

# CLASSIFICAÇÃO DE EVENTOS EM REDES DE DISTRIBUIÇÃO DE ENERGIA UTILIZANDO TRANSFORMADA WAVELET E MODELOS NEURAI AUTÔNOMOS\*

Vitor Hugo Ferreira<sup>1</sup>, André Lazzaretti<sup>2</sup>, Hugo Vieira Neto<sup>3</sup>, Rodrigo Riella<sup>2</sup>, Julio Omori<sup>4</sup>

<sup>1</sup>Universidade Federal Fluminense – UFF

<sup>2</sup>Instituto de Tecnologia para o Desenvolvimento – LACTEC

<sup>3</sup>Universidade Tecnológica Federal do Paraná – UTFPR

<sup>4</sup>Companhia Paranaense de Energia – COPEL

lazzaretti@lactec.org.br, vitor@vm.uff.br, hvieir@utfpr.edu.br, riella@lactec.org.br, julio.omori@copel.com

**Resumo** – Este trabalho apresenta um método para classificação automática de oscilografias correspondentes a faltas e eventos relacionados à qualidade de energia em redes de distribuição de energia elétrica. O método proposto é dividido em dois estágios: pré-processamento e classificação dos eventos. No primeiro estágio, os sinais das oscilografias são decompostos utilizando a transformada *wavelet*. A energia presente em cada sub-banda do domínio *wavelet* é então calculada para compor vetores de características, que são utilizados como entradas para os classificadores do segundo estágio. Os classificadores investigados são baseados em redes neurais artificiais *feed-forward* do tipo *Multi-Layer Perceptron (MLP)* e *Support Vector Machines (SVM)*, as quais são capazes de promover de maneira automática a seleção de entradas e o controle de complexidade da rede simultaneamente. Experimentos usando dados simulados forneceram resultados promissores para a classificação de eventos de qualidade de energia.

**Palavras-chave** – Classificação de faltas, *wavelets*, seleção de entradas, controle de complexidade, inferência Bayesiana aplicada a MLPs, Máquinas de Vetor de Suporte.

**Abstract** – This work proposes a method for automatic classification of oscillographies corresponding to faults and events related to service quality in electricity distribution networks. The proposed method is divided in two stages: pre-processing and classification of events. In the first stage, oscillography signals are decomposed using the wavelet transform. The energy present in each sub-band of the wavelet domain is then computed in order to compose feature vectors, which are fed to the automatic classifiers of the second stage. The classifiers investigated are based on Multi-Layer Perceptron (MLP) feed-forward artificial neural networks and Support Vector Machines (SVM), which are able to promote feature selection and network complexity control simultaneously. Experiments using simulated data yielded promising results in service quality event classification.

**Keywords** – Fault classification, wavelets, feature selection, complexity control, Bayesian inference applied to MLPs, Support Vector Machines.

## 1. INTRODUÇÃO

Em função do crescente uso de dispositivos eletrônicos de potência e controles industriais, houve um aumento da preocupação, principalmente por parte das concessionárias de energia, com eventos relacionados à Qualidade de Energia Elétrica (QEE). Tais eventos são caracterizados por alterar a forma de onda de tensão ou corrente, devido a curtos-circuitos ou má operação de equipamentos instalados nos consumidores de energia [2]. Alterações nas formas de onda podem por vezes resultar em danos aos consumidores, forçando as concessionárias a monitorar o fornecimento da energia para que esta seja entregue de forma adequada [3] nos vários estágios do sistema elétrico de potência.

Tendo em vista essa realidade, a Companhia Paranaense de Energia (COPEL) realizou em conjunto com o Instituto de Tecnologia para o Desenvolvimento (LACTEC) um projeto dentro do programa de pesquisa e desenvolvimento da Agência Nacional de Energia Elétrica (ANEEL), intitulado “Sistema de Monitoramento da Qualidade de Energia Elétrica para Barras de Distribuição”, cujo objetivo é o de possibilitar o monitoramento contínuo da tensão em barras das subestações de distribuição de energia, armazenando oscilografias dos eventos de QEE [4].

A detecção de eventos é feita com base em valores pré-programados de níveis de tensão, ou seja, toda vez que a forma de onda de tensão extrapola os limites especificados, é feito o armazenamento da forma de onda, para sua análise posterior. A dificuldade encontrada nessa abordagem é o grande volume de dados gerados pelos monitores de QEE, dificultando a análise dos mesmos por parte dos engenheiros de proteção e manutenção. A grande quantidade de dados impede que seja feita uma identificação manual para cada oscilografia registrada, resultando em falta de classificações adequadas para os vários eventos, na maioria das oscilografias armazenadas.

