

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

VITOR FERREIRA*, ANDRÉ LAZZARETTI†, HUGO VIEIRA NETO‡, RODRIGO RIELLA†, JULIO OMORI§

**Universidade Federal Fluminense – UFF*
Rua Passo da Pátria, 156 – Niterói-RJ

†*Instituto de Tecnologia para o Desenvolvimento – LACTEC*
Avenida Comendador Franco, 1341 – Curitiba-PR

‡*Universidade Tecnológica Federal do Paraná – UTFPR*
Avenida Sete de Setembro, 3165 – Curitiba-PR

§*Companhia Paranaense de Energia – COPEL*
Rua Jose Izidoro Biazzetto, 158 – Curitiba-PR

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

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.

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.

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. 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 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 Desenvolvi-

mento (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 (Riella et al., 2008).

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 regis-

trada, resultando em falta de classificações adequadas para os vários eventos, na maioria das oscilografias armazenadas.

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*), com o intuito de ser utilizado para treinar e testar o estágio de classificação do sistema proposto. Grande parte das oscilografias reais armazenadas é resultante de faltas no sistema elétrico (cerca de 70%) e o restante resultante de eventos de qualidade de energia. Essa proporção foi mantida nas simulações utilizadas no presente trabalho.

Foi escolhido um pré-processamento baseado no cálculo da transformada *wavelet* (TW) (Dag e Ucak, 2004; Khorashadi-Zadeh e Aghaebrabimi, 2006), em função do reduzido tamanho do vetor de características obtido, quando comparado com o vetor produzido pela transformada de Fourier (TF), fato que facilita o processamento dos dados como um todo, sem perdas na capacidade de classificação. Em seguida, na etapa de classificação, 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 σ_i no caso da SVM) (Ferreira e da Silva, 2007). Vale ressaltar que essa metodologia (modelos autônomos), tanto para a SVM, quanto para a MLP, 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 (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 TW 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. A *wavelet*-mãe utilizada na decomposição foi a Daubechies-8 ('db8'). A justificativa para a utilização dessa *wavelet*-mãe é que *wavelets* dessa família já são utilizadas para classificar eventos dessa natureza, apresentando bons resultados. 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 TW, 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 treino e teste separados, e em seguida utilizando-se um conjunto para avaliação do desempenho da classificação como um todo, tanto para a rede MLP quanto para a SVM.

Para os conjuntos de treino e teste 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 no processo de classificação.

3 Inferência Bayesiana aplicada a MLPs

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 $D = \{X, Y\}$; $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, $Y = \{d_1, \dots, d_N\}$, o objetivo do treinamento do modelo sob o ponto de vista da inferência Bayesiana reside na determinação do vetor de parâmetros \mathbf{w} que maximize a probabilidade a posteriori $p(\mathbf{w}|X, Y)$ dada por

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

onde $p(\mathbf{w}|X)$ representa a probabilidade *a priori* do vetor de parâmetros \mathbf{w} , $p(Y|X)$ é um fator de normalização e $p(Y|\mathbf{w}, X)$ é a função de verossimilhança, relacionada com a distribuição de probabi-

lidade de \mathbf{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

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

onde \mathbf{w}_i representa o grupo contendo M_i parâmetros, com α_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 (Bishop, 1995).

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

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

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 α_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 α_i podem ser encontrados nos trabalhos de Mackay (1992) e Bishop (1995). A definição empírica de limiares de relevância pode ser encontrada com maior riqueza de detalhes no trabalho de Ferreira e da Silva (2007).

De posse das estimativas de \mathbf{w} e α_i , a maximização da evidência para os modelos pode ser utilizada para avaliação analítica da adequação de um conjunto aninhado de modelos, i.e. MLPs com diferentes números de neurônios na camada oculta, aos dados $D = \{X, Y\}$. Tal avaliação é dada pela evidência para os modelos, cujo logaritmo é dado pela expressão

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

onde γ_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 γ o número efetivo de parâmetros estimados do modelo, dados por

$$\gamma = \sum_{i=1}^g \gamma_i; \quad \gamma_i = \alpha_i \sum_{j=1}^{M_i} (w_{ij}^*)^2. \quad (8)$$

4 Máquinas de vetor de suporte

As máquinas de vetor de suporte (SVM) foram originalmente desenvolvidas para solução de problemas de classificação através da aplicação do conceito de hiperplano ótimo, baseado na maximização da margem de separação ρ . A motivação para a maximização de ρ encontra fundamento em uma medida de complexidade conhecida como dimensão de Vapnik e Chervonenkis, popularmente denominada dimensão VC (Vapnik, 1998). Matematicamente, a saída de uma SVM pode ser dada por

$$\begin{aligned} f(\mathbf{x}, \mathbf{W}, b) &= \text{sgn}[\mathbf{W}^T \Phi(\mathbf{x}) + b]; \\ \mathbf{W} &= [W_1, \dots, W_N]^T; \\ \Phi(\mathbf{x}) &= [\phi_1(\mathbf{x}), \dots, \phi_N(\mathbf{x})]^T, \end{aligned} \quad (9)$$

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:

$$\text{sgn}[a] = \begin{cases} 1, & \text{se } a \geq 0 \\ 0, & \text{se } a < 0 \end{cases} \quad (10)$$

Assim, a maximização da margem de separação ρ pode ser formulada pelo seguinte problema de otimização restrita (Vapnik, 1998):

$$\min_{\mathbf{W}, b, \xi} E_s(\mathbf{W}) = \frac{1}{2} \mathbf{W}^T \mathbf{W} + C \sum_{i=1}^N \xi_i, \quad (11)$$

s.a.

