter estimation using a novel bagging method based on distance correlation metric. *Nonlinear Dynamics*, 82(1-2), 201–215.

Ayala Solares, J.R., Wei, H.L., and Billings, S.A. (2019). A novel logistic-NARX model as a classifier for dynamic binary classification. *Neural Computing and Applications*, 31(1), 11–25.

Baldacchino, T., Anderson, S.R., and Kadirkamanathan, V. (2012). Structure detection and parameter estimation for NARX models in a unified EM framework. *Automatica*, 48(5), 857–865.

Barredo Arrieta, A., Díaz-Rodríguez, N., and Del Ser, J. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115.

Billings, S.A. (2013). *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons.

Bollen, M.H.J., Das, R., Djokic, S., Ciufo, P., Meyer, J., Ronnberg, S.K., and Zavodam, F. (2017). Power Quality Concerns in Implementing Smart Distribution-Grid Applications. *IEEE Transactions on Smart Grid*, 8(1), 391–399.

Chawla, N.V., Bowyer, K.W., Hall, L.O., and Kegelmeyer, W.P. (2002). SMOTE: Synthetic Minority Oversampling Technique. *Journal of Artificial Intelligence Research*, 16(2), 321–357.

Duda, R.O., Hart, P.E., and Stork, D.G. (2012). *Pattern classification*. John Wiley & Sons.

Ferreira, D.D. (2005). HOS-based method for classification of power quality disturbances. *Electronics letters*, 41(2).

Ferreira, D.D., Nepomuceno, E.G., Cerqueira, A.S., and Mendes, T.M. (2017). A model validation scale based on multiple indices. *Electrical Engineering*, 99(1), 325–334.

IEEE (2019). IEEE Recommended Practice for Monitoring Electric Power Quality. *IEEE Std 1159-2019 (Revision of IEEE Std 1159-2009)*, 1–98.

Kuhn, M. and Johnson, K. (2013). *Applied predictive modeling*. Springer.

Kukreja, S.L., Löfberg, J., and Brenner, M.J. (2006). A least absolute shrinkage and selection operator (LASSO) for nonlinear system identification. *IFAC Proceedings Volumes*, 39(1), 814–819.

Martins, S.A.M., Nepomuceno, E.G., and Barroso, M.F.S. (2013). Improved structure detection for polynomial NARX models using a multiobjective error reduction ratio. *Journal of Control, Automation and Electrical Systems*, 24(6), 764–772.

Mendel, J. (1991). Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. *Proceedings of the IEEE*, 79(3), 278–305.

Mishra, M. (2019). Power quality disturbance detection and classification using signal processing and soft computing techniques: A comprehensive review. *International Transactions on Electrical Energy Systems*, 29(8).

Nagata, E.A., Ferreira, D.D., Bollen, M.H., Barbosa, B.H., Ribeiro, E.G., Duque, C.A., and Ribeiro, P.F. (2020). Real-time voltage sag detection and classification for power quality diagnostics. *Measurement*, 164.

Nagata, E.A., Ferreira, D.D., Duque, C.A., and Cequeira, A.S. (2018). Voltage sag and swell detection and segmentation based on Independent Component Analysis. *Electric Power Systems Research*, 155, 274–280.

Naik, C.A. and Kundu, P. (2014). Power quality disturbance classification employing S-transform and three-module artificial neural network. *International Transactions on Electrical Energy Systems*, 24(9), 1301–1322.

Nepomuceno, E.G. and Martins, S.A.M. (2016). A lower bound error for free-run simulation of the polynomial NARMAX. *Systems Science & Control Engineering*, 4(1), 50–58.

Pan, D., Zhao, Z., Zhang, L., and Tang, C. (2017). Recursive clustering K-nearest neighbors algorithm and the application in the classification of power quality disturbances. In *2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)*, 1–5. IEEE.

Roscher, R., Bohn, B., Duarte, M.F., and Garcke, J. (2020). Explainable Machine Learning for Scientific Insights and Discoveries. *IEEE Access*, 8, 42200–42216.

Sette, S. and Boullart, L. (2001). Genetic programming: principles and applications. *Engineering Applications of Artificial Intelligence*, 14(6), 727–736.

Söderström, T. and Stoica, P. (1989). *System identification*. Prentice-Hall International.

Wei, H.L., Billings, S.A., and Liu, J. (2004). Term and variable selection for non-linear system identification. *International Journal of Control*, 77(1), 86–110.

Zhang, H., Liu, P., and Malik, O. (2003). Detection and classification of power quality disturbances in noisy conditions. *IEE Proceedings - Generation, Transmission and Distribution*, 150(5), 567.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320.