\*Versão estendida do artigo originalmente publicado nos anais do IX Congresso Brasileiro de Redes Neurais / Inteligência Computacional [1].

Dessa maneira, com o intuito de se obter uma base de dados com oscilografias de eventos já classificados, inicialmente foi construído um modelo simplificado de uma subestação de distribuição, bem como de seus alimentadores, utilizando o *software* ATP (*Alternative Transient Program*) [5], com o intuito de ser utilizado para treinar e testar o estágio de classificação do sistema proposto. O uso do ATP para esse tipo de modelagem já é bastante conhecido e consolidado – espera-se que os resultados obtidos com essas simulações possam servir de base para o desenvolvimento de um classificador a ser aplicado em dados reais. Grande parte dos eventos reais é resultante de faltas no sistema elétrico [2], sendo que cerca de 70% destas ocorrem no sistema de distribuição [6] e o restante é devido a eventos de qualidade de energia. Essa proporção foi mantida nas simulações realizadas no presente trabalho.

O processo de classificação utilizado já vem sendo muito discutido e aplicado em diversos trabalhos na área, principalmente nos níveis de transmissão [7, 8] e distribuição [9, 10], consistindo basicamente em estágios de pré-processamento, classificação e pós-processamento [11]. No primeiro estágio são extraídas características dos sinais em análise, através do uso de uma transformação dos sinais no domínio do tempo para o domínio da frequência – via transformada de Fourier – ou para o domínio tempo-frequência – via transformada *wavelet*. A classificação é realizada frequentemente por uma rede neural artificial e, seguida de um pós-processamento – normalmente utilizando lógica *fuzzy* – com o intuito de se efetuar o processo final de decisão, atribuindo um “grau de confiança” ao resultado dado pela rede neural para cada classe do problema.

No presente trabalho foi escolhido um pré-processamento baseado no cálculo da transformada *wavelet* [9, 10], em função do reduzido tamanho do vetor de características obtido, quando comparado com o vetor produzido pela transformada de Fourier, fato que facilita o processamento dos dados como um todo, sem perdas na capacidade de classificação [12]. Em seguida, na etapa de classificação propriamente dita, foram avaliados dois algoritmos baseados em redes neurais – um MLP (*Multi-Layer Perceptron*) e uma SVM (*Support Vector Machine*) – os quais incluem métodos automáticos e acoplados para seleção de entradas e controle de complexidade, incluindo o tamanho da estrutura – número de neurônios na camada oculta no caso do MLP e hiperparâmetros  $C$  e  $\sigma_l$  no caso da SVM [13, 14]. Vale ressaltar que o uso de modelos autônomos, tanto para o MLP quanto para a SVM, é a principal contribuição do presente trabalho.

## 2. ESQUEMA DE CLASSIFICAÇÃO

Inicialmente, foi utilizado o ATP para modelar a subestação de distribuição escolhida para a análise e gerar a base de dados dos eventos a serem classificados. Esses eventos foram escolhidos com base nos eventos mais comuns (curtos-circuitos e eventos associados com qualidade de energia) que ocorrem nos alimentadores dessa subestação e correspondem a um total de 15 tipos diferentes de classes: falta monofásica para a terra (nas fases A, B e C), faltas bifásicas para a terra (nas fases AB, BC e CA), faltas trifásicas para a terra, faltas bifásicas sem envolver terra (também nas fases AB, BC e CA), faltas trifásicas sem envolver terra, abertura do disjuntor de um alimentador, religamento automático de um alimentador, chaveamento de banco de capacitores e sinais sem distúrbio aparente.

Em seguida, foi feito o cálculo da transformada *wavelet* em 10 níveis para a tensão das três fases na barra da subestação sob uma taxa de amostragem de 7680 Hz, com objetivo de extrair o máximo de características dos sinais em análise, minimizando a perda de informação relevante [12]. A *wavelet*-mãe utilizada na decomposição foi a Daubechies-8 (‘db8’), pois *wavelets* dessa família vem sendo utilizadas para classificar eventos dessa natureza e apresentam bons resultados [12]. Com o intuito de reduzir a dimensionalidade da entrada para o estágio de classificação, foi efetuado o cálculo da energia nas várias sub-bandas da transformada *wavelet*, resultando em um vetor de entrada para as redes neurais com 33 elementos (dez valores para os sinais detalhe e um para o sinal aproximação, para cada uma das três fases).