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

Em (11), o primeiro termo da função objetivo é responsável pelo controle de complexidade do modelo por meio da maximização de ρ . O segundo termo está relacionado com o erro de classificação para o conjunto de dados, visto que para dados corretamente classificados ξ_i é igual a zero. O hiperparâmetro C é responsável pelo equilíbrio entre a complexidade do modelo e o ajuste dos dados de treinamento, sendo desta forma denominado parâmetro de regularização.

O problema de otimização quadrática formulado em (11) pode ser resolvido utilizando o método dos multiplicadores de Lagrange, cuja formulação dual é dada por

$$\max_{\alpha} \Psi(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N d_i d_j K(\mathbf{x}_i, \mathbf{x}_j) \alpha_i \alpha_j, \quad (12)$$

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 α 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,

$$K(\mathbf{x}_i, \mathbf{x}_j) = [\Phi(\mathbf{x}_i)]^T \Phi(\mathbf{x}_j). \quad (13)$$

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

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\sum_{l=1}^N \sigma_l^2 (x_{il} - x_{jl})^2}, \quad (14)$$

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

No ponto ótimo de (12) nem todos α_i^* são não-nulos. Os vetores para os quais α_i^* são diferentes de zero são os chamados vetores de suporte, os quais definem a superfície de decisão da SVM pela expressão:

$$f(\mathbf{x}, \mathbf{W}, b) = \text{sgn} \left[\sum_{i=1}^{N_S} \alpha_i d_i K(\mathbf{x}_i, \mathbf{x}) + b \right], \quad (15)$$

onde N_S representa o número de vetores de 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 de suporte, as SVM apresentam alguns hiperparâmetros a serem especificados pelo usuário, quais sejam: constante de regularização C e hiperparâmetros do *kernel* σ_i^2 . 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, porém é computacionalmente intensivo. Por outro lado, o limite superior utilizado neste trabalho é analítico, desenvolvido por Vapnik e Chappelle (2000) tendo por base o conceito de extensão dos vetores de suporte (*span of support vectors*). Neste trabalho são utilizados algoritmos genéticos (Goldberg, 1989) para minimização desse limite e a consequente estimativa de C e σ_i^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). O acerto individual representa o resultado da classificação em cada estágio independentemente do processo como um todo. Já o 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: Classificação em Três Estágios – MLP

Classificador	Acerto Indiv.	Acerto Total	Nº de Neurônios	Nº 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 de 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 neste trabalho, devido a limitações de espaço. Para cada par de classes é determinada uma quantidade de vetores de suporte e de entradas selecionadas, sendo que a média representa de forma aproximada as características de complexidade do modelo (equivalente ao número de neurônios do MLP) e entradas relevantes para o processo de classificação.

Tabela 2: Classificação em Três Estágios – SVM

Classificador	Acerto Indiv.	Acerto Total	Nº Médio Vet. Sup.	Nº Médio 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

A análise das tabelas 1 e 2 resume 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 do modelo. Além disso, 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.

O acerto final foi de 87,1 % para MLP e 91,8 % para SVM. A separação do processo de classificação em três estágios fornece um aumento no acerto total da classificação, tanto para MLP quanto para SVM, quando comparado com um classificador único de 15 classes (86,3% para MLP e 89,5% para SVM). Além disso, essa separação fornece flexibilidade na análise de eventos de mesma natureza, como no caso das faltas e eventos de qualidade de energia.

6 Conclusão

Este trabalho apresentou a viabilidade da aplicação de um método baseado em redes neurais 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 estrutura e seleção de entradas das redes neurais. 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 na qual são feitas as aquisições de dados. Finalmente, a aplicação de um classificador – 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 de 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

- Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, UK.
- Dag, O. e Ucak, C. (2004). Fault classification for power distribution systems via a combined wavelet-neural approach, *Proceedings of the 2006 International Conference on Power System Technology*, Vol. 2, pp. 1309–1314.
- Ferreira, V. H. e da Silva, A. P. A. (2007). Toward estimating autonomous neural network load forecasters, *IEEE Transactions on Power Systems* **22**(4): 1554–1562.
- Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Professional, Reading, MA.
- Khorashadi-Zadeh, H. e Aghaebrahimi, M. R. (2006). A novel approach to fault classification and fault location for medium voltage cables based on artificial neural network, *International Journal of Computational Intelligence* **2**(2): 90–93.
- Mackay, D. J. C. (1992). *Bayesian Methods for Adaptive Models*, PhD thesis, California Institute of Technology, Pasadena, CA.
- Riella, R. J., Ferrari, V. P., Paulillo, G., Ortega, M. R. e Pereira, J. G. (2008). Desenvolvimento de um sistema de monitoramento contínuo da qualidade da energia elétrica para subestações de distribuição, *Anais do XVIII Seminário Nacional de Distribuição de Energia Elétrica – SENDI*, Olinda.
- Vapnik, V. e Chappelle, O. (2000). Bounds on error expectation for support vector machines, *Neural Computation* **12**(9): 2013–2036.
- Vapnik, V. N. (1998). *Statistical Learning Theory*, Wiley-Interscience, New York, NY.