O processo de classificação é dividido em três etapas, sendo a primeira delas um classificador de duas classes, que separa faltas dos demais eventos de interesse. A partir dessa separação, são utilizados outros dois classificadores independentes, sendo um para as faltas e outro para os demais eventos. Cada classificador foi treinado e avaliado utilizando-se conjuntos de dados distintos para treino e teste, e em seguida utilizando-se um terceiro conjunto para avaliação do desempenho da classificação como um todo, tanto para a rede MLP quanto para a SVM.

Para os conjuntos de dados utilizados nos dois métodos de classificação, foram variados os instantes de ocorrência dos eventos, as curvas de carga dos alimentadores, a distância da subestação em que ocorrem os eventos (de pontos mais próximos até pontos mais distantes) e, no caso das faltas, foi variada a resistência de falta, resultando em um total de 6480 instâncias, sendo 432 para cada classe. Nas seções a seguir são abordados os aspectos teóricos envolvidos nos processos de classificação.

## 3. INFERÊNCIA BAYESIANA APLICADA A MLPS

A aplicação de inferência Bayesiana ao desenvolvimento de MLPs foi proposta originalmente por Mackay [15]. Com base na maximização da evidência, os três níveis hierárquicos de inferência são explorados, desde a estimação dos parâmetros até a escolha do modelo mais provável à luz dos dados, passando pelo cálculo dos hiperparâmetros, cuja análise permite o desenvolvimento do método encapsulado de seleção de entradas [13].

Definida a estrutura a ser utilizada num MLP – número de camadas ocultas, número de neurônios por camada e tipo de função de ativação de cada neurônio – e dado o conjunto de pares entrada-saída  $D = \{X, Y\}$ ;  $X = \{x_1, \dots, x_N\}$ ;  $Y = \{d_1, \dots, d_N\}$ , o objetivo do treinamento do modelo sob o ponto de vista da inferência Bayesiana reside na determinação do vetor de parâmetros  $w$  que maximize a probabilidade a posteriori  $p(w|X, Y)$  dada por:

$$p(w|X, Y) = \frac{p(Y|w, X)p(w|X)}{p(Y|X)}, \quad (1)$$

onde  $p(\mathbf{w}|X)$  representa a probabilidade *a priori* do vetor de parâmetros  $\mathbf{w}$ ,  $p(Y|X) = \int p(Y|\mathbf{w}, X)p(\mathbf{w}|X)d\mathbf{w}$  é um fator de normalização e  $p(Y|\mathbf{w}, X)$  é a função de verossimilhança, relacionada com a distribuição de probabilidade de  $x_i$  pertencer a uma dada classe. Considerando uma dicotomia, a função de verossimilhança é dada por:

$$p(Y|\mathbf{w}, X) = \prod_{i=1}^N f(\mathbf{x}_i, \mathbf{w})^{d_i} [1 - f(\mathbf{x}_i, \mathbf{w})]^{1-d_i}, \quad (2)$$

onde  $d_i \in [0, 1]$ , com  $f(\mathbf{x}_i, \mathbf{w})$  respondendo pela saída do MLP, dada por:

$$f(\mathbf{x}_i, \mathbf{w}) = \frac{1}{1 + e^{\left[ -\sum_{i=1}^M w_i \varphi_i \left( \sum_{j=1}^N w_{ij} x_j + b_i \right) \right]}}. \quad (3)$$

Na equação (3),  $\varphi_i(a) : \mathbb{R} \rightarrow \mathbb{R}$  representa a função de ativação do  $i$ -ésimo neurônio da camada oculta,  $b_i$  o *bias* deste neurônio,  $M$  o número de neurônios na camada oculta, e  $\mathbf{w}$  o vetor incluindo todos os parâmetros do modelo, i. e. pesos sinápticos e *bias*.

Para escolha de  $p(\mathbf{w}|X)$  deve ser levado em consideração o fato de que é esperado que grupos diferentes de pesos, por exemplo pesos ligando uma determinada entrada ao modelo e pesos conectando neurônios da camada oculta à saída, apresentem comportamentos distintos ao longo do processo de estimação. Desta forma, é razoável definir distribuições específicas para cada conjunto de pesos, dando origem à distribuição *a priori* dada por:

$$p(\mathbf{w}|X) = \prod_{i=1}^g p(\mathbf{w}_i) = \frac{1}{\prod_{i=1}^g \left( \frac{2\pi}{\alpha_i} \right)^{\frac{M_i}{2}}} e^{\left( -\frac{1}{2} \sum_{i=1}^g \alpha_i \|\mathbf{w}_i\|^2 \right)}, \quad (4)$$

onde  $\mathbf{w}_i$  representa o grupo contendo  $M_i$  parâmetros, com  $\alpha_i$  respondendo pelo hiperparâmetro dado pelo inverso da variância da distribuição Gaussiana com vetor média nulo, utilizada para representação *a priori* de  $\mathbf{w}_i$ , e  $g$  relacionado com o número de grupos de parâmetros. Um agrupamento específico dos parâmetros do modelo, incluindo um grupo para os pesos que ligam cada entrada ao modelo, dá origem ao método automático de estimação da relevância das entradas [16].

Tendo as distribuições  $p(Y|\mathbf{w}, X)$  e  $p(\mathbf{w}|X)$ , a probabilidade *a posteriori*  $p(\mathbf{w}|X, Y)$  é dada por:

$$p(\mathbf{w}|X, Y) = \frac{1}{Z_S} e^{-S(\mathbf{w})}, \quad (5)$$

sendo

$$S(\mathbf{w}) = -\sum_{i=1}^n d_n \ln f(\mathbf{x}_i, \mathbf{w}) + (1 - d_n) \ln [1 - f(\mathbf{x}_i, \mathbf{w})] + \frac{1}{2} \sum_{i=1}^g \left( \alpha_i \sum_{j=1}^{M_i} w_{ij}^2 \right), \quad (6)$$

onde  $Z_s = \int e^{-S(\mathbf{w})} d\mathbf{w}$  é um fator de normalização.

Portanto, para maximizar  $p(\mathbf{w}|X, Y)$  é necessário minimizar  $S(\mathbf{w})$ . Para tal, é utilizada a chamada aproximação da evidência, a qual além de fornecer estimativas para  $\mathbf{w}$ , fornece também estimativas para  $\alpha_i$ , cuja comparação com limiares de relevância empíricos pode ser utilizada para identificação e retirada de entradas irrelevantes. Maiores detalhes sobre a estimação de  $\mathbf{w}$  e  $\alpha_i$  podem ser encontrados nos trabalhos de Mackay [15] e Bishop [16]. A definição empírica de limiares de relevância pode ser encontrada com maior riqueza de detalhes nos trabalhos de Ferreira e da Silva [13, 14].

De posse das estimativas de  $\mathbf{w}$  e  $\alpha_i$ , a maximização da evidência para os modelos pode ser utilizada para avaliação analítica da adequação de um conjunto aninhado de modelos, i. e. vários MLP com diferentes números de neurônios nas suas camadas ocultas, aos dados  $D = \{X, Y\}$ . Tal avaliação é dada pela evidência para os modelos, cujo logaritmo é dado pela expressão:

$$\ln p(Y|H_h) = -S(\mathbf{w}) - \frac{1}{2} \ln |\mathbf{A}(\mathbf{w})| + \frac{1}{2} \sum_{i=1}^g M_i \alpha_i + 2 \ln m + \ln m! + \frac{1}{2} \sum_{i=1}^g \ln \left( \frac{2}{\gamma_i} \right) + \frac{1}{2} \ln \left( \frac{2}{N - \gamma} \right), \quad (7)$$

onde  $\gamma_i$  representa o número efetivo de parâmetros estimados para o  $i$ -ésimo grupo de pesos  $\mathbf{w}_i^* = [w_{i1}^*, \dots, w_{iM}^*]^T$  e  $\gamma$  o número efetivo de parâmetros estimados do modelo, dados por:

$$\gamma = \sum_{i=1}^g \gamma_i; \quad (8)$$

$$\gamma_i = \alpha_i \sum_{j=1}^{M_i} (w_{ij}^*)^2.$$

#### 4. MÁQUINAS DE VETOR SUPORTE

As máquinas de vetor suporte (SVM) foram desenvolvidas com base em um novo paradigma da área de aprendizado de máquina conhecido como aprendizado estatístico. Diferentemente da abordagem clássica para problemas de classificação, que

necessita de uma quantidade elevada de dados em conjunto com a inserção de conhecimento prévio sobre o problema, a teoria do aprendizado estatístico foi desenvolvida para solução de problemas cuja quantidade de dados disponíveis é reduzida, onde pouco ou até mesmo nenhum conhecimento prévio pode ser utilizado, características estas comumente encontradas em aplicações reais [17].

A teoria das SVM foi originalmente desenvolvida para solução de problemas de classificação através da aplicação do conceito de hiperplano ótimo, baseando-se na maximização da margem de separação  $\rho$ . A motivação para a maximização de  $\rho$  encontra fundamento em uma medida de complexidade conhecida como dimensão de Vapnik e Chervonenkis, popularmente denominada dimensão  $VC$  [17]. De acordo com o dilema *bias*-variância [18], o desempenho do modelo para novos dados pode ser decomposto em duas parcelas conflitantes – *bias* e variância, as quais estão relacionadas com o ajuste aos dados disponíveis e com o nível de flexibilidade da função estimada, respectivamente. Modelos excessivamente ajustados aos padrões de treinamento irão apresentar *bias* reduzido, porém elevada variância em virtude do grau de complexidade fornecido. Analogamente, modelos com elevada dimensão  $VC$ , apesar de se ajustarem de forma satisfatória aos dados de treinamento, apresentarão reduzida capacidade de generalização.

Apesar de não possuir expressão analítica geral para modelos não-lineares, a dimensão  $VC$  de hiperplanos  $F(\mathbf{X}_i, \mathbf{w})$  com margem de separação  $\rho$  é limitada superiormente [17] por:

$$VC[F(\mathbf{X}, \mathbf{w})] \leq \frac{R^2}{\rho^2}, \quad (9)$$

onde  $R$  é o raio da menor hipersfera que engloba a imagem dos padrões de treinamento no espaço de características.

O conceito de hiperplano ótimo de separação pode ser expandido para problemas de classificação de padrões não-linearmente separáveis através do mapeamento do espaço original de representação para um espaço de dimensão superior, onde a probabilidade do problema ser linearmente separável é elevada. Assim, as SVM podem ser vistas como máquinas lineares aplicadas a um espaço de representação expandido, de dimensão maior que o espaço de representação original do problema, com o mapeamento que governa esta expansão sendo obtido de maneira intrínseca. Matematicamente, para problemas de classificação binária a saída de uma SVM pode ser dada por:

$$\begin{aligned} f(\mathbf{x}, \mathbf{W}, b) &= \text{sgn}[\mathbf{W}^T \Phi(\mathbf{x}) + b]; \\ \mathbf{W} &= [W_1, \dots, W_N]^T; \\ \Phi(\mathbf{x}) &= [\phi_1(\mathbf{x}), \dots, \phi_N(\mathbf{x})]^T, \end{aligned} \quad (10)$$

onde  $\Phi(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^N$  representa o mapeamento não-linear das entradas no espaço de características, com  $\mathbf{W}$  respondendo pelo conjunto de parâmetros do modelo,  $b$  pelo *bias* e  $\text{sgn}[a]$  representando a função sinal:

$$\text{sgn}[a] = \begin{cases} 1, & \text{se } a \geq 0 \\ 0, & \text{se } a < 0 \end{cases} . \quad (11)$$

Assim, a maximização da margem de separação  $\rho$  pode ser formulada pelo seguinte problema de otimização restrita [17]:

$$\min_{\mathbf{W}, b, \xi} E_s(\mathbf{W}) = \frac{1}{2} \mathbf{W}^T \mathbf{W} + C \sum_{i=1}^N \xi_i, \quad (12)$$

s.a.

$$\begin{cases} d_i[\mathbf{W}^T \Phi(\mathbf{x}) + b] \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases}, \quad i = 1, 2, \dots, N.$$

Na equação (12), o primeiro termo da função objetivo é responsável pelo controle de complexidade do modelo por meio da maximização de  $\rho$ . O segundo termo está relacionado com o erro de classificação para o conjunto de dados, visto que para dados corretamente classificados  $\xi_i$  é igual a zero. O hiperparâmetro  $C$  é responsável pelo equilíbrio entre a complexidade do modelo e o ajuste aos dados de treinamento, sendo desta forma denominado parâmetro de regularização [18].

O problema de otimização quadrática formulado em (12) pode ser resolvido utilizando o método dos multiplicadores de Lagrange, cuja formulação dual é dada por:

$$\max_{\alpha} \Psi(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N d_i d_j K(\mathbf{x}_i, \mathbf{x}_j) \alpha_i \alpha_j, \quad (13)$$

s.a.

$$\begin{cases} 0 \leq \alpha_i \leq C \\ \sum_{i=1}^N \alpha_i d_i = 0 \end{cases}, \quad i = 1, 2, \dots, N.$$

onde  $\alpha$  representa o conjunto de multiplicadores de Lagrange e  $K(\mathbf{x}_i, \mathbf{x}_j)$  o *kernel* (núcleo) do produto interno no espaço de características, ou seja:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j). \quad (14)$$

Existem diversos tipos de *kernel*  $K(\mathbf{x}_i, \mathbf{x}_j)$ , os quais devem obedecer às condições do teorema de Mercer [19]. Neste trabalho é utilizado o *kernel* Gaussiano dado por:

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{\left[-\sum_{l=1}^N \sigma_l^2 (x_{il} - x_{jl})^2\right]}, \quad (15)$$

onde  $\sigma_l^2$ ,  $l = 1, 2, \dots, N$ , representam os hiperparâmetros do *kernel*.

No ponto ótimo da equação (13) nem todos  $\alpha_i^*$  são não-nulos. Os vetores para os quais  $\alpha_i^*$  são diferentes de zero são os chamados vetores suporte, os quais definem a superfície de decisão da SVM pela expressão:

$$f(\mathbf{x}, \mathbf{W}, b) = \text{sgn} \left[ \sum_{i=1}^{N_S} \alpha_i d_i K(\mathbf{x}_i, \mathbf{x}) + b \right], \quad (16)$$

onde  $N_S$  representa o número de vetores suporte.

Apesar de apresentarem preocupação com o controle de complexidade da estrutura na sua formulação e produzirem como subproduto do algoritmo de treinamento a estrutura do modelo por meio do número de vetores suporte, as SVM apresentam alguns hiperparâmetros a serem especificados pelo usuário, quais sejam: constante de regularização  $C$  e hiperparâmetros  $\sigma_l^2$  do *kernel*. Comumente selecionados via validação cruzada, neste trabalho estes hiperparâmetros são selecionados por meio da minimização de um limite superior do erro de generalização estimado via validação cruzada única (*leave-one-out*). Este método de reamostragem fornece uma estimativa quase não-tendenciosa para a capacidade de generalização do modelo [19], porém é computacionalmente intensivo. Por outro lado, o limite superior  $T[f(\mathbf{x}, \mathbf{W}, b)]$  utilizado neste trabalho é analítico, desenvolvido por Vapnik e Chappelle [20] tendo por base o conceito de extensão dos vetores suporte (*span of support vectors*), sendo dado pela expressão:

$$T[f(\mathbf{x}, \mathbf{W}, b)] = \sum_{i=1}^{N_S} \alpha_i S_i^2 \quad (17)$$

onde  $S_i^2$  representa a extensão do  $i$ -ésimo vetor suporte, dada por:

$$S_i^2 = \frac{1}{\left(\tilde{\mathbf{K}}^{-1}\right)_{ii}}. \quad (18)$$

Na equação (18),  $\left(\tilde{\mathbf{K}}^{-1}\right)_{ii}$  representa o  $i$ -ésimo elemento da diagonal da inversa da matriz  $\tilde{\mathbf{K}}$ , dada por:

$$\tilde{\mathbf{K}} = \begin{bmatrix} \mathbf{K}_{N_S} & \mathbf{u} \\ \mathbf{u}^T & \mathbf{0} \end{bmatrix}, \quad (19)$$

onde  $\mathbf{K}_{N_S}$  representa a matriz com o núcleo do produto interno entre todos os vetores suporte e  $\mathbf{u} \in \mathbb{R}^{N_S}$  um vetor unitário.

Diante da característica multimodal da equação (17) [21], confirmada para problemas de regressão nos trabalhos de Ferreira e da Silva [13, 14], neste trabalho são utilizados algoritmos genéticos [22] para minimização de  $T[f(\mathbf{x}, \mathbf{W}, b)]$  visando estimar  $C$  e  $\sigma_l^2$ .

## 5. RESULTADOS

A tabela 1 apresenta o resultado da classificação em três estágios utilizando MLP, além das características dos modelos em cada um dos estágios – número de neurônios na camada intermediária e número de entradas selecionadas. A taxa de acerto individual representa o resultado da classificação em cada estágio independentemente do processo como um todo. A taxa de acerto total utiliza um conjunto de teste que leva em conta a divisão inicial do classificador 1, bem como os erros provenientes desse classificador. Uma observação importante é que o acerto individual e o acerto total do classificador 1 são iguais, uma vez que este constitui o primeiro estágio do processo.

Tabela 1: Taxas de acerto em três estágios – MLP

Classificador	Acerto Individual	Acerto Total	Número de Neurônios	Número de Entradas
1 (2 classes)	96,1 %	96,1 %	15	33
2 (4 classes)	89,7 %	85,2 %	17	32
3 (11 classes)	93,1 %	92,1 %	18	33

Na tabela 2 são apresentados os resultados dos mesmos experimentos de classificação, porém utilizando SVM nos três estágios. Neste caso são apresentados os números médios das entradas selecionadas e dos vetores suporte. A utilização de valores médios foi escolhida em função do treinamento do modelo SVM ser realizado em pares de classes, dificultando a apresentação desses resultados com maiores detalhes. Para cada par de classes é determinada uma quantidade de vetores suporte e de entradas

Tabela 2: Taxas de acerto em três estágios – SVM

Classificador	Acerto Individual	Acerto Total	Número Médio de Vetores de Suporte	Número Médio de Entradas
1 (2 classes)	92,3 %	92,3 %	1191	21
2 (4 classes)	93,8 %	87,6 %	153	23
3 (11 classes)	86,1 %	78,6 %	55	26

selecionadas, sendo que a média representa de forma aproximada as características de complexidade do modelo – informação equivalente ao número de neurônios do MLP – e entradas relevantes para o processo de classificação.

Nas duas abordagens, o primeiro classificador distingue faltas (faltas monofásicas para a terra, faltas bifásicas para a terra, faltas trifásicas para a terra, faltas bifásicas sem envolver terra e faltas trifásicas sem envolver terra) dos demais eventos (abertura do disjuntor de um alimentador, religamento automático de um alimentador, chaveamento de banco de capacitores e sinais sem distúrbio aparente). A primeira linha das tabelas 1 e 2 mostra a taxa de acertos para esta primeira dicotomia. A segunda e terceira linhas mostram tanto as taxa de acertos dos classificadores 2 e 3 quanto as suas respectivas taxas de acerto acumuladas, onde passam a ser considerados eventuais erros cometidos pelo primeiro estágio. As tabelas 1 e 2 mostram que os modelos apresentam boa capacidade de classificação, especialmente o primeiro estágio de distinção entre faltas e demais eventos, onde a taxa de acerto chega a 96% para o caso do MLP.

A tabela 3 mostra a síntese dos resultados dos métodos de classificação testados. Nela são considerados os resultados do processo global do classificador em três estágios, tanto para MLP quanto para SVM, contrastados com os resultados obtidos para um classificador único de 15 classes, também para ambas as abordagens. Apesar de apresentar taxa de acerto maior para cada estágio independente, o sistema de classificação baseado em MLP apresentou taxa de acerto ligeiramente inferior em relação ao baseado em SVM. Por outro lado, a complexidade dos modelos baseados em SVM mostrou-se mais elevada que a complexidade dos modelos baseados em MLP, que considerou praticamente a totalidade das entradas como relevantes.

Tabela 3: Taxas de acerto global – MLP e SVM

Método de Classificação	Acerto Total
MLP (três estágios)	87,1 %
MLP (estágio único)	86,3 %
SVM (três estágios)	91,8 %
SVM (estágio único)	89,5 %

De acordo com a tabela 3, a separação do processo de classificação em três estágios fornece um sensível aumento na taxa de acerto total da classificação, tanto para modelos baseados em MLP quanto para modelos baseados em SVM, quando comparado com a taxa de acerto de classificadores únicos de 15 classes. Além disso, essa separação em três estágios fornece flexibilidade na análise de eventos de mesma natureza, como no caso das faltas e demais eventos relacionados à qualidade de energia.

## 6. CONCLUSÃO

Este trabalho apresentou a viabilidade da aplicação de um método baseado em redes neurais artificiais do tipo MLP e SVM, que dispõem de autonomia para classificação de eventos em redes de distribuição de energia elétrica, na medida em que incluem técnicas automáticas e acopladas para controle de complexidade da sua estrutura e seleção de entradas relevantes. Isso permite que um sistema que utilize esse método de classificação possa operar de forma independente dos demais estágios do processo, tais como pré-processamento e definição dos conjuntos de treino e teste que, para o problema em questão, dependem da subestação de distribuição específica na qual são feitas as aquisições de dados.

As tabelas 1, 2 e 3 resumem as principais características dos métodos aqui apresentados para a classificação de eventos em redes de distribuição de energia elétrica. Essas características estão relacionadas à autonomia das redes neurais no que diz respeito à seleção do modelo mais adequado aos dados apresentados, levando em conta as entradas mais importantes para o modelo em questão e tornando o processo de classificação independente de qualquer conhecimento prévio. As entradas mais relevantes podem servir de base para uma análise das características em frequência dos sinais de entrada das redes neurais, uma vez que essas entradas são o resultado de um processo de decomposição *wavelet* das tensões na barra da subestação.

Finalmente, a aplicação de um classificador de eventos em redes de distribuição de energia elétrica – especialmente no caso da classificação de faltas – permite que posteriormente sejam implementadas técnicas de localização de faltas baseadas nas oscilografias que foram classificadas, auxiliando com isso as equipes de manutenção das concessionárias de energia. Como trabalho futuro, pretende-se investigar a influência do uso de outras *wavelets*-mãe e outras formas de pré-processamento no desempenho dos classificadores.

## REFERÊNCIAS

- [1] V. Ferreira, A. Lazzaretti, H. Vieira Neto, R. Riella e J. Omori. “Classificação de Eventos em Redes de Distribuição de Energia Utilizando Transformada Wavelet e Modelos Neurais Autônomos”. In *Anais do IX Congresso Brasileiro de Redes Neurais / Inteligência Computacional*, Ouro Preto, Outubro de 2009.
- [2] J. Morais, Y. Pires, C. Cardoso and A. Klautau. “Data Mining Applied to the Electric Power Industry: Classification of Short-Circuit Faults in Transmission Lines”. In *Proceedings of the 7th International Conference on Intelligent Systems Design and Application*, pp. 943–948, October 2007.
- [3] ANEEL. *Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional – PRODIST, Módulo 8 – Qualidade de Energia Elétrica*. Agência Nacional de Energia Elétrica, 2007.
- [4] R. J. Riella, V. P. Ferrari, G. Paulillo, M. R. Ortega e J. G. Pereira. “Desenvolvimento de um Sistema de Monitoramento Contínuo da Qualidade da Energia Elétrica para Subestações de Distribuição”. In *Anais do XVIII Seminário Nacional de Distribuição de Energia Elétrica – SENDI*, Olinda, Outubro de 2008.
- [5] EMTP. *Alternative Transients Program (ATP) Rule Book*. Can Am EMTP User’s Group, 1995.
- [6] R. C. Dugan, M. F. McGranaghan, S. Santoso and H. W. Beaty. *Electrical Power System Quality*. McGraw-Hill, New York, second edition, 2002.
- [7] K. M. Silva, B. A. Souza and N. S. D. Brito. “Fault Detection and Classification in Transmission Lines Based on Wavelet Transform and ANN”. *IEEE Trans. on Power Delivery*, vol. 21, no. 4, pp. 2058–2063, October 2006.
- [8] H. Zheng-You, C. Xiaoqing and L. Guoming. “Wavelet Entropy Measure Definition and Its Application for Transmission Line Fault Detection and Identification”. In *Proceedings of the 2006 International Conference on Power System Technology*, pp. 1–6, October 2006.
- [9] O. Dag and C. Ucak. “Fault Classification for Power Distribution Systems via a Combined Wavelet-Neural Approach”. In *Proceedings of the 2006 International Conference on Power System Technology*, volume 2, pp. 1309–1314, November 2004.
- [10] H. Khorashadi-Zadeh and M. R. Aghaebrahimi. “A Novel Approach to Fault Classification and Fault Location for Medium Voltage Cables Based on Artificial Neural Network”. *International Journal of Computational Intelligence*, vol. 2, no. 2, pp. 90–93, Spring 2006.
- [11] J. Arrillaga, B. C. Smith, N. R. Watson and A. R. Wood. *Power System Harmonics*. Wiley, 2003.
- [12] O. Delmont Filho. “Um Algoritmo para Detecção, Localização e Classificação de Distúrbios em Qualidade da Energia Elétrica Utilizando a Transformada Wavelet”. Tese de doutorado, EESC/USP, São Carlos, 2007.
- [13] V. H. Ferreira and A. P. A. da Silva. “Toward Estimating Autonomous Neural Network Load Forecasters”. *IEEE Trans. on Power Systems*, vol. 22, no. 4, pp. 1554–1562, November 2007.
- [14] V. H. Ferreira. “Desenvolvimento de Modelos Neurais Autônomos para Previsão de Carga Elétrica”. Tese de doutorado, COPPE/UFRJ, Rio de Janeiro, 2008.
- [15] D. J. C. Mackay. “Bayesian Methods for Adaptive Models”. Ph.D. thesis, California Institute of Technology, Pasadena, 1992.
- [16] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, 1995.
- [17] V. N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, New York, 1998.
- [18] V. Cherkassky and F. F. Mulier. *Learning from Data: Concepts, Theory and Methods*. Wiley, 1998.
- [19] B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press, 2001.
- [20] V. Vapnik and O. Chapelle. “Bounds on Error Expectation for Support Vector Machines”. *Neural Computation*, vol. 12, no. 9, pp. 2013–2036, September 2000.
- [21] O. Chapelle, V. Vapnik, O. Bousquet and S. Mukherjee. “Choosing Multiple Parameters for Support Vector Machines”. *Machine Learning*, vol. 46, no. 1-3, pp. 131–159, January 2002.
- [22] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Professional, Reading, 1